frontiers | Research Topics

Epitope mapped vaccines and diagnostics for emerging pathogens

Edited by
Tarek A. Ahmad and Adriana Harbuzariu

Published in
Frontiers in Medicine
Frontiers in Immunology

ISSN 1664-8714
ISBN 978-2-83251-301-9
DOI 10.3389/978-2-83251-301-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Epitope mapped vaccines and diagnostics for emerging pathogens

Topic editors

Tarek A. Ahmad — Bibliotheca Alexandrina, Egypt

Adriana Harbuzariu — Emory University, United States

Citation

Ahmad, T. A., Harbuzariu, A., eds. (2023). *Epitope mapped vaccines and diagnostics for emerging pathogens*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-301-9

Table of contents

frontiers in Medicine

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fmed.2021.825876

Evaluation of the Whole Proteome of *Achromobacter xylosoxidans* to Identify Vaccine Targets for mRNA and Peptides-Based Vaccine Designing Against the Emerging Respiratory and Lung Cancer-Causing Bacteria

Taimoor Khan[1†], Muhammad Abdullah[2], Tayyba Fatima Toor[3], Fahad N. Almajhdi[4], Muhammad Suleman[5], Arshad Iqbal[5], Liaqat Ali[6], Abbas Khan[1†], Yasir Waheed[7] and Dong-Qing Wei[1,8,9*]*

[1] Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [2] Amna Inayat Medical College, Lahore, Pakistan, [3] Central Park Medical College, Lahore, Pakistan, [4] Department of Botany and Microbiology, College of Sciences, King Saud University, Riyadh, Saudi Arabia, [5] Centre for Biotechnology and Microbiology, University of Swat, Kanju, Pakistan, [6] Division of Biology, Kansas State University, Manhattan, KS, United States, [7] Foundation University Medical College, Foundation University Islamabad, Islamabad, Pakistan, [8] State Key Laboratory of Microbial Metabolism, Joint Laboratory of International Laboratory of Metabolic and Developmental Sciences, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [9] Peng Cheng Laboratory, Shenzhen, China

Achromobacter xylosoxidans is a rod-shaped Gram-negative bacterium linked with causing several infections which mostly includes hematological malignancies. It has been recently reported to be associated with the development and progression of lung cancer and is an emerging respiratory disease-causing bacterium. The treatment of individuals infected with *A. xylosoxidans* bacteremia is difficult due to the fact that this pathogen has both intrinsic and acquired resistance mechanisms, typically resulting in a phenotype of multidrug resistance (MDR). Efforts are needed to design effective therapeutic strategies to curtail the emergence of this bacterium. Computational vaccine designing has proven its effectiveness, specificity, safety, and stability compared to conventional approaches of vaccine development. Therefore, the whole proteome of *A. xylosoxidans* was screened for the characterization of potential vaccine targets through subtractive proteomics pipeline for therapeutics design. Annotation of the whole proteome confirmed the three immunogenic vaccine targets, such as (E3HHR6), (E3HH04), and (E3HWA2), which were used to map the putative immune epitopes. The shortlisted epitopes, specific against Cytotoxic T Lymphocytes, Helper T-cell Lymphocytes, and linear B-Cell, were used to design the mRNA and multi-epitopes vaccine (MEVC). Initial validations confirmed the antigenic and non-allergenic properties of these constructs, followed by docking with the immune receptor, TLR-5, which resulted in robust interactions. The interaction

OPEN ACCESS

Edited by:
Tarek A. Ahmad, Bibliotheca Alexandrina, Egypt

Reviewed by:
Muhammad Tahir Ul Qamar, Guangxi University, China
Maha A. El Demellawy, City of Scientific Research and Technological Applications, Egypt
Mounir M. Salem Bekhit, King Saud University, Saudi Arabia

***Correspondence:**
Yasir Waheed
yasir_waheed_199@hotmail.com
Dong-Qing Wei
dqwei@sjtu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases–Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *30 November 2021*
Accepted: *29 December 2021*
Published: *04 February 2022*

Citation:
Khan T, Abdullah M, Toor TF, Almajhdi FN, Suleman M, Iqbal A, Ali L, Khan A, Waheed Y and Wei D-Q (2022) Evaluation of the Whole Proteome of Achromobacter xylosoxidans to Identify Vaccine Targets for mRNA and Peptides-Based Vaccine Designing Against the Emerging Respiratory and Lung Cancer-Causing Bacteria. Front. Med. 8:825876. doi: 10.3389/fmed.2021.825876

pattern that followed in the docking complex included formation of 5 hydrogen bonds, 2 salt bridges, and 165 non-bonded contacts. This stronger binding affinity was also assessed through using the mmGBSA approach, showing a total of free binding energy of −34.64 kcal/mol. Further validations based on *in silico* cloning revealed a CAI score of 0.98 and an optimal percentage of GC contents (54.4%) indicated a putatively higher expression of the vaccine construct in *Escherichia coli*. Moreover, immune simulation revealed strong antibodies production upon the injection of the designed MEVC that resulted in the highest peaks of IgM+ IgG production (>3,500) between 10 and 15 days. In conclusion the current study provide basis for vaccine designing against the emerging *A. xylosoxidans,* which demands further experimental studies for *in vitro* and *in vivo* validations.

Keywords: *Achromobacter xylosoxidans*, lungs cancer, vaccine targets, mRNA-based vaccines, immune simulation

INTRODUCTION

Achromobacter xylosoxidans is a motile, oxidase++, aerobic, and gram-negative rod-shaped bacterium extensively dispersed in the environment and reported to be associated with healthcare infection particularly the hematological malignancies (1). This bacterium has been reported to be associated with bacteremia in patients suffering from cystic fibrosis (2). The emerging bacterium is also associated with lung disease severity in children and lung inflammation in patients (3, 4). The bacterium mainly colonizes in the patient's airway and follows a complex mechanism of evolutionary dynamics related with host/pathogen interactions (5). It was originally isolated from patients with otitis media, where associated complications including pharyngitis, pneumonia, peritonitis, and urinary tract infections (6). Moreover, *A. xylosoxidans* has been recently reported to be associated with the development and progression of lung cancer. It is also reported as an emerging respiratory disease-causing bacterium (7). Infection in the lungs, with *A. xylosoxidans,* has been linked to a variety of health complications, notably IgM deficiency and acute myelogenous leukemia (AML) (8, 9).

The treatment of individuals infected with *A. xylosoxidans* bacteremia is difficult due to the fact that this pathogen has both intrinsic and acquired resistance mechanisms, typically resulting in a phenotype of multidrug resistance (MDR) (10). It is resistant to all the aminoglycosides and rifampin, as well as trimethoprim-sulfamethoxazole, ciprofloxacin, and other quinolones, with varying resistance to trimethoprim-sulfamethoxazole, ciprofloxacin, and other quinolones. Carbapenems and antipseudomonal penicillin are typically effective against most isolates (11–13). However, there is no specific treatment for this multi-disease-causing bacterium. Efforts are needed to design effective therapeutic strategies to curtail the emergence of this bacterium. Designing small molecule inhibitors is a time taking process, however, vaccination has been used to prevent and to cure a range of microbial infections, with the goal of generating adaptive immune responses by delivering antigenic components to the immune system (14). Classical vaccines, like inactivated pathogens, subunit vaccines, and live attenuated provide long-lasting protection, but they limit quick and large-scale production possibilities (15). The use of computational tools has significantly reduced the cost and time of developing peptides-based therapeutics. Computational vaccine designing has proven its effectiveness, specificity, safety, and stability compared to the conventional approaches of vaccine development (14, 16–19). Immunoinformatic strategy has been adopted for designing vaccine against a number of pathogens including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Mayaro virus, human norovirus, and Shigella spp (14, 16–18, 20).

Computational methods are of great interest in exploring biological mechanisms and designing therapeutics against the emerging pathogens (21, 22). This also involves the application of computational vaccinology approaches to design vaccine therapeutics. In this scientific study, screening of the whole proteome for vaccine targets prioritization, followed by mapping of cytotoxic T-cell Epitopes, helper T-cell epitopes, and linear B cell epitopes, extracted from the antigenic proteins of *A. xylosoxidans,* was performed. The shortlisted proteins selected from the extracellular membrane are reported to be involved in attachment, pathogenesis, replication, and disease severity. Antigenicity, immunogenicity, and allergenicity of the selected epitopes were predicted to design a potent peptide vaccine composed of potential antigenic messenger RNA (mRNA) and peptides-based multi-epitopes vaccines. The current findings will aid the development of potential peptides-based vaccine candidate. This research could pave the way for the development of a dynamic and efficient multi epitopes-based vaccine that contain a unique mix of numerous *A. xylosoxidans*-derived antigenic peptides with different roles during the *A. xylosoxidans* infection. Overall, this may also advance the therapeutics research to combat the emerging pathogen.

METHODOLOGY

Proteome Subtraction

The total proteome (UniProt ID: UP000006876) of *A. xylosoxidans* was accessed and the chromosome region (Genome accession: CP002287) coding 6,445 proteins was

downloaded from UniProtKB (https://www.uniprot.org/) (23). The whole proteome sequences were then subjected to subtractive proteomics approach for the identification of target proteins to design putative vaccine candidates against the target proteins of *A. xylosoxidans*. Firstly, each protein was screened on the basis of a subcellular localization with the utility of an online server, CELLO (http://cello.life.nctu.edu.tw/) (24). The predictions of this integrated server are based on a two-level support vector machine (SVM) system to determine the protein localization. Secondly, to exclude pathogen and human host (Homo sapiens, ID: 9606) homologous proteins; BLASTp tool with default parameters (e-value: 10-5) was utilized. This was performed to remove homologous proteins and to further process the non-homologous proteins in vaccine target prioritization (25). Furthermore, the Cluster Data Base with High Tolerance (CD-HIT) suite (http://weizhongli-lab.org/cd-hit/) (26) was utilized to exclude overlapping protein sequences in the whole proteome. This analysis based on an improved clustering algorithm was performed to remove duplicated proteins with a cut-off value of 0.8 representing 80% identity (27). Moreover, shortlisting of immunogenic target proteins was performed with the utility of two web servers, VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (28) and Algpred2 (https://webs.iiitd.edu.in/raghava/algpred2/) (29). The analysis was deployed to characterize the pathogenic proteins on the basis of antigenicity and allergenicity status. Finally, the shortlisted three target proteins were analyzed during additional investigations.

Epitopes Prioritization in *A. xylosoxidans* Proteome

The process of putative epitopes screening was initiated with the utility of NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/) server (30) to predict Cytotoxic T-lymphocytes. The method, trained on 886 known MHC class I ligands and 12 super-types, helps to predict the CTL epitopes based on TAP (Transport Associated with Antigen Processing). This was followed by prediction of HTL epitopes using the IEDB online server (http://www.iedb.org/) by using seven reference alleles set of human HLAs (31). The server predicts and characterizes the HTL epitopes based on binding affinity by depicting percentile ranks, whereas an epitope with a lower percentile rank indicates a higher binding affinity. Furthermore, to investigate the interferon-gamma producing helper T lymphocyte (HTL), epitopes were screened using an IFN- epitope (http://crdd.osdd.net/raghava/ifnepitope/) server (32). The analysis was performed with the utility of server integrated algorithms, which classifies the HTL epitopes based on SVM scores to differentiate between IFN-positive (inducing) and IFN-negative (non-inducing) epitopes (32). This was proceeded with prediction of linear B-cell epitopes with the utility of ABCPred online server (https://webs.iiitd.edu.in/raghava/abcpred/) (33). These investigations are based on comparative analysis with experimentally proved continuous B cell epitopes with higher accuracy (34). Additionally, the final shortlisting of these epitopes for inclusion in vaccine designs were preceded by antigenic and allergenic potential evaluation. The highly antigenic and non-allergenic epitopes were included in the process of *in silico* vaccine designing.

Proteome-Wide mRNA-Based Vaccine Design

The selected highly antigenic epitopes that are predicted for each target protein were then utilized in the design of an mRNA-based vaccine against *A. xylosoxidans*. Herein, 2 T-cell epitopes from each target protein were selected and a total of 6 CTL epitopes were included in the final whole proteome specific mRNA-based vaccine design (35). Similarly, a total of 6 HTL and 6 B-cell epitopes were included in the final mRNA vaccine candidate. Furthermore, an addition of 5′ m7G Cap, 5′ untranslated region (UTR), Kozak sequence, signal peptide, and linked epitopes with suitable linkers (AAY, PMGLP, and GGGGS) was consecutively performed (36). A stop codon was then added before the 3′ UTR region and was followed by the poly A Tail. The same assembly of epitopes was used in the design of multi-epitopes-based and proteome-wide peptides-based vaccine design.

Proteome-Wide Peptides-Based Vaccine Assembly and Three-Dimensional Structure Modeling

The final multi-epitopes-based putative vaccine candidate against *A. xylosoxidans* was designed based on the same assembly of the highly antigenic and non-allergenic 6 CTL, 6 HTL, and 6 B-cell epitopes. The three different classes of epitopes were added with a suitable adjuvant and were linked together by using different linkers (EAAK, sAAY, GPGPG, and KK, respectively) (16, 22, 37). This addition of linkers and adjuvant is important to retain the independent immunogenic activity of epitopes and to prevent differentiation after inclusion in the vaccine construct (38). After finalizing the assembly of the whole proteome-based vaccine construct, Robetta online server (http://robetta.bakerlab.org) (39) was utilized for modeling of the 3D structure. The server is capable of modeling multi-chain complexes by using the integrated RoseTTAFold or comparative modeling approach to model (3D) the input amino acid sequences.

Validation of the Peptides-Based Vaccine Design

To validate the accuracy of the 3D structure design, several protein structure validation servers were employed. This included the utility of proSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (40) server to predict the Z-score of the 3D model. The deviations in Z-score from the normal range are indicative of errors in the tertiary structure of modeled proteins. Additionally, ERRAT online server (https://www.doe-mbi.ucla.edu/errat/) was utilized to predict the quality factor based on the atomic bonding contacts. Finally, the available online server, PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) (41), was used to assess the stereo-chemical properties and to predict the overall structure geometry through Ramachandran Plot analysis (42, 43). Finally, the different physiochemical properties of the peptides-based vaccine construct were evaluated by using the ProtParam online tool

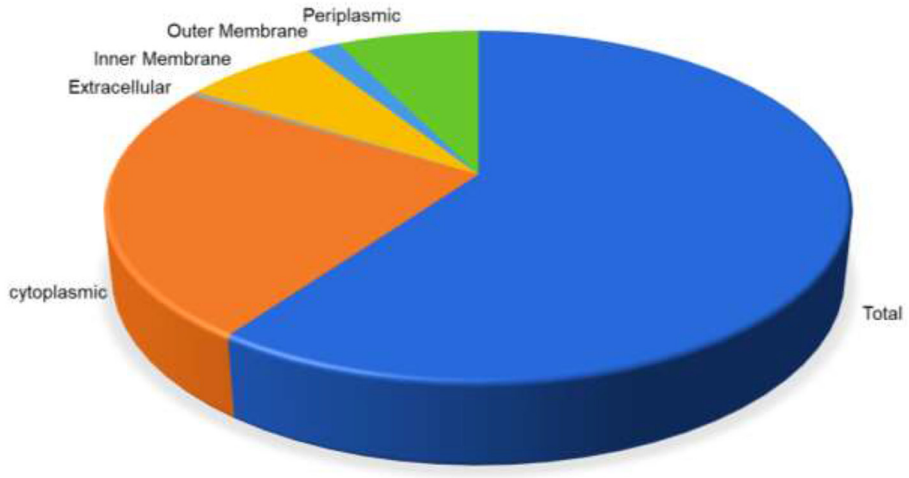


FIGURE 1 | Distribution of proteins based on the location inside the cell. Periplasmic (green) are 7.32%, outer membrane (light blue) 1.72%, inner membrane (yellow) 6.94%, extracellular (gray) 0.45%, while cytoplasmic proteins are 23.66%, respectively.

TABLE 1 | The prioritized outer-membrane target proteins with antigenicity scores of ≥0.8.

UniProt accession	Target outer membrane protein	Antigenicity scores
E3HHR6	Uncharacterized protein	0.80
E3HH04	Outer membrane transport protein OMPP1/FadL/TodX family protein	0.85
E3HWA2	Filamentous hemagglutinin family N-terminal domain protein 1	0.88

(http://web.expasy.org/protparam/) (44). This was performed to depict several topographies of the constructed vaccine including molecular weight, theoretical PI, instability index, and other related properties.

Docking and Interaction Analysis of the Vaccine

To identify the interacting patterns, molecular docking of the proposed vaccine with human TLR-5 (Toll-like receptor-5) (45) was performed using the HawkDock server (http://cadd.zju.edu.cn/hawkdock/). This server, by deploying a hybrid docking method, efficiently provides information about the binding interactions of the complex (42). Additionally, the server also offers an mm-GBSA analysis for binding free energy calculations. The structure of TLR-5, utilized in the docking complex, was retrieved from RCSB using the Accession ID: 3j0a. The structure of TLR-5 was visualized and was prepared in PyMOL before docking with the proposed vaccine.

In silico Cloning of the MEVC

The *in silico* cloning analysis were initiated with the acquisition of reverse-translated optimized DNA sequence for the peptides-based vaccine construct. This was performed with the utility of Java codon adaptation tool (JCat tool) to ensure the production of the multi-epitope subunit vaccine in an appropriate expression host (43). After selecting appropriate host (E. coli strain K-12) for the proposed vaccine, the GC content and the CAI score of the optimized DNA sequence were also determined. Furthermore, the optimized DNA sequence, after choosing two restriction enzyme sites (XhoI and EcoRI), was inserted in the pET-28a (+) expression vector to obtain the cloned plasmid by using the Snapgene software.

Immune Simulation

The validation of potential immunogenic response induced by the proposed peptides-based vaccine against the *A. xylosoxidans* was also performed. This was achieved with the utility of immune simulation approach by using C-ImmSim (http://150.146.2.1/C-IMMSIM/index.php) (46). The server is capable of depicting putative immune responses against the desired antigenic protein constructs. Several machine learning methods, with scoring matrix PSSM-based systems, are deployed to forecast antibody responses. This evaluation includes counts of antibodies, cytokines, and interferons produced against an injected antigen (47). The method is widely deployed to depict the experimental feasibility of computationally designed vaccines against the target organisms.

Molecular Dynamics Simulation

Molecular dynamics (MD) simulation of the vaccine-TLR complex was performed to check the stability of the complex using AMBER20 simulation package for 20ns (48, 49). The parameters were used as previously used by Abbas et al. (50, 51). For stability and residual flexibility estimation, CPPTRAJ and PTRAJ modules were used (52).

TABLE 2 | The shortlisted immune epitopes with antigenicity and allergenicity status evaluated for each target protein.

Target protein accession ID	Type of epitopes	Total shortlisted epitopes	Epitope sequences	Antigenicity scores	Antigenicity status	Allergenicity
E3HHR6	CTL	2	SVDVNTLVY TLGSAIWFY	1.2 1.1	Antigenic	Non-allergenic
E3HHR6	HTL	2	VGTNIIAGYMAGLRS FAWEPFVVAPVGRYD	0.5 0.7	Antigenic	Non-allergenic
E3HHR6	B-cell	2	EVEIYGRNTDWHGQTL PGMAGDHQTGIGDVTL	0.7 1.1	Antigenic	Non-allergenic
E3HH04	CTL	2	TTLDVGYTY NASGLGNAY	0.4 0.9	Antigenic	Non-allergenic
E3HH04	HTL	2	YDSNAHILGIQLSSR ALGANYKFAPNWKWK	0.9 2	Antigenic	Non-allergenic
E3HH04	B-cell	2	TGLGGSRPTGGNGGDA SWTGWSSIPNLKIRNS	2.3 0.5	Antigenic	Non-allergenic
E3HWA2	CTL	2	ETSPITPTY QSDASMTLQ	0.4 1.1	Antigenic	Non-allergenic
E3HWA2	HTL	2	AGALRSNQGNRIEAG GSAAQLQSSKGMTLS	1 1.1	Antigenic	Non-allergenic
E3HWA2	B-cell	2	PGTIEVRSDKDSGRDS GGSIVHGGTSLAGKDL	1.5 0.6	Antigenic	Non-allergenic

TABLE 3 | The full-length peptides-based MEVC design with explored physiochemical properties.

Vaccine type	Proteome wide peptides based MEVC Construct	Number of amino acids	Molecular weight (kd)	Theoretical pI	Aliphatic index	Hydropathicity (GRAVY)	Antigenicity score	Antigenicity status	Allergenicity
mRNA based vaccine	SVDVNTLVYAAYTLGSAIWFYAAYTTLDVGYTYAAYNASGLGNAYAAYETSPITPTYAAYQSDASMTLQGPGPGVGTNIIAGYMAGLRSGPGPGFAWEPFVVAPVGRYDGPGPGYDSNAHILGIQLSSRGPGPGALGANYKFAPNWKWKGPGPGAGALRSNQGNRIEAGGPGPGGSAAQLQSSKGMTLSKKEVEIYGRNTDWHGQTLKKPGMAGDHQTGIGDVTLKKTGLGGSRPTGGNGGDAKKSWTGWSSIPNLKIRNSKKPGTIEVRSDKDSGRDSKKGGSIVHGGTSLAGKDL	297	30 kd	9.54	62.53	−0.465	1	Antigen	Non-allergen
Peptides based vaccine	MRVLYLLFSFLFIFLMPLPGVFGGIGDPVTCLKSGAICHPVFCPRRYKQIGTCGLPGTKCCKKPEAAKSVDVNTLVYAAYTLGSAIWFYAAYTTLDVGYTYAAYNASGLGNAYAAYETSPITPTYAAYQSDASMTLQGPGPGVGTNIIAGYMAGLRSGPGPGFAWEPFVVAPVGRYDGPGPGYDSNAHILGIQLSSRGPGPGALGANYKFAPNWKWKGPGPGAGALRSNQGNRIEAGGPGPGGSAAQLQSSKGMTLSKKEVEIYGRNTDWHGQTLKKPGMAGDHQTGIGDVTLKKTGLGGSRPTGGNGGDAKKSWTGWSSIPNLKIRNSKKPGTIEVRSDKDSGRDSKKGGSIVHGGTSLAGKDL	365	37 kd	9.55	67.70	−0.283	1	Antigen	Non-allergen

RESULTS

Cellular Localization and Immunogenic Potential Based Protein Targets Prioritization

The whole proteome sequence of pathogenic *A. xylosoxidans* was subjected to the mining of the therapeutic targets. This was performed to shortlist the novel protein targets for computational vaccine designs. The approach is widely utilized to identify the genome-wide therapeutic targets in several diseases (53). Herein, subtractive proteomics pipeline was followed to shortlist putative vaccine targets against *A. xylosoxidans*. This was initiated with the screening of subcellular localization for the whole proteome, which contains a total of 6,445 proteins. The screening was

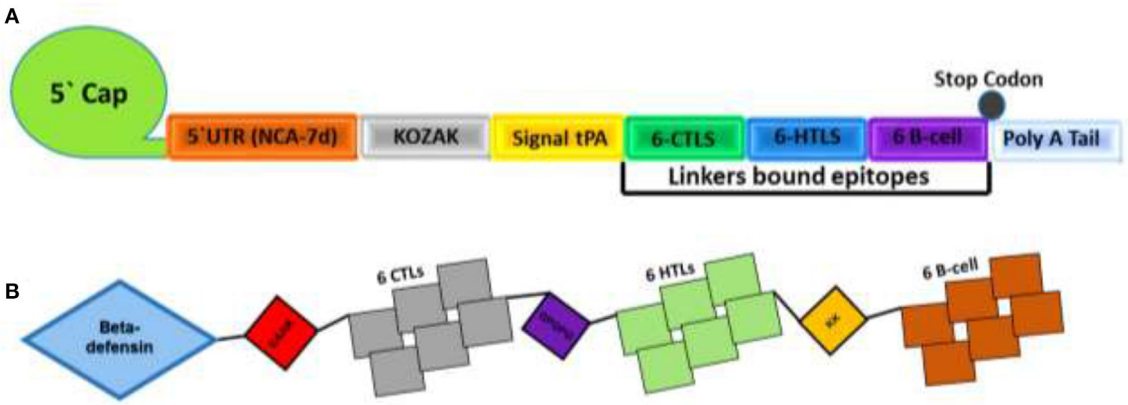


FIGURE 2 | Topological representation of the designed vaccines. **(A)** Final mRNA vaccine construct with an organized order of elements from left (N-terminal) to right (C-terminal). Each structural element, such as 5′ cap, 5′ UTR, Kozak Sequence, Signal protein, CTL epitopes, HTL epitopes, B-cell epitopes, 3′ UTR, and a Poly A tail, is represented by a different color scheme. **(B)** shows the final peptides-based vaccine construct with the adjuvant attached at N terminal and different epitopes are linked by EAAK, AAY, GPGPG and KK linkers.

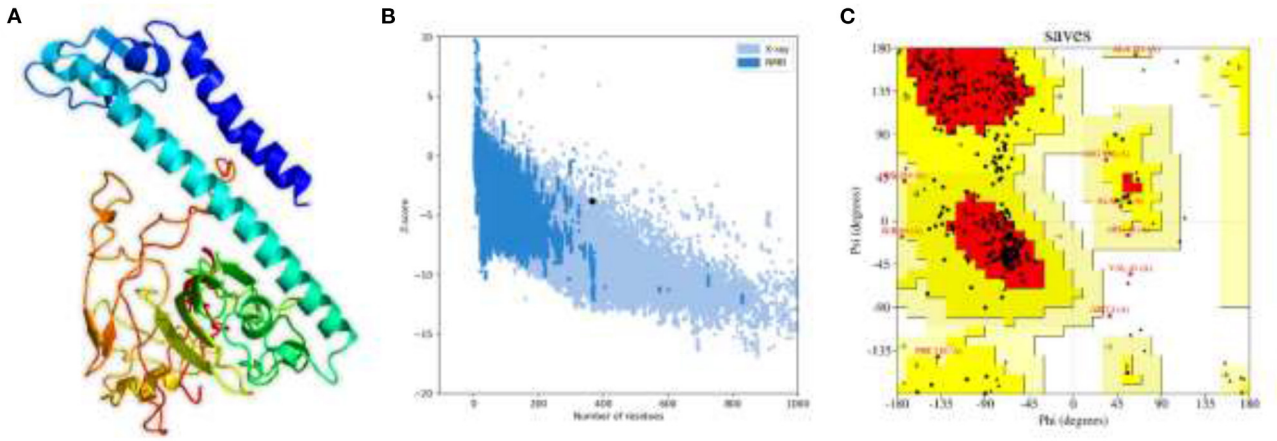


FIGURE 3 | The 3D structural model and validation of the final multi-epitopes-based vaccine design against *A. xylosoxidans*. **(A)** The Robetta generated 3D model of the vaccine construct, **(B)** the ProSA-web that generated a Z-score graph, and **(C)** the PROCHECK-generated Ramachandran-plot for validation of the structure.

performed through the online available protein localization tool "CELLO" (http://cello.life.nctu.edu.tw/). This resulted in a proteome-wide localization for each individual protein classified as periplasmic, inner membrane, and outer-membrane proteins, as shown in **Figure 1**. This helped in the shortlisting of 185 outer membrane proteins as the prioritized vaccine targets which are processed in further analysis (54, 55). These proteins were then screened for removal of human homologous proteins through using BLASTp tool, followed by paralogs screening using CD-HIT. Furthermore, the shortlisted proteins were also screened for antigenicity and allergenicity profiles, to remove non-antigenic and allergenic proteins. This characterization resulted in the selection of the highly immunogenic three target proteins with antigenicity score of >0.8. These highly antigenic targets included an outer membrane transport protein, an uncharacterized protein, and a Filamentous hemagglutinin family N-terminal domain protein, as shown in **Table 1**. Same protein targets were then analyzed to screen putative immune epitopes (T-Cell epitopes, HTL epitopes, and B-Cell epitopes) for inclusion in the final vaccine designs.

Prioritization of Highly Immunogenic Epitopes for Each Target Protein

The identification of immune epitopes through advanced computational approaches ultimately aids in designing the highly antigenic vaccine designs against the pathogenic organisms. The selection of the target proteins was followed by screening of putative immunogenic epitopes (CTL, HTL, and B-cell). The analysis was initiated with identification of CTL epitopes in each target protein sequence. These epitopes are the essential factors involved in the recognition through MHC molecules for production of immune response. Similarly, the antigenic

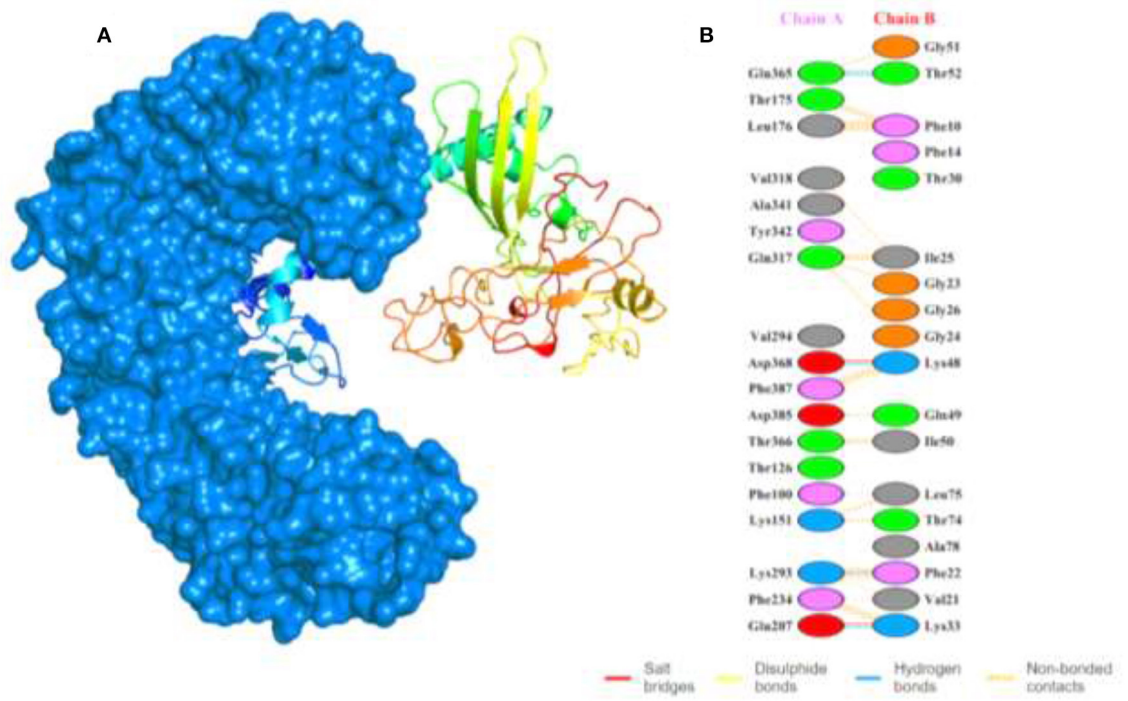


FIGURE 4 | Shows the docking complex of the vaccine structure and human TLR-5. **(A)** The docking complex with human TLR-5 structure represented in surface view (marine blue color), while the multi-colored attached structure represents the designed MEVC. **(B)** The binding patterns formed in the docking complex between the two interacting partners.

HTL and B-cell epitopes recognition by the host immune system are also vital in mediating adaptive immunity. Next, the shortlisted three proteins of *A. xylosoxidans* were also screened for identification of HTL epitopes with an IFN induction potential. This was followed by the identification of B-cell epitopes for each target protein. During the screening of T-Cell Epitopes, a total of 9, 10, and 70 MHC-I binders were identified for each target protein, respectively; whereas the number of identified HTL epitopes were 1,953, 2,982, and 1,070 while B-Cell epitopes were 31, 47, and 540, respectively, characterized for each of the three target proteins. Additionally, the epitopes selected among the identified immune epitopes was subjected to further analysis for antigenicity and allergenicity potential before the inclusion in the final vaccine designs. The final shortlisted epitopes after immunogenic potential evaluation based on high antigenicity scores are given in **Table 2.** All the shortlisted epitopes (CTL, HTL, and B cell) from **Table 2** were included in the MEVC designs against *A. xylosoxidans*.

Designing mRNA- and Peptides-Based MEVC

The proteome-wide identification of putative targets and the mapping of vital epitopes was followed by mRNA and by peptides-based vaccine designing. The highly antigenic epitopes, joined together through different linkers, were then used in the full-length mRNA- and peptides-based MEVC design shown in **Table 3**. The mRNA-based vaccine against *A. xylosoxidans* was designed by targeting the three highly antigenic target proteins. The assembly of final mRNA vaccine construct comprised of a total of 18 immune epitopes (6 CTL epitopes, 6 HTL epitopes, and 6 linear B-cell epitopes). This was initiated with the N terminal 5'm7G cap followed by NCA-7d (5' UTR), Kozak sequence, and a signal peptide tPA (tissue Plasminogen Activator). The assembly of mRNA vaccine also includes the addition of a stop codon, S27a+R3U (3' UTR) sequence, and a Poly (A) tail (120 nucleotides long). The graphical representation and arrangement of the mRNA vaccine design is shown in **Figure 2A**. The peptides-based whole proteome-wide vaccine construct against *A. xylosoxidans* of length 365 amino acids was also evaluated for different physiochemical properties, along with antigenic and allergenic potential to ensure its application in further experimental designs (**Figure 2B**).

Structural Modeling and Validation of MEVC

Structural vaccinology approaches, by using computationally predicted epitopes, have been already deployed in several experimental models (56, 57). Here, in the whole proteome-wide, final peptides-based MEVC construct of length 365 amino acids, involving 18 immune epitopes joined through linkers and N-terminal adjuvant, was constructed. Next, the 3D structure of the final vaccine construct was generated with the utility of Robetta server as shown in **Figure 3A**. The best 3D predicted model was visualized using PyMOL software and was then validated through two different servers to finalize the most accurate model.

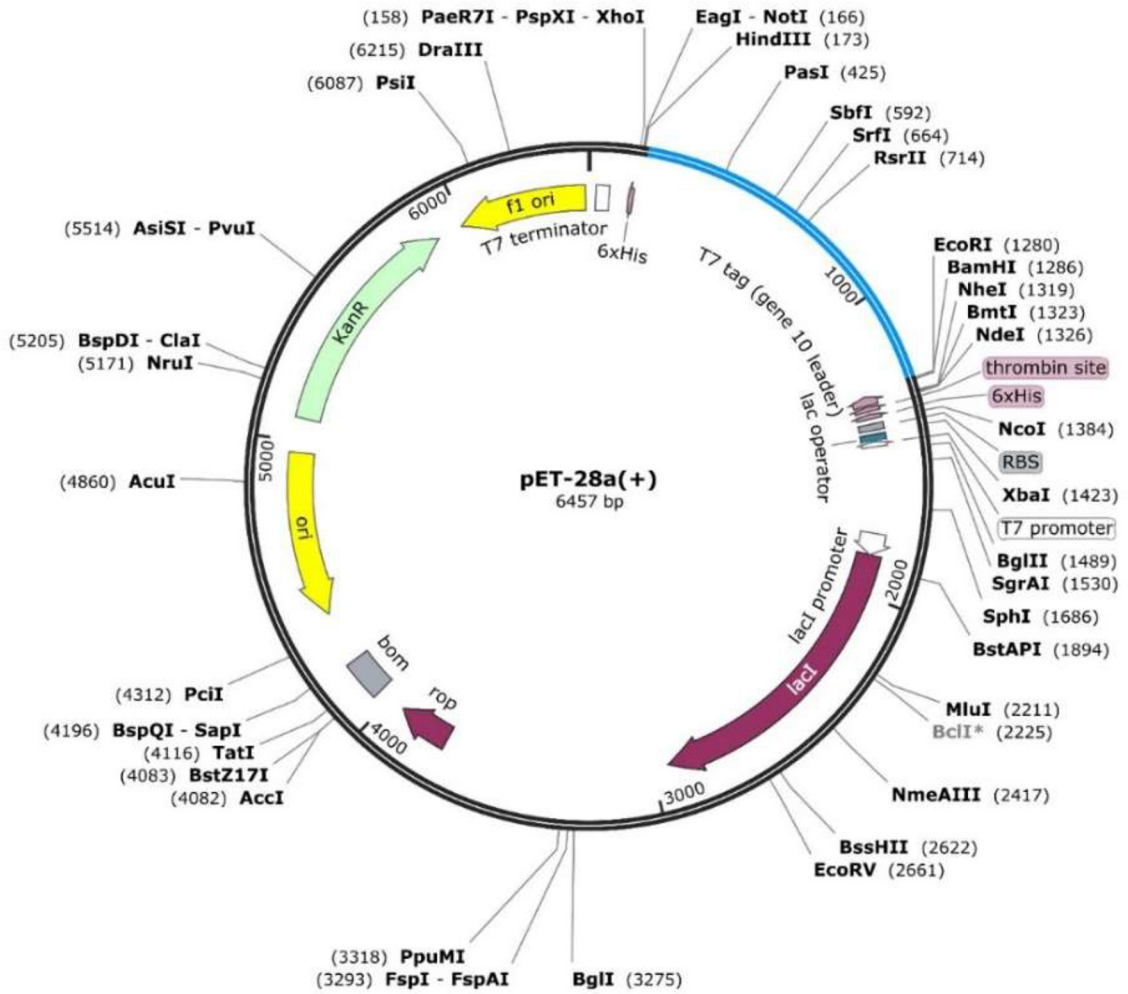


FIGURE 5 | The designed plasmid map for the proposed vaccine construct by using restriction enzymes (XhoI and ECOR1) in the pet28a (+) vector.

This model validation was performed with the use of ProSA-web and PROCHECK servers. The ProSA-web analysis revealed a Z-score of−3.8 (**Figure 3B**) and the ERRAT predicted high quality scores that indicated the validity of the predicted 3D structure. Moreover, PROCHECK resulted in the Ramachandran-plot that showed 72.5% of the residues in the most favored regions and only 0.7% in the dis-allowed region (**Figure 3C**).

Docking Complex Analysis of MEVC-TLR5 and Interaction Analysis

To evaluate the MEVC-TLR5 interactions, we utilized the HawkDock server to generate several docking models with docking complex scores. The best docking complex (**Figure 4A**) with lowest acquired energy scores (−5673.28) were then subjected to interaction analysis by utilizing PDBsum. This was performed to calculate the number of different binding interactions including hydrogen bonds, salt bridges, and non-bonding contacts in the docking complex as shown in **Figure 4B.** The interaction pattern, followed in the docking complex, included formation of 5 hydrogen bonds, 2 salt bridges, and 165 non-bonded contacts. Furthermore, the MM-GBSA analysis was also performed to reveal the total and individual binding free energies involved in the regulation of the formation of the docking complex. The energies are as follows: Van der Waals energy (−103.99 kcal/mol), Electrostatic energy (−1833.88 kcal/mol), Gibbs free energy (1915.62 kcal/mol), Surface Area (−12.39 kcal/mol), and the total binding energy −34.64 kcal/mol, which shows that robust binding of the vaccine to the TLR-5.

In silico Cloning Design of WP-MEVC

The constructed peptides-based vaccine design was also subjected to the acquisition of an improved DNA sequence for the expression in the *E. coli* strain K-12 with the utility of Jcat server. The optimized nucleotide sequences with a calculated higher CAI score of 0.98 and an optimal percentage of GC contents (54.4%) indicated a putatively higher expression of the vaccine construct in *E. coli*. This was followed by the selection of the restriction enzymes (XhoI and ECORI) used in the cloning of the optimized

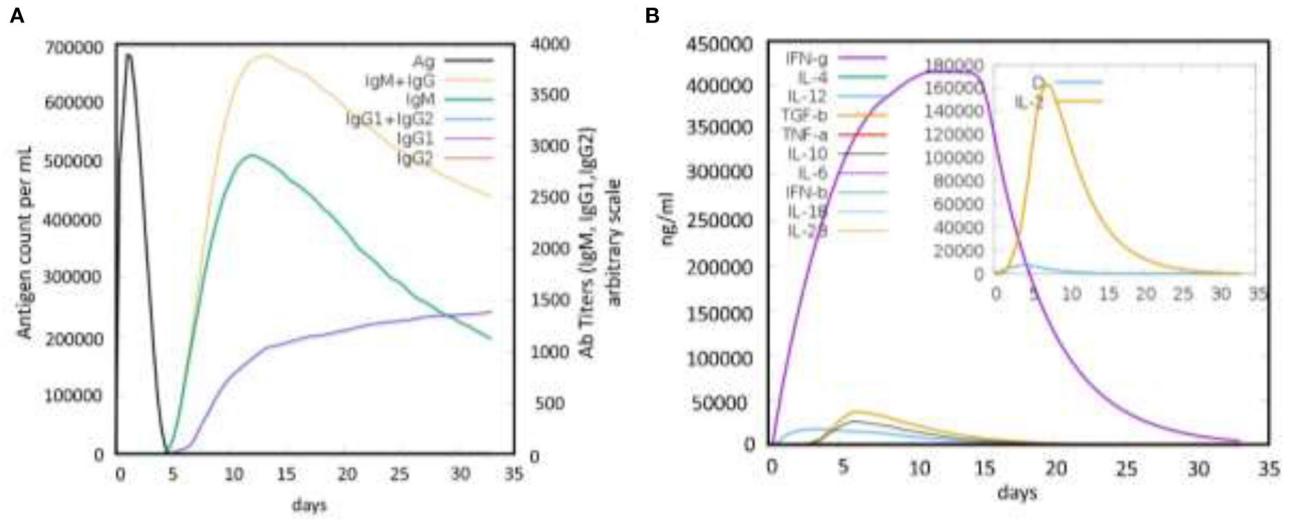


FIGURE 6 | The immune simulation graphs showing immune response produced against the antigenic peptides-based vaccine design. **(A)** The graph showing antigen count/ml/day plotted against Ab titers. **(B)** The production of cytokines and interleukins against the whole proteome-wide peptides-based vaccine design.

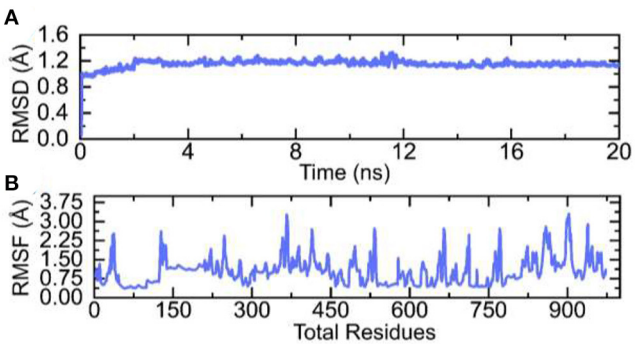


FIGURE 7 | RMSD and RMSF graphs of the vaccine-TLR complex over the simulation time. **(A)** RMSD and **(B)** RMSF of the complex.

DNA sequence in pET28a (+) vector. The insertion of the desired vaccine sequence, followed by designing of pET28a (+) plasmid, was performed in Snapgene (**Figure 5**).

Immune Simulation of the Proposed Vaccine

Finally, the immune response induction potential of the designed vaccine was also evaluated for the highly antigenic MEVC construct during the *in silico* immune simulation. The analysis revealed a strong induction of antibodies production against the vaccine construct as an antigen. This analysis showed a higher antigen counts of >600,000 between days 1-2 and were completely neutralized until the 5th day. This was followed with an increasing trend of antibodies produced against the antigenic vaccine, achieving the highest peaks (>3,500) of IgM+ IgG production between 10 and 15 days (**Figure 6A**). Similarly, higher immunoglobulins (IgM, IgG1, and IgG2), specific antibody titers, were also observed during the same time period. These results reflected the putative immunogenic potential of the designed MEVC to trigger an enhanced immune response against *A. xylosoxidans*. Additionally, the peptides-based vaccine design was also evaluated for production of different Interleukins (IL), Interferon (IFN), transforming growth factors (TGF), and transforming necrosis factor (TNF). The highest among these factors, produced against the antigenic vaccine, was IFN-g with the highest peak (>400,000) observed during 10 to 15 days. The concentrations of different cytokines and interleukins produced are presented in **Figure 6B**. However, an advanced experimental analysis of the designed vaccine candidate and its potential to produce adaptive immunity against human pathogenic *A. xylosoxidans* may be clarified through further *in vitro* and *in vivo* demonstrations.

MD Simulation

Molecular dynamics (MD) simulation of the vaccine-TLR complex was performed to check the stability and flexibility of the complex. The stability of the complex was calculated as room mean square deviation (RMSD), which revealed the stable behavior of the complex over the simulation time. The RMSD graph of the complex is given in **Figure 7A**. On the hand the residual flexibility was calculated as room mean square fluctuation (RMSF), which revealed different flexibility index of each residue (**Figure 7B**). Overall, the flexibility level is well acceptable and demonstrates the favorable dynamic behavior.

CONCLUSIONS

In conclusion, the whole proteome of *A. xylosoxidans* was studied to shortlist the three putative vaccine targets. This was followed by the prediction of different immune epitopes (CTL, HTL, and B-cell) for each of the target protein. The highly immunogenic peptides were then topographically arranged to design a proteome-wide mRNA and MEVC. The final vaccine construct was then subjected to a 3D modeling and

structural validation. The evaluations of the vaccine structure and exploration of physiochemical features also reflected the ability of a potential immunization. Moreover, strong interaction patterns (more hydrogen bonds), through molecular docking and experimental feasibility, were confirmed through *in silico* cloning and immune simulations. The study provides new insights into target specific immune epitopes identification, and its utility in mRNA and peptides-based vaccine designs. This also suggests further experimental validation of the proposed vaccine designs. However, the experimental processing may face challenges including protein expression solubility during synthesis. Still, the vaccine designs may serve as pre-validated therapeutic option with a potential utility against *A. xylosoxidans*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

AUTHOR CONTRIBUTIONS

TK, FA, TT, MS, and AK: conceptualization. MA: data curation. TK, FA, MS, and AK: formal analysis. TK, FA, TT, and AI: investigation. MA and TT: methodology. AI and D-QW: resources. MA, MS, AI, and LA: software. LA and D-QW: supervision. MS, LA, AK, and YW: validation. LA: visualization. AK, YW, and D-QW: writing—original draft. YW and D-QW: writing—review and editing.

FUNDING

FA is funded from Researchers Supporting Project of King Saud University, Riyadh, Saudi Arabia (No. RSP-2021/198).

REFERENCES

1. Duggan JM, Goldstein SJ, Chenoweth CE, Kauffman CA, Bradley SF. Achromobacter xylosoxidans bacteremia: report of four cases and review of the literature. *Clin Infect Dis.* (1996) 23:569-76. doi: 10.1093/clinids/23.3.569
2. Mandell WF, Garvey GJ, Neu HC. Achromobacter xylosoxidans bacteremia. *Rev Infect Dis.* (1987) 9:1001-5. doi: 10.1093/clinids/9.5.1001
3. Marsac C, Berdah L, Thouvenin G, Sermet-Gaudelus I, Corvol H. *Achromobacter xylosoxidans* airway infection is associated with lung disease severity in children with cystic fibrosis. *ERJ Open Res.* (2021) 7:00076-2021. doi: 10.1183/23120541.00076-2021
4. Firmida MC, Marques EA, Leão RS, Pereira RHV, Rodrigues ERA, Albano RM, et al. *Achromobacter xylosoxidans* infection in cystic fibrosis siblings with different outcomes. *Respir Med Case Rep.* (2017) 20:98-103. doi: 10.1016/j.rmcr.2017.01.005
5. Menetrey Q, Sorlin P, Jumas-Bilak E, Chiron R, Dupont C, Marchandin H. *Achromobacter xylosoxidans* and *Stenotrophomonas maltophilia*: emerging pathogens well-armed for life in the cystic fibrosis patients' lung. *Genes.* (2021) 12:610. doi: 10.3390/genes12050610
6. Barragán EP, Pérez JS, Corbella L, Orellana MÁ, Fernández-Ruiz M. *Achromobacter xylosoxidans* bacteremia: clinical and microbiological features in a 10-year case series. *Revista Española de Quimioterapia.* (2018) 31:268.
7. Claassen SL, Reese JM, Mysliwiec V, Mahlen SD. *Achromobacter xylosoxidans* infection presenting as a pulmonary nodule mimicking cancer. *J Clin Microbiol.* (2011) 49:2751-4. doi: 10.1128/JCM.02571-10
8. Sebanayagam V, Nguyen P, Nassar MA, Soubani A. Nosocomial *Achromobacter xylosoxidans* infection presenting as a cavitary lung lesion in a lung cancer patient. *Cureus.* (2020) 12:e9818. doi: 10.7759/cureus.9818
9. Aisenberg G, Rolston KV, Safdar A. Bacteremia caused by Achromobacter and Alcaligenes species in 46 patients with cancer 1989–2003. *Cancer.* (2004) 101:2134-40. doi: 10.1002/cncr.20604
10. Yamamoto M, Nagao M, Hotta G, Matsumura Y, Matsushima A, Ito Y, et al. Molecular characterization of IMP-type metallo-β-lactamases among multidrug-resistant *Achromobacter xylosoxidans*. *J Antimicrob Chemother.* (2012) 67:2110-3. doi: 10.1093/jac/dks179
11. Lobo LJ, Tulu Z, Aris RM, Noone PG. Pan-resistant *Achromobacter xylosoxidans* and *Stenotrophomonas maltophilia* infection in cystic fibrosis does not reduce survival after lung transplantation. *Transplantation.* (2015) 99:2196-202. doi: 10.1097/TP.0000000000000709
12. Mathy V, Grohs P, Compain F. *In vitro* activity of β-lactams in combination with avibactam against multidrug-resistant *Pseudomonas aeruginosa*, *Stenotrophomonas maltophilia*, and *Achromobacter xylosoxidans* isolates from patients with cystic fibrosis. *J Med Microbiol.* (2018) 67:1217-20. doi: 10.1099/jmm.0.000801
13. Ulloa ER, Kousha A, Tsunemoto H, Pogliano J, Licitra C, LiPuma JJ, et al. Azithromycin exerts bactericidal activity and enhances innate immune mediated killing of MDR *Achromobacter xylosoxidans*. *Infect Microb Dis.* (2020) 2:10-7. doi: 10.1097/IM9.0000000000000014
14. Ali A, Khan A, Kaushik AC, Wang Y, Ali SS, Junaid M, et al. Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). *Sci Rep.* (2019) 9:1-12. doi: 10.1038/s41598-018-37070-z
15. Plotkin SA. Vaccines: the fourth century. *Clin Vaccine Immunol.* (2009) 16:1709-19. doi: 10.1128/CVI.00290-09
16. Khan A, Khan S, Saleem S, Nizam-Uddin N, Mohammad A, Khan T, et al. Immunogenomics guided design of immunomodulatory multi-epitope subunit vaccine against the SARS-CoV-2 new variants, and its validation through in silico cloning and immune simulation. *Comput Biol Med.* (2021) 133:104420. doi: 10.1016/j.compbiomed.2021.104420
17. Khan S, Khan A, Rehman AU, Ahmad I, Ullah S, Khan AA, et al. Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression. *Infect Genet Evolut.* (2019) 73:390-400. doi: 10.1016/j.meegid.2019.06.006
18. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, et al. Immunoinformatics approaches to explore Helicobacter Pylori proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. *Sci Rep.* (2019) 9:1-13. doi: 10.1038/s41598-019-49354-z
19. Khan A, Khan T, Ali S, Aftab S, Wang Y, Qiankun W, et al. SARS-CoV-2 new variants: Characteristic features and impact on the efficacy of different vaccines. *Biomed Pharmacother.* (2021) 143:112176. doi: 10.1016/j.biopha.2021.112176
20. Khan A, Junaid M, Kaushik AC, Ali A, Ali SS, Mehmood A, Wei D-Q. Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PLoS ONE.* (2018) 13:e0196484. doi: 10.1371/journal.pone.0196484
21. Khan T, Khan A, Ali SS, Ali S, Wei D-Q. A computational perspective on the dynamic behaviour of recurrent drug resistance mutations in the pncA gene from *Mycobacterium tuberculosis*. *RSC Adv.* (2021) 11:2476-86. doi: 10.1039/D0RA09326B
22. Khan T, Khan A, Wei D-Q. MMV-db: vaccinomics and RNA-based therapeutics database for infectious hemorrhagic fever-causing mammarenaviruses. *Database.* (2021) 2021. doi: 10.1093/database/baab063

23. Consortium U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* (2019) 47:D506-15. doi: 10.1093/nar/gky1049
24. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins Struct Function Bioinform.* (2006) 64:643-51. doi: 10.1002/prot.21018
25. Mahram A, Herbordt MC. NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Transact Reconfig Technol Syst.* (2015) 7:1-20. doi: 10.1145/2629691
26. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* (2012) 28:3150-2. doi: 10.1093/bioinformatics/bts565
27. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* (2006) 22:1658-9. doi: 10.1093/bioinformatics/btl158
28. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* (2007) 8:1-7. doi: 10.1186/1471-2105-8-4
29. Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GP. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform.* (2020) 22:1–12. doi: 10.1093/bib/bbaa294
30. Stranzl T, Larsen MV, Lundegaard C, Nielsen, M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics.* (2010) 62:357-68. doi: 10.1007/s00251-010-0441-4
31. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, et al. Immune epitope database analysis resource. *Nucl Acids Res.* (2012) 40:W525-30. doi: 10.1093/nar/gks438
32. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct.* (2013) 8:1-15. doi: 10.1186/1745-6150-8-30
33. Saha S, Raghava GPS. Prediction methods for B-cell epitopes. In: *Immunoinformatics*. Totowa, NJ: Springer (2007). p. 387-94.
34. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct Funct Bioinform.* (2006) 65:40-8. doi: 10.1002/prot.21078
35. Cai X, Li JJ, Liu T, Brian O, Li J. Infectious disease mRNA vaccines and a review on epitope prediction for vaccine design. *Brief Funct Genom.* (2021) 20:289–303. doi: 10.1093/bfgp/elab027
36. Kanekiyo M, Ellis D, King NP. New vaccine design and delivery technologies. *J Infect Dis.* (2019) 219:S88-96. doi: 10.1093/infdis/jiy745
37. Khan T, Khan A, Nasir SN, Ahmad S, Ali SS, Wei D-Q. CytomegaloVirusDb: multi-Omics knowledge database for cytomegaloviruses. *Comput Biol Med.* (2021) 135:104563. doi: 10.1016/j.compbiomed.2021.104563
38. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discov Today.* (2020) 25:1034-42. doi: 10.1016/j.drudis.2020.03.006
39. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucl Acids Res.* (2004) 32:W526-31. doi: 10.1093/nar/gkh468
40. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucl Acids Res.* (2007) 35:W407-10. doi: 10.1093/nar/gkm290
41. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein-structure coordinates. *Int Table Crystallogr.* (2012) F:684–7. doi: 10.1107/97809553602060000882
42. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, et al. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucl Acids Res.* (2019) 47:W322-30. doi: 10.1093/nar/gkz397
43. 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. *Nucl Acids Res.* (2005) 33:W526-31. doi: 10.1093/nar/gki376
44. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: *Proteom Protocols Handbook.* Totowa, NJ (2005). p. 571-607.
45. Invernizzi R, Lloyd CM, Molyneaux PL. Respiratory microbiome and epithelial interactions shape immunity in the lungs. *Immunology.* (2020) 160:171-82. doi: 10.1111/imm.13195
46. Rapin N, Lund O, Castiglione F. Immune system simulation online. *Bioinformatics.* (2011) 27:2013-14. doi: 10.1093/bioinformatics/btr335
47. Delves PJ, Roitt IM. The immune system. *N Engl J Med.* (2000) 343:37-49. doi: 10.1056/NEJM200007063430107
48. Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem.* (2005) 26:1668-88. doi: 10.1002/jcc.20290
49. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun.* (1995) 91:1-41. doi: 10.1016/0010-4655(95)00041-D
50. Khan A, Ashfaq Ur R, Junaid M, Li C-D, Saleem S, Humayun F, et al. Dynamics insights into the gain of flexibility by helix-12 in ESR1 as a mechanism of resistance to drugs in breast cancer cell lines. *Front Mol Biosci.* (2020) 6:159. doi: 10.3389/fmolb.2019.00159
51. Khan A, Umbreen S, Hameed A, Fatima R, Zahoor U, Babar Z, et al. In silico mutagenesis-based remodelling of SARS-CoV-1 Peptide (ATLQAIAS) to Inhibit SARS-CoV-2: structural-dynamics and free energy calculations. *Interdisciplin Sci Comput Life Sci.* (2021) 1-14. doi: 10.1007/s12539-021-00447-2
52. Roe DR. Cheatham III TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput.* (2013) 9:3084-95. doi: 10.1021/ct400341p
53. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Sci Rep.* (2018) 8:1-19. doi: 10.1038/s41598-018-26689-7
54. Hara Y, Mohamed R, Nathan S. Immunogenic Burkholderia pseudomallei outer membrane proteins as potential candidate vaccine targets. *PLoS ONE.* (2009) 4:e6496. doi: 10.1371/journal.pone.0006496
55. Leow CY, Kazi A, Ismail CMKH, Chuah C, Lim BH, Leow CH, et al. Reverse vaccinology approach for the identification and characterization of outer membrane proteins of Shigella flexneri as potential cellular-and antibody-dependent vaccine candidates. *Clin Exp Vaccine Res.* (2020) 9:15-25. doi: 10.7774/cevr.2020.9.1.15
56. Ren S, Guan L, Dong Y, Wang C, Feng L, Xie Y. Design and evaluation of a multi-epitope assembly peptide vaccine against Acinetobacter baumannii infection in mice. *Swiss Med Weekly.* (2019) 149:w20052. doi: 10.4414/smw.2019.20052
57. Guedes RLM, Rodrigues CMF, Coatnoan N, Cosson A, Cadioli FA, Garcia HA, et al. A comparative in silico linear B-cell epitope prediction and characterization for South American and African Trypanosoma vivax strains. *Genomics.* (2019) 111:407-17. doi: 10.1016/j.ygeno.2018.02.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MMSB declared a shared affiliation with one of the author, FA, to the handling editor at the time of review.

frontiers in Medicine

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fmed.2022.829467

In silico Designing of an Epitope-Based Vaccine Against Common *E. coli* Pathotypes

Mohamed A. Soltan[1], Mohammed Y. Behairy[2], Mennatallah S. Abdelkader[3], Sarah Albogami[4], Eman Fayad[4], Refaat A. Eid[5], Khaled M. Darwish[6], Sameh S. Elhady[7], Ahmed M. Lotfy[8] and Muhammad Alaa Eldeen[9]**

[1] Department of Microbiology and Immunology, Faculty of Pharmacy, Sinai University, Ismailia, Egypt, [2] Department of Microbiology and Immunology, Faculty of Pharmacy, University of Sadat City, Sadat City, Egypt, [3] Department of Microbiology and Immunology, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt, [4] Department of Biotechnology, College of Science, Taif University, Taif, Saudi Arabia, [5] Department of Pathology, College of Medicine, King Khalid University, Abha, Saudi Arabia, [6] Department of Medicinal Chemistry, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt, [7] Department of Natural Products, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, [8] Department of Pharmacology, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Egypt, [9] Division of Cell Biology, Histology and Genetics, Department of Zoology, Faculty of Science, Zagazig University, Zagazig, Egypt

Escherichia coli (*E. coli*) is a Gram-negative bacterium that belongs to the family Enterobacteriaceae. While *E. coli* can stay as an innocuous resident in the digestive tract, it can cause a group of symptoms ranging from diarrhea to live threatening complications. Due to the increased rate of antibiotic resistance worldwide, the development of an effective vaccine against *E. coli* pathotypes is a major health priority. In this study, a reverse vaccinology approach along with immunoinformatics has been applied for the detection of potential antigens to develop an effective vaccine. Based on our screening of 5,155 proteins, we identified lipopolysaccharide assembly protein (LptD) and outer membrane protein assembly factor (BamA) as vaccine candidates for the current study. The conservancy of these proteins in the main *E. coli* pathotypes was assessed through BLASTp to make sure that the designed vaccine will be protective against major *E. coli* pathotypes. The multitope vaccine was constructed using cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and B cell lymphocyte (BCL) epitopes with suitable linkers and adjuvant. Following that, it was analyzed computationally where it was found to be antigenic, soluble, stable, and non-allergen. Additionally, the adopted docking study, as well as all-atom molecular dynamics simulation, illustrated the promising predicted affinity and free binding energy of this constructed vaccine against the human Toll-like receptor-4 (*h*TLR-4) dimeric state. In this regard, wet lab studies are required to prove the efficacy of the potential vaccine construct that demonstrated promising results through computational validation.

Keywords: *Escherichia coli*, reverse vaccinology, immunoinformatics, epitope mapping, multitope vaccine

OPEN ACCESS

Edited by:
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

Reviewed by:
Ahmed Rakib,
University of Tennessee Health Science Center (UTHSC), United States
Angus Oli,
Nnamdi Azikiwe University, Nigeria
Paul Bogdan,
University of Southern California, United States

****Correspondence:***
Mohamed A. Soltan
mohamed.mohamed@su.edu.eg
Muhammad Alaa Eldeen
dr.muhammadalaa@gmail.com

Specialty section:
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *05 December 2021*
Accepted: *21 January 2022*
Published: *04 March 2022*

Citation:
Soltan MA, Behairy MY, Abdelkader MS, Albogami S, Fayad E, Eid RA, Darwish KM, Elhady SS, Lotfy AM and Alaa Eldeen M (2022) In silico Designing of an Epitope-Based Vaccine Against Common E. coli Pathotypes. Front. Med. 9:829467. doi: 10.3389/fmed.2022.829467

INTRODUCTION

According to its pathogenicity in humans, *E. coli* can be classified into an extraintestinal infection-causing *E. coli* (primarily uropathogenic *E. coli*, UPEC, and neonatal meningitis *E. coli*, NMEC) and strains that cause enteric infections (divided into 6 pathotypes; enteropathogenic *E. coli* [EPEC], enterohemorrhagic *E. coli* [EHEC], enteroaggregative *E. coli* [EAEC], enteroinvasive *E. coli* [EIEC], enterotoxigenic *E. coli* [ETEC] and diffusely adherent *E. coli* [DAEC]). Additionally, two further pathotypes have emerged recently; adherent invasive *E. coli* [AIEC] that is predicted to be associated with Crohn's disease and does not produce any toxins, while the other pathotype, Shiga toxin-producing enteroaggregative *E. coli* [STEAEC], causes food poisoning due to the action of Shiga toxin and was reported to be the causative agent for the 2011 Germany *E. coli* outbreak (1). People can be easily infected with *E. coli* by swallowing a small amount of it in water, vegetables, or meat where the fecal-oral route is the major way of transmission (2). Another common way of categorizing *E. coli* is by serotype through detection of surface antigens O and H. Currently, more than 190 serogroups have been identified (3).

Currently, pathogenic *E. coli* is a major public health concern because of possessing a low infectious dose and simple transmission through food and water (4). Excessive usage of different classes of antibiotics, especially in cases that do not require this form of treatment, contributed largely to the appearance of antibiotic resistance in bacterial strains that were previously sensitive even for the traditional classes of antibiotics (5). Antibiotic resistance of *E. coli* is extensively studied for two major reasons; firstly, it represents the most common infective Gram-negative pathogen for humans (6), secondly, resistant *E. coli* can affect other bacteria in the gastrointestinal tract through transferring antibiotics resistance determinants (7). The resistance rates are demonstrating continuous inclement throughout the last few years (8). An alternative solution to fight against this pathogen is designing an effective vaccine against its common pathotypes. While this solution was adopted by many research groups around the world, there is no officially approved vaccine against pathogenic *E. coli* in the market (9).

Many trials have been performed to generate an effective vaccine for various *E. coli* pathotypes. Regarding intestinal *E. coli*, Shiga toxin-based vaccine (10), attenuated bacteria-based vaccine (11), and polysaccharide-based vaccine (12) have been proposed for EHEC while autotransporter-based vaccine (13) and adhesion based vaccine (14) were proposed for ETEC. Moving to extraintestinal *E. coli*, capsular-based vaccine (15) and iron scavenger receptors-based vaccine (16) were designed. The previous approaches for designing an effective vaccine are based mainly on studying the virulence properties for each pathotype of *E. coli* and as we mentioned above none of these vaccines have been approved by FDA yet. In the current study, we planned to merge two advanced approaches, reverse vaccinology and immunoinformatics for a vaccine design where these approaches would be applied on conserved protein candidates in various *E. coli* pathotypes to generate a vaccine with possible cross-reactivity.

Reverse vaccinology is a novel approach for vaccine designing that demonstrated a great development in the last few years as a result of the revolution in sequencing techniques and the data availability in highly organized databases. Additionally, computational tools that can deal with this large amount of data and translate it to valuable outcomes that can be used in applied research also showed great progress (17). This approach for vaccine design has been utilized for a protein-based vaccine design against viral, bacterial, and fungal infectious agents (18–20). Moreover, it was also applied to design an effective vaccine for animals against ETEC, a common pathotype of *E. coli* that leads to post-weaning diarrhea in pigs (21). Recent studies showed that multitope vaccines had superiority over single-component ones regarding their effectiveness (22). In the current study, we aim to investigate the whole proteome of *E. coli* and select potential protein candidates through the reverse vaccinology approach. Following that, immunogenic epitopes from these candidates would be mapped and gathered to initiate a multitope vaccine that would be assessed computationally for its, physicochemical, chemical, and immunological characteristics to be nominated as a potential vaccine candidate against pathogenic *E. coli*.

MATERIALS AND METHODS

We divided the current study into two main stages (**Figure 1**) wherein the first stage we handled the proteome of *E. coli* with successive filtration steps of reverse vaccinology approach to nominate protein candidates for vaccine design while in the second stage we constructed, analyzed, and assessed the multiepitope vaccine that was created based on B and T cell epitopes of protein candidates from the first stage.

Data Retrieval

The amino acid sequences of the whole proteome of *E. coli* O157:H7 str. Sakai was retrieved with Genbank assembly accession No: GCF_000008865.2. This strain was selected as it represents the reference pathogenic *E. coli* strain on NCBI.

Prediction of Essential Exoproteome and Secretome

Living organisms cannot survive without their essential genes (23). Therefore, the first filtration step was detecting these genes using Geptop 2.0, where a list of essential proteins was detected out of 5,155 proteins (the complete proteome of the studied strain) (24). The default essentiality score cutoff (0.24) was used for this process. Following that, PSORTb v3.0.2 online server (25) was employed to predict the subcellular localization of the filtered genes. The overall result of this filtration phase was a list of exoproteome and secretome essential proteins.

Prediction of Virulence Protein

Essential exoproteome and secretome were applied to virulenpred (26) to estimate the virulence potential.

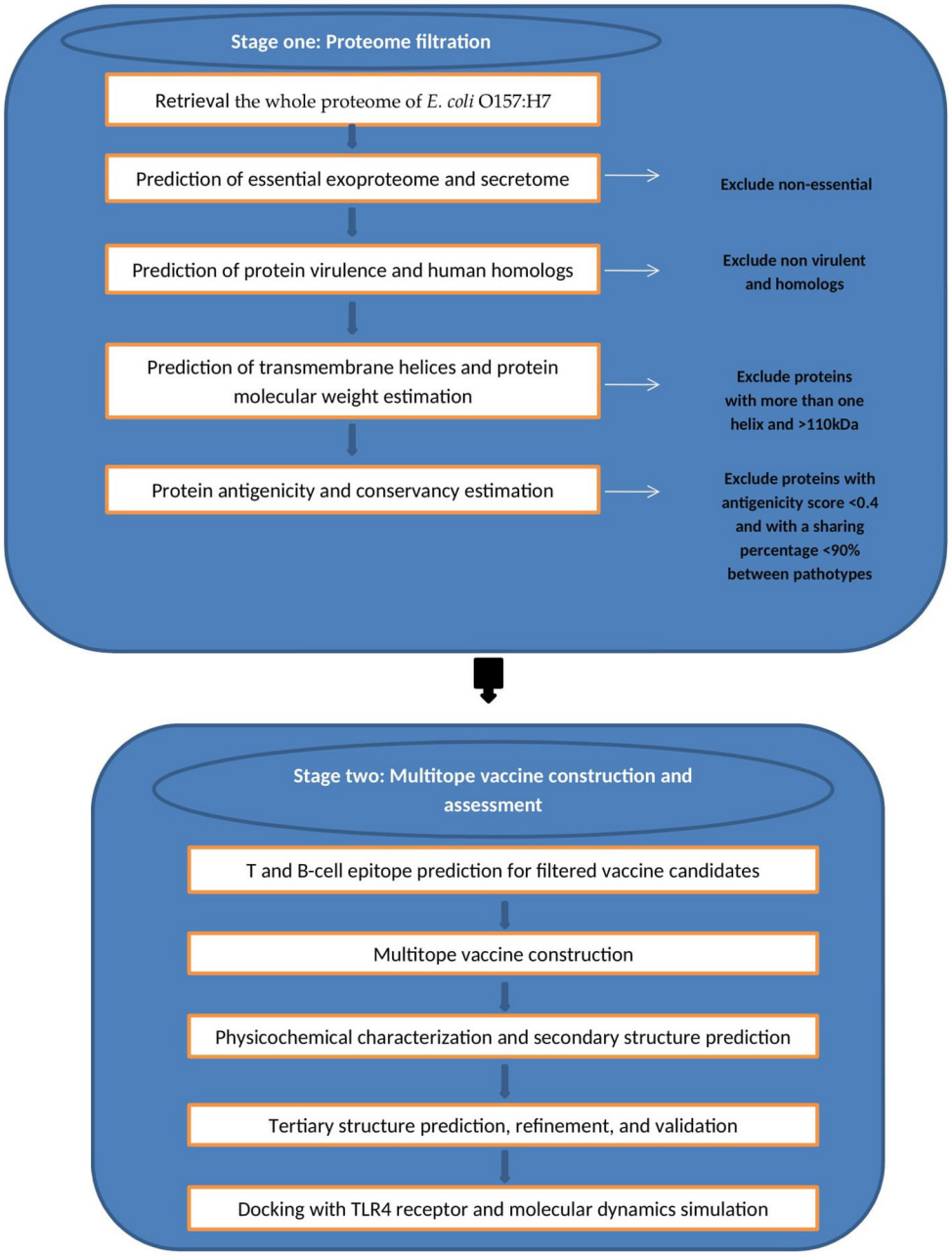


FIGURE 1 | Graphical representation of the methodology employed for the multi-epitope vaccine design against *E. coli*.

Determination of Human Homologs

Protein list from the above stage was analyzed against human proteome on NCBI using BLASTp to check for human homologs, and those with ≥35% identity were removed from our list.

Prediction of Transmembrane Helices and Protein Molecular Weight Estimation

The presence of transmembrane helices and proteins' molecular weight estimation were detected through TMHMM (27), and Expasy tool (28), respectively. Those with molecular weight <110 kDa and had ≤1 transmembrane helix were selected for antigenicity assessment in the next step (29).

Protein Antigenicity Analysis

Protein Antigenicity was detected by Vaxijen V 2.0 (30), where proteins with antigenicity score more than 0.4 were considered antigenic.

Analysis of Protein Conserved Identity With Various *E. coli* Pathotypes

Recommended proteins from previous steps were assessed for their conservation among 10 *E. coli* strains where BLASTp analysis was performed for each protein against a representative strain of each *E. coli* pathotype to confirm that the designed vaccine will have the ability to act against any pathogenic *E. coli* which is a major aim of this study. Proteins with a sharing percentage <90% were eliminated and considered not conserved.

T-Cell and B-Cell Epitope Prediction

After completing the whole reverse vaccinology filtration steps, successful candidates were transferred to the epitope mapping stage where the filtered proteins were submitted to SignalP-5.0 Server to anticipate the location of signal peptides. Moving to the next step, the mature polypeptides were mapped for B and T cell epitopes through the online server of the Immune Epitope Database (IEBD) (31). The first epitopes to be predicted were CTLs, which were anticipated against HLA allele reference set that provided more than 97% in terms of population coverage (32). Secondly, HTLs were also predicted against HLA reference set to cover more than 99% in terms of population coverage (33). Moreover, HTL epitopes were analyzed for their ability to induce γ-INF vs. other cytokines through the webserver (http://crdd.osdd.net/raghava/ifnepitope/), IL-4 induction ability was estimated through the webserver (https://webs.iiitd.edu.in/raghava/il4pred/), while IL-10 induction ability was assessed through the webserver (https://webs.iiitd.edu.in/raghava/il10pred/). For both HTLs and CTLs, the affinity of the filtered peptides toward respective alleles was assessed where the peptide 3D structure was estimated *via* PEP FOLD 3 webserver (34); concurrently, the 3D structure of HLA-B*44:03 (PDB ID 4JQX) and HLA-DRB1*04:01 (PDB ID 5JLZ) was retrieved from the protein data bank to act as receptors for MHC-I and MHC-II epitopes, respectively, and the docking study was performed through AutoDock Vina (35). The final assessment for HTLs and CTLs was the conservancy prediction where the web-based tool from IEDB analysis resources was employed for this process (36). The last set of epitopes, BCLs were finally estimated through IEBD. After prediction, several characteristics for every single epitope were predicted to nominate the best candidates that would be used for the multitope construct. These characteristics were the population coverage, conservancy profile, antigenicity score, allergenicity, and toxicity probabilities; where IEDB analysis tools, Vaxijen, AllerTOP, and ToxinPred were employed, respectively, for these assessments.

Multitope Vaccine Construction

In order to construct a potential multitope vaccine, the best six candidates of CTL, HTL, and BCL epitopes from the step of epitope mapping were linked through GGGS, GPGPG, and KK amino acid linkers, respectively, where these linkers were incorporated to provide *in vivo* separation of the assembled epitopes (37). Apart from the epitopes, PADRE sequence and β defensin adjuvant were added to finalize the designed vaccine construct. This potential vaccine construct was assessed for its immunogenicity score, allergenicity, and toxicity probabilities through the same servers that were employed previously for single epitopes estimation.

Physicochemical Features, Protein Solubility Assessment, and Secondary Structure Prediction

ProtParam, a tool available on Expasy server (28) was employed to assess the physicochemical properties of the designed potential vaccine. The propensity upon overexpression in *E. coli* and the protein secondary structure of the multitope construct were anticipated *via* SOLpro server (38) and PSIPRED 4.0 webserver (39) respectively.

Tertiary Structure Prediction, Refinement, and Validation

It is essential to predict the potential vaccine 3D structure to be able to assess its binding with a toll-like receptor. The current study utilized 3Dpro webserver (40) for this purpose. Computational estimation of a protein tertiary structure is performed through a molecule bending and twisting to create a structure with the least possible energy and maximum state of stability where the employed server runs this process *via* an analysis of the structural similarity between the protein sequence under prediction and the data available on PDB. Following tertiary structure prediction, GalaxyRefine server (41) was employed to perform refinement for the generated structure in terms of stability and protein-energy, and the enhanced structures vs. the original one were assessed by Ramachandran plot analysis (42) and ProSA (43).

Conformational B-Cell Epitope Prediction

The conformational B-cell epitopes were analyzed for the multitope design after 3D structure prediction and refinement. The ElliPro Server (http://tools.iedb.org/ellipro), which is a reliable tool for detection of B cell epitopes against a specific antigen, was employed for this prediction (44).

Disulfide Engineering of the Designed Vaccine

Prior to initiating a docking study for the constructed potential vaccine, it was recommended to enhance the stability of the 3D format of the designed construct. Disulfide bonds were proved to enhance protein geometric conformation and consequently elevate its stability. Disulfide by Design 2.0 (45) was assigned for this process.

Docking of Designed Vaccine With *h*TLR-4

The current study employed molecular docking as a prediction tool for assessment of preferred orientation of the ligand, the current study vaccine construct, to its corresponding receptor and estimate the binding affinity (46). Inflammations triggered by *E. coli* are involved mainly with TLR-4 (47). Hence, the *h*TLR-4 (PDB id: 4G8A) was chosen as a receptor for the potential epitope-based constructed vaccine, the ligand, and ClusPro 2.0 server (48) was utilized to run this docking study. This server predicts the best docking models by performing billions of conformations, clustering of the 1,000 lowest energy

structures generated and removing steric clashes. The 2 PDB files were uploaded to ClusPro server and the docking process was performed using default parameters.

Dihedral Coordinate-Based Normal-Mode Analyses

The iMODS server was employed to investigate and analyze the collective flexibility as well as motion functions of the constructed epitope vaccine in relation to the bound *h*TLR-4 protein target since the latter server possesses the advantage of being fast and efficient (http://www.imods.chaconlab.org/) (49). This server can predict several values such as eigenvalues which reflected a harder deformation when this value is high (50). The atoms and residues of both the bound *h*TLR-4 protein and epitope vaccine ligand were continuously indexed where atoms of number range 1–9,567 and 9,568–15,225 were, respectively, assigned for the epitope-bound *h*TLR-4 (the initial 27–627 amino acids) and epitope vaccine itself (the following 1–380 amino acids).

Molecular Dynamics Simulations

The molecular dynamics computational approach was applied to describe the epitope/*h*TRL-4 molecular behaviors in addition to measuring the stability of such protein-protein complex (51). The *h*TLR-4 can exist in a monomeric state as well as m-shaped dimeric architecture where the latter is prior to the initiation of downstream signal transduction (52, 53). Additionally, reported crystalline *h*TLR-4 showed a significant *N*-glycosylation profile with several oligosaccharide moieties being linked to their surface at conserved residues (54, 55). In these regards, four docked *h*TLR-4/epitope vaccine complexes were investigated for the impact of glycosylation and/or *h*TLR-4 oligomerization state on the binding affinity as well as the thermodynamic stability of the constructed epitope vaccine. Simulated complexes were dimeric or monomeric *h*TLR-4 states being either glycosylated or not (Glycosylated dimer = GlyDim; Sole dimer = SolDim; Glycosylated monomer = GlyMon, Sole monomer = SolMon). Each of the latter adopted models were individually subjected to 100 ns explicit molecular dynamics runs using GROMACS-2019 (http://www.gromacs.org/) (56) and under CHARMM36m forcefield for the protein simulations (57–60). Under periodic boundary conditions, the TIP3P water model 3D-box was used to solvate the investigated protein-protein complex model with 10 Å marginal distances (61). The standard ionization state of both protein amino acids was assigned under physiological pH 7.4, whereas the whole constructed system charge was *via* sufficient chloride and potassium ion numbers introduced through Monte-Carlo ion-placement method (62). The atomic counts of the four constructed models were 438642, 424612, 242818, and 237898 atoms for GlyDim; SolDim; GlyMon, and SolMon, respectively.

The constructed system was minimized throughout 5 ps using a steepest-descent algorithm (63), and subsequently subjected to double-staged equilibration for 100 ps/stage under a constant number of particles, Volume, and Temperature (NVT; 303.15 K, Berendsen temperature coupling regulation) and a constant number of particles, Pressure, and Temperature (NPT; 303.15 K and 1 atm. Pressure, Parrinello-Rahman barostat regulation) ensembles for the first and second stage, respectively. Throughout both minimization and equilibration stages, the original protein foldings were preserved and all heavy atoms were restrained at 1,000 kJ/mol.nm^2 force constant. Finally, the minimized/equilibrated systems were produced through 100 ns explicit molecular dynamics runs under NPT ensemble using Particle-Mesh Ewald algorithm for long-range electrostatic interactions computation. Linear constraint LINCS method modeled the covalent bond lengths at 2 fs integration time step sizes (64). van der Waals and Coulomb's non-bounded interactions were truncated at 10 Å *via* Verlet cut-off schemes (65). The MD simulations were performed using Aziz® Supercomputer (King Abdulaziz University's High-Performance Computing Center), *via* 5 nodes with 24 CPUs/node and 8 MPI processes/node. Adopted nodes run CentOS-6.4 with dual Intel® E5-2695v2 (24 cores/node; i.e. 2.4 GHz 12 Cores) offering 96 GB Memory/node. The MD simulation of the dimer *h*TLR-4 states took ∼228 and 218 h for the glycosylated and non-glycosylated proteins, respectively. While as the monomeric states took nearly 132 and 128 h for the glycosylated and non-glycosylated proteins, respectively.

The trajectory-oriented analytical parameters; root-mean-square deviation (RMSD), RMS-fluctuation (RMSF), radius of gyration (Rg), and solvent-accessible surface area (SASA) were computed through molecular dynamics trajectory analysis using GROMACS built-in analytical scripts. The free binding energies, as well as residue-wise energy contribution between *h*TLR-4 protein target and epitope vaccine ligand, were estimated *via* Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) using GROMACS/*g_mmpbsa* scripts (66). The SASA-only model of the free-binding energy ($\Delta G_{Total} = \Delta G_{\text{MolecularMechanics}} + \Delta G_{Polar} + \Delta G_{Apolar}$) was used across the whole 100 ns molecular dynamics runs. Important MM/PBSA parameters for polar/solvation calculations were set at 1.40 Å solvent probe radius, 80 pdie solvent dielectric constants, 1 vdie standard vacuume, and 2 pdie solute dielectric constants. Regarding only-SASA non-polar solvation; 1.40 Å SASA solvent probe radius, 0.0227 kJ/mol.Å^2 solvent surface tension, and 3.8493 kJ/mol offset constant, were used. Finally, parameters for the continuum-integral-based model were set as 1.25 Å solvent probe radius, 200 quadrature points/Å^2, and 0.0334 Å^3 bulk solvent density. Representing ligand-protein conformations at specific timeframes was done *via* Schrödinger-Pymol V.2.0.6 graphical package.

Immune Simulation of the Designed Vaccine

The C-ImmSim server (67) was employed to predict the stimulated immune response against the designed vaccine through a computational approach. We followed the technique of prime—booster—booster for this investigation and that was achieved by injecting the designed vaccine three times with 4 weeks intervals. This approach was applied to obtain a long-lasting immune response.

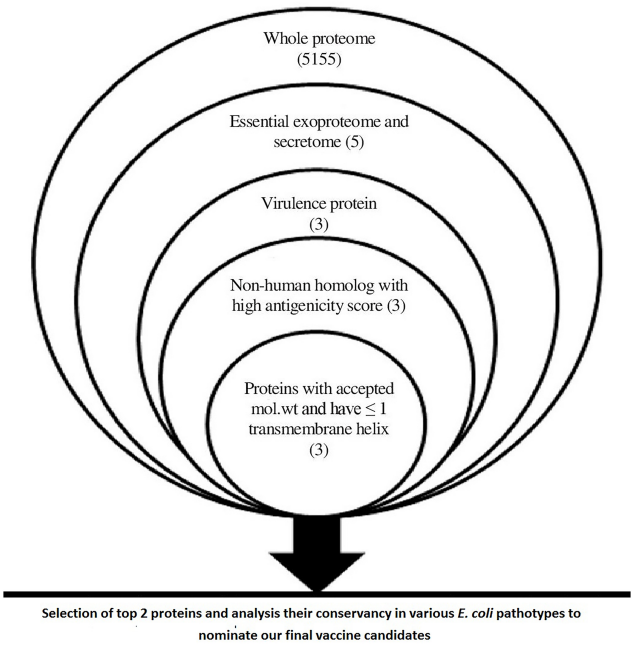


FIGURE 2 | A chart summarizing applied filtration steps for the nomination of potential vaccine candidates by applying *in silico* reverse vaccinology technique.

RESULTS

The main output of the filtration steps that have been applied in the current study to define potential candidates are shown in **Figure 2**.

Proteome Analysis for Selection of Vaccine Candidates

Data within **Figure 2** demonstrates the results of filtration steps of *E. coli* O157:H7 str. Sakai proteome where three proteins (lipopolysaccharide assembly protein LptD, outer membrane protein assembly factor BamA, and lipopolysaccharide assembly protein LptE) matched all the requirements. We selected the top 2 proteins that passed all the filtration steps and had the highest antigenicity score to obtain an accepted length of multiepitope vaccine after connecting the best-predicted T and B cell epitopes from these 2 proteins so we nominated LptD and BamA as our final vaccine candidates, characteristics of these candidates are shown in **Supplementary Table 1**.

Conservation Analysis of Nominated Vaccine Candidates Among *E. coli* Pathotypes

Conservation analysis of our nominated vaccine candidates showed that these proteins were found with high conservation among various *E. coli* pathotypes (**Supplementary Table 2**) which confirmed their ability to protect against different *E. coli* infections, therefore we designed the multitope vaccine based on these 2 proteins.

T Cell Epitopes

In order to select the best T cell epitopes for constructing the multitope vaccine, the top 100 generated peptides from IEBD per each protein candidate were estimated for their antigenicity score, allergenicity, and toxicity probabilities, and the top 10 peptides, which demonstrated the lowest percentile rank, the highest binding affinity and antigenicity score more than 0.4 were grouped in **Table 1** (for MHC-I peptides) and **Table 2** (for MHC-II peptides). The binding affinity of the selected candidates of CTLs and HTLs was assessed through a docking study where the generated docked complexes for CTLs are shown in **Supplementary Figure 1** and the docked complexes for HTLs are shown in **Supplementary Figure 2**. The binding scores of all complexes are shown in **Supplementary Table 3**. These scores ranged between −7.6 and −8.5 for CTLs, and from −7.1 to −8.1 for HTLs. The receptors of this docking study were deposited in the protein data bank with docked peptides that were firstly removed before running the current docking analysis then docked again separately to their respective receptor to act as a control for our docking study. The binding score for MHC-I control was −7.5 and that of MHC-II was −7.6, By comparing the binding energy score for our candidate list and the controls we can confirm that these candidates are promising to be selected for the construction of the multitope vaccine. Another significant factor that was considered in the selection of the epitopes for constructing the multitope vaccine was the population coverage. The IEDB population coverage analysis tool was employed for this process. The whole list of predicted epitopes for both protein candidates showed the epitopes arranged in descending way based on their binding affinity to the different alleles therefore we selected the top 10%, that would represent the epitopes with high binding affinity to respective alleles, and collected these alleles per each epitope to analyze the population coverage for single epitopes (**Supplementary Table 4**) then we analyzed the population coverage for the combined CTLs, combined HTLs and the multitope vaccine (**Supplementary Table 5**).

B-Cell Epitope Identification

Bepipred Linear Epitope Prediction 2.0 was used as a prediction method. It identified 30 and 21 B-cell epitopes for BamA and LptD proteins, respectively. Peptides with a length between 9:18 amino acids were analyzed for their antigenicity and peptides with antigenicity score >0.4 were tested for their allergenicity and toxicity (**Table 3**).

Construction of Multitope Vaccine

From **Tables 1**–**3**, six epitopes per each table (three from each protein candidate) were chosen based to constitute the basis of the multitope vaccine (graphical representation for the constructive components is shown in **Supplementary Figure 3**). Moreover, β-defensin and PADRE peptide were also incorporated to finalize a potential vaccine sequence of 380 amino acids in length and its sequence was as the following:

TABLE 1 | Top-ranked T-cell epitopes (MHC-I peptides) of BamA and LptD proteins.

Epitope	Protein	Antigenicity	Allergenicity	Toxicity	Immunogenicity	Conservancy (%)
KTDDFTFNY	BamA	1.74	Non	Non	0.2750	100
FKTDDFTFNY	BamA	1.54	Non	Non	0.3020	100
MSAGIALQW	BamA	1.41	Non	Non	0.1361	100
NVDAGNRFY	BamA	0.41	Allergen	Non	0.1855	100
AELSVTNPY	BamA	0.97	Non	Non	−0.1052	100
AEIQQINIV	BamA	0.78	Non	Non	0.0114	100
YANSVRTSF	BamA	0.53	Allergen	Non	−0.1435	100
RMSAGIALQW	BamA	1.5	Allergen	Non	0.1080	100
QADDADLSDY	BamA	0.82	Non	Non	−0.0478	100
YSDPSNIRM	BamA	0.82	Non	Non	−0.0285	100
RTGDDNITW	LptD	1.7	Allergen	Non	0.1838	100
FSEQNTSSY	LptD	0.43	Non	Non	−0.2901	100
KLDESVNRV	LptD	0.79	Non	Non	0.0104	100
RIYGQAVHF	LptD	0.66	Non	Non	0.0106	100
SPEYIQATL	LptD	0.49	Allergen	Non	0.1050	100
ATSNSSIEY	LptD	1.13	Allergen	Non	−0.2062	100
KVGPVSIFY	LptD	0.71	Non	Non	0.0650	100
TLEPRAQYLY	LptD	1.04	Non	Non	−0.0088	100
IYDDAAVERF	LptD	0.5	Allergen	Non	0.2593	100
KQADSMLGV	LptD	0.8	Allergen	Non	−0.2809	100

TABLE 2 | Top-ranked T-cell epitopes (MHC-II peptides) of BamA and LptD proteins.

Epitope	Protein	Antigenicity	Allergenicity	Toxicity	INF-γ	IL-4	IL-10	Conservancy (%)
DPSNIRMSAGIALQW	BamA	1.26	Non-allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
PSNIRMSAGIALQWM	BamA	1.22	Non-allergenic	Non-toxic	Positive	Non-inducer	Non-inducer	100
SNIRMSAGIALQWMS	BamA	1.03	Non-allergenic	Non-toxic	Positive	Non-inducer	Non-inducer	100
KLAGDLETLRSYYLD	BamA	0.66	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
QKLAGDLETLRSYYL	BamA	0.62	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
NIRMSAGIALQWMSP	BamA	1.25	Non-allergenic	Non-toxic	Positive	Non-inducer	Non-inducer	100
QRVAVGAALLSMPVR	BamA	0.51	Non-allergenic	Non-toxic	Positive	Non-inducer	Non-inducer	100
DYTNKSYGTDVTLGF	BamA	0.79	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
NKSYGTDVTLGFPIN	BamA	0.99	Non-allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
TNKSYGTDVTLGFPI	BamA	1.07	Non-allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
GPVSIFYSPYLQLPV	LptD	0.68	Non-allergenic	Non-toxic	Positive	Non-inducer	Non-inducer	100
VQLNYRYASPEYIQA	LptD	1.11	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
VGPVSIFYSPYLQLP	LptD	0.69	Allergenic	Non-toxic	Positive	Inducer	Inducer	100
VSIFYSPYLQLPVGD	LptD	0.45	Allergenic	Non-toxic	Positive	Inducer	Inducer	100
LNYRYASPEYIQATL	LptD	0.88	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
KVGPVSIFYSPYLQL	LptD	0.88	Non-allergenic	Non-toxic	Positive	Inducer	Inducer	100
NYRYASPEYIQATLP	LptD	0.66	Allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
YLPYYWNIAPNMDAT	LptD	1.76	Allergenic	Non-toxic	Positive	Inducer	Inducer	100
AKYTTTNYFEFYLPY	LptD	1.13	Non-allergenic	Non-toxic	Positive	Inducer	Non-inducer	100
SSIEYRRDEDRLVQL	LptD	0.81	Non-allergenic	Non-toxic	Positive	Non-inducer	Inducer	100

"EAAAKGIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCS TRGRKCCRRKKEAAAKAKFVAAWTLKAAAGGGSKTDDFT FNYGGGSAELSVTNPYGGGSAEIQQINIVGGGSFSEQNTSSY GGGSRIYGQAVHFGGGSTLEPRAQYLYGPGPGDPSNIRMS AGIALQWGPGPGQRVAVGAALLSMPVRGPGPGTNKSYGT DVTLGFPIGPGPGKVGPVSIFYSPYLQLGPGPGAKYTTTNYF EFYLPYGPGPGSSIEYRRDEDRLVQLKKASGVRVGESLDRTK KIRFEGNDTSKDAVKKTDTQRVPGSPKKKEAPGQPEPVKK PSYFNDFDNKYGSSTDGYKKQTNLDWYNSRNTTKLDESKK AKFVAAWTLKAAAGGGS".

TABLE 3 | Predicted B-cell epitopes of BamA and LptD proteins.

BamA				LptD			
Epitope	**Antigenicity**	**Allerge-nicity**	**Toxicity**	**Epitope**	**Antigenicity**	**Allerge-nicity**	**Toxicity**
PVRTGDTVNDEDIS	1.24	Allergen	Non	KEAPGQPEPV	0.96	Non	Non
ASGVRVGESLDRT	0.97	Non	Non	DKVYEDEHPNDDSS	0.41	Non	Non
IRFEGNDTSKDAV	1.14	Non	Non	PSYFNDFDNKYGSSTDGY	1.22	Non	Non
TDTQRVPGSP	0.49	Non	Non	QVFSEQNTSSYS	0.42	Non	Non
FQADDADLSDYTNK	0.56	Non	Non	QTNLDWYNSRNTTKLDES	0.75	Non	Non

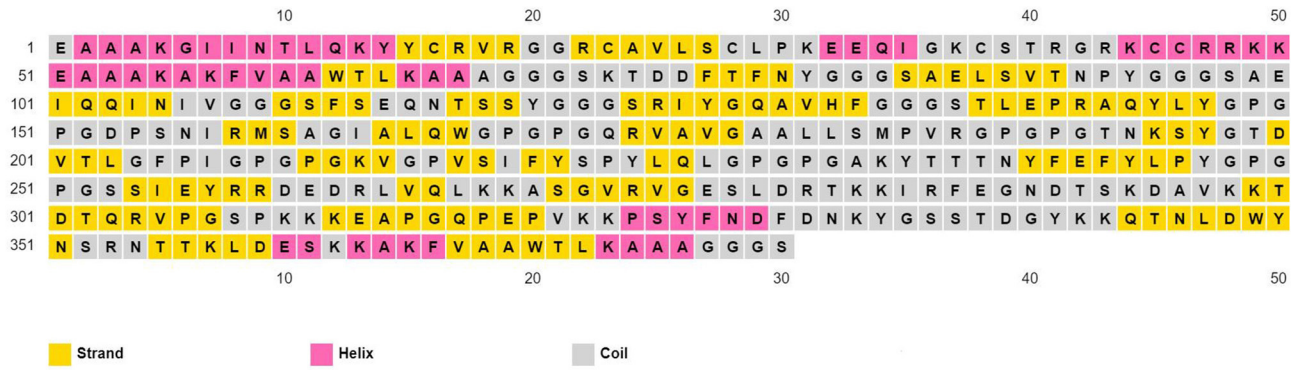


FIGURE 3 | Secondary structure prediction of designed multitope vaccine using PESIPRED server.

This construct was assessed to be non-allergen with an antigenicity score of 1.07 (estimated by VaxiJen v2.0) and 0.959 (estimated by ANTIGENpro).

Physicochemical Properties Assessment and Secondary Structure Prediction

The physicochemical properties of the predicted vaccine construct were detected by using the ProtParam server and demonstrated in **Supplementary Table 6**. The designed vaccine had a SOLpro SVM score of 0.95; therefore it was predicted to be soluble as SOLpro values > 0.5 are considered as soluble. Vaccine secondary structure prediction demonstrated the presence of 14.2% helix, 37.9% strand, and 47.9% coil structure (**Figure 3**).

Tertiary Structure Prediction, Refinement, and Validation

The assessment of the primary 3D structure, through Ramachandran plot analysis and ProSA online server, demonstrated that 87.6, 6.3, and 6.1% of residues were located in favored, allowed, and outlier regions, and the *Z*-score −3.45, respectively. While these values may be considered acceptable for a predicted 3D structure, we continued with refinement for better structure creation. Protein refinement occurred through the help of GalaxyRefine web server and the best model, regarding scores improvement, **Figure 4A** demonstrated the following scores' enhancement, the Z-score enhanced from −3.45 to −3.9 (**Figure 4B**) while the Ramachandran plot analysis scores became 96%, 2.4%, and 1.6% for residues in favored, allowed, and outlier regions, respectively (**Figure 4C**).

Conformational B-Cell Epitope Prediction

The tertiary structure and folding of the designed vaccine may generate new conformational B-cell epitopes and for this purpose, we used ElliPro server conformational. In the current assessment, the server predicts 9 new epitopes and their scores were between 0.514 and 0.84 (**Supplementary Table 7**). The predicted 3D models of the generated epitopes are shown in **Supplementary Figure 4**.

Vaccine Disulfide Engineering

Usage of DbD2 server for disulfide bond assign demonstrated that 26 pairs of amino acids are eligible to make disulfide bonds while in terms of other parameters such as energy and Chi3 value, this number reduced to only 2 pairs. Therefore, 4 mutations were performed at the residues pairs of SER89-ARG187 and PRO324-PHE330. The followed values of energy and Chi3 to recommend disulfide engineering were below 2.2 and from −87 to +97, respectively.

Molecular Docking of the Vaccine With TLR4

The ClusPro 2.0, which was employed for the docking study, generated 30 models and the model number 0.00 (**Figure 5**) exhibited the lowest binding energy score of −1420.9 kcal/mol which implicate a good affinity and stability of the constructed complex.

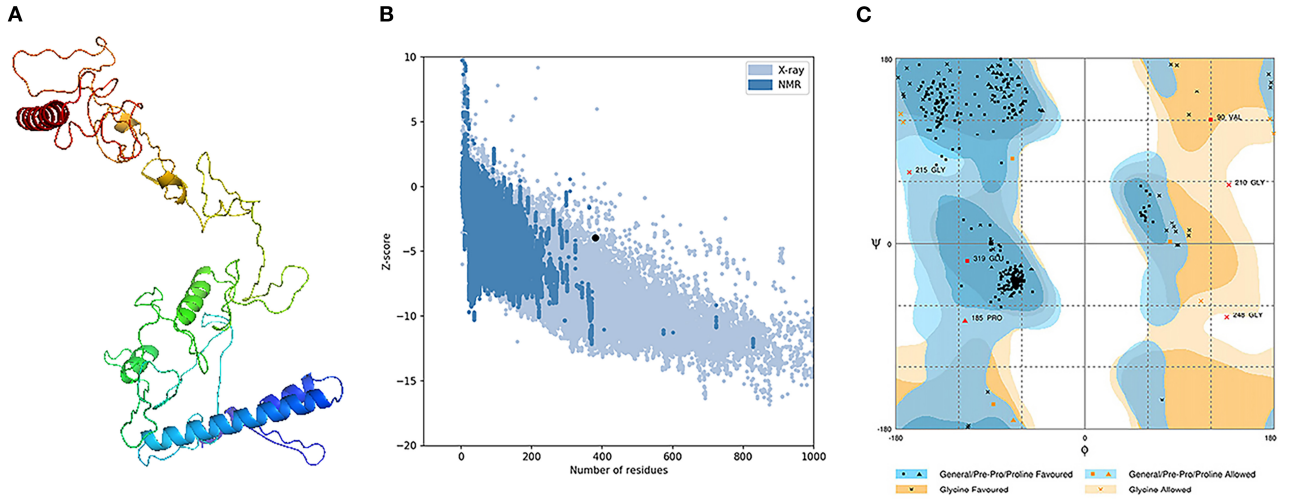


FIGURE 4 | Structural analysis of the predicted 3D structure of the vaccine. **(A)** The three-dimensional structure of vaccine obtained after molecular refinements; **(B)** ProSA-web evaluation of the vaccine structure; **(C)** Ramachandran plot analysis of the protein structure after molecular refinements.

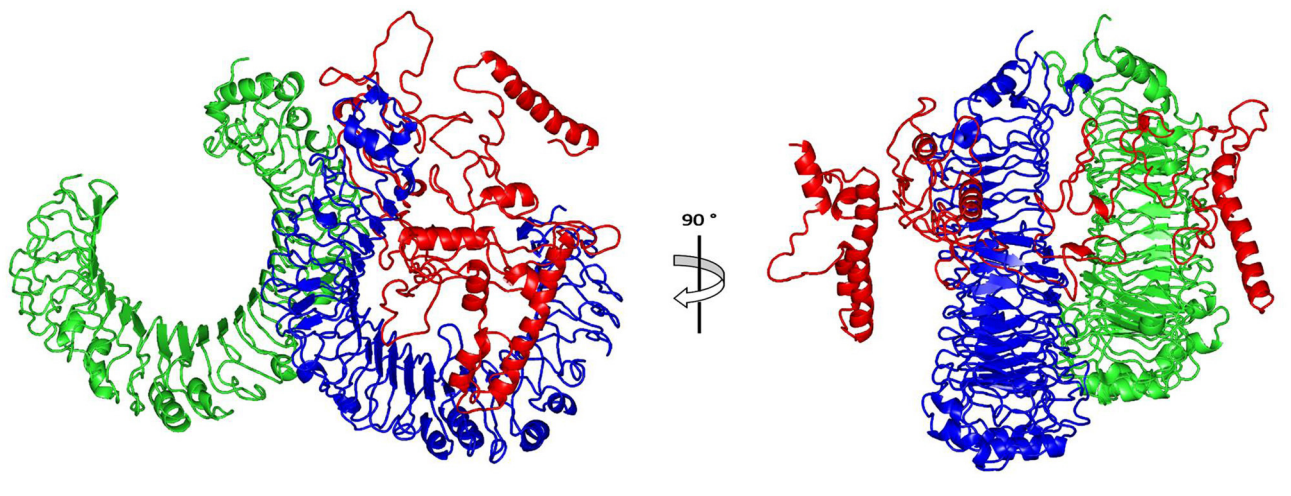


FIGURE 5 | Docked complex of vaccine constructs with human TLR4; vaccine constructs in red color and TLR4 receptor in green and blue colors.

Dihedral Coordinate-Based Normal-Mode Analyses

Within the iMOS server finding analysis, B-factor correlates with the relative magnitude of atom displacement around conformational equilibria. Values were significantly higher for the epitope vaccine ligand (atom index; 9,568–15,225), particularly at its respective carboxy terminus, in relation to that of the *h*TLR-4 target protein (atom index; 1–9,567) (**Figure 6A**). The B-factor results were recapitulated *via* the complex deformability index presented in **Figure 6B** where each vaccine residue, particularly at the carboxy end, showed individual distortions being higher than those of the *h*TLR-4 target protein. The estimated eigenvalue, which represents the motion stiffness of the complex, was $1.95e^{-06}$, where being in inverse order in relation to variance predicting the significantly higher mobility of the vaccine as compared to the *h*TLR-2 complex across collective functional motions (**Figures 6C,D**). The iMOS provided the covariance matrix illustrating the coupled residue pairs demonstrating anti-correlated (blue color), correlated (red color), or uncorrelated (white color) motions. The *h*TLR-4 depicted lower predicted correlated residue-pair motions than did the epitope vaccine ligand, however, the latter protein showed less anti-correlated motions (**Figure 6E**). Finally, the obtained elastic-network model explains the differential flexibility patterns among both investigated proteins (**Figure 6F**). Represented in different colors, the elastic-network model describes the atom pairs linked *via* springs relying on the stiffness degree between them. Stiffer strings were correlated to dark gray colors. The target *h*TLR-4 protein showed continuous dark-gray bands along the normal distribution of stiffer string, while the

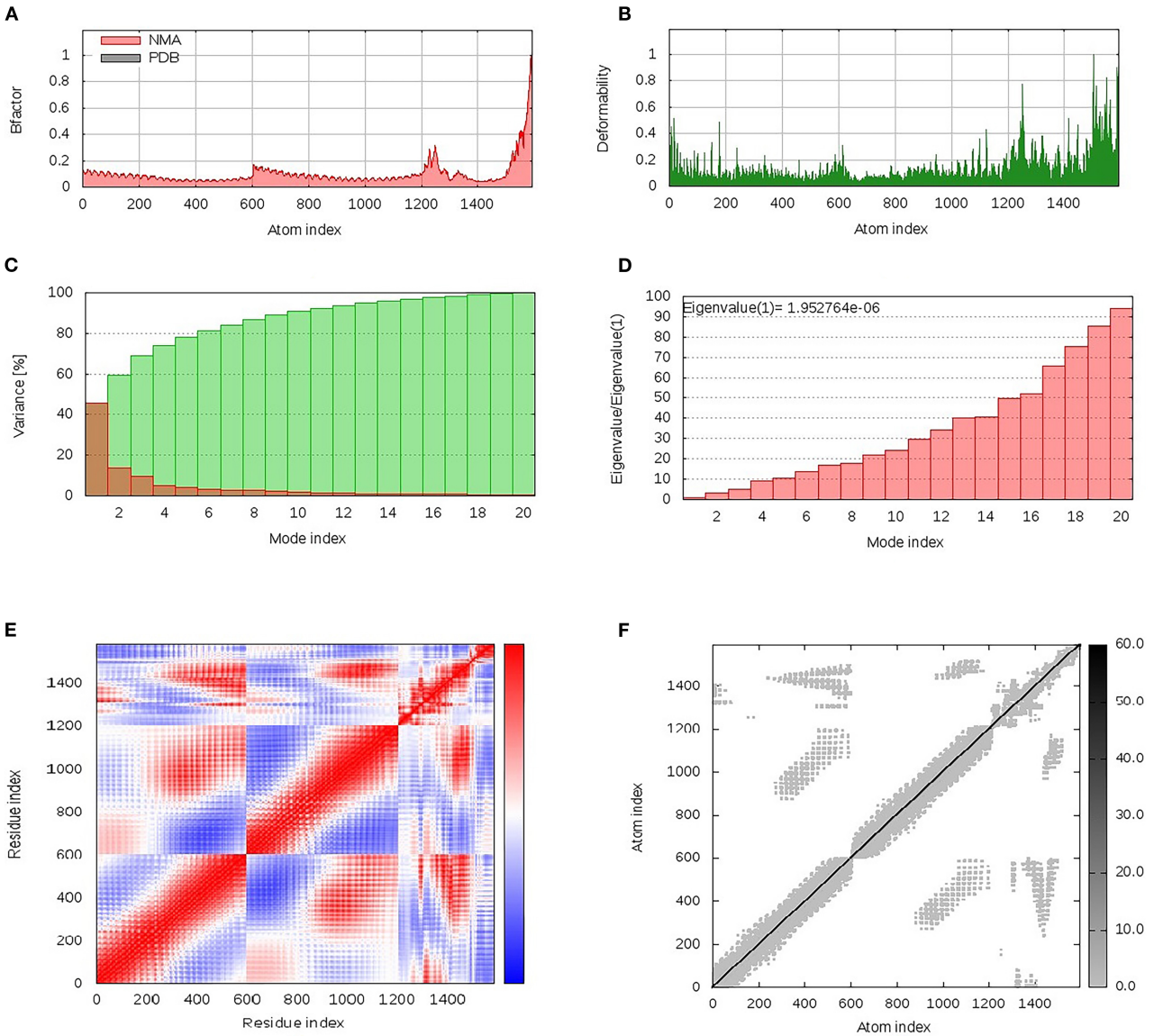


FIGURE 6 | Dihedral coordinate-based normal-mode analyses of multitope vaccine-*h*TLR-4 complex; ligand-receptor interaction was assessed throughout comparative **(A)** B-factor indices, **(B)** deformabilities, **(C)** variance, **(D)** eigenvalues, **(E)** covariance of residue indices, and **(F)** elastic network analysis.

residues of the epitope vaccine illustrated non-continuous gray bands around the same immobility normal string, particularly for those residues near the carboxy terminus.

Molecular Dynamics Simulation Analyses

Throughout the 100 ns molecular dynamics runs, each simulated protein depicted thermodynamic behavior being typical through molecular dynamics simulations (**Figure 7**). The monitored proteins' RMSD deviations, in relation to corresponding alpha-carbon (*Cα*-RMSD) of the reference protein, showed an initial increase over the initial frames owing to the release of constraints at the beginning of the simulation stage. Steady RMSD tones were depicted for almost all protein models beyond the first 30 ns and till the end of the molecular dynamics runs (i.e., for >75 ns). The RMSD trajectories for the epitope vaccine and respective bound hTLR-4 were around two-fold differences the thing that ensured sufficient protein convergence as well as significant ligand accommodation at the target pocket. The latter findings ensured the adequacy of the 100 ns MD simulation timeline to grasp sufficient thermodynamic information within efficient computational cost and without the need for more time extension. Regarding comparative *Cα*-RMSDs analysis between the *h*TLR-4 target proteins and their corresponding in-bound epitope vaccine ligands, higher RMSD trajectories were illustrated for the ligand proteins (**Figures 7A,B**). The latter was obvious since the obtained average *Cα*-RMSD values across

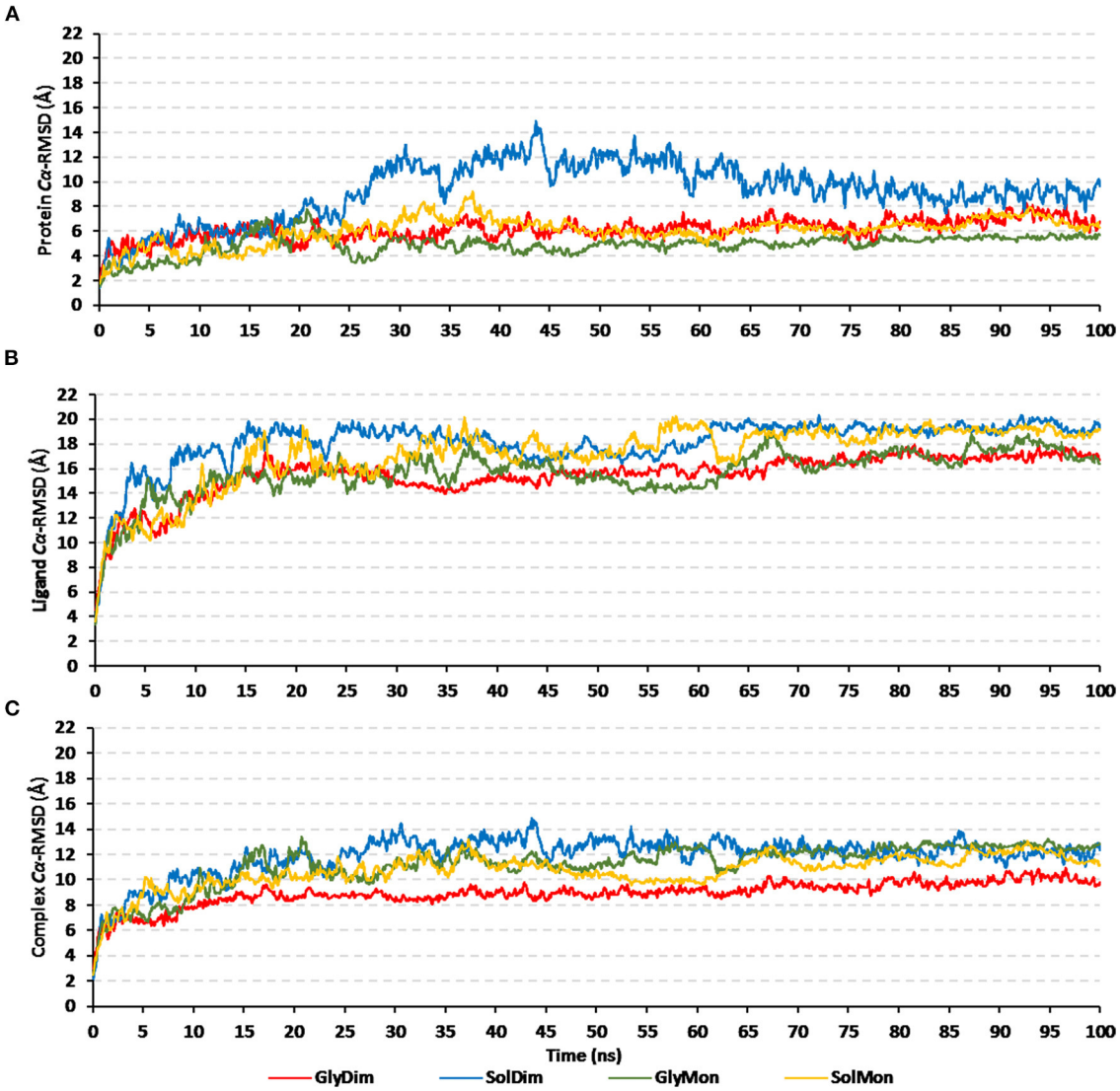


FIGURE 7 | RMSD trajectory analysis for the four investigated epitope vaccine ligand-bound *h*TLR-4 target models across 100 ns explicit molecular dynamics simulation runs. Time-evolution of **(A)** Proteins' *Cα*-RMSD, **(B)** Ligands' *Cα*-RMSD, and **(C)** Complexes' *Cα*-RMSD, along molecular dynamics timeframes (ns).

the protein's trajectory plateau and till MD simulation end courses were 8.58 ± 0.72 and 15.92 ± 0.32 Å for the *h*TLR-4 and vaccine ligands, respectively. Concerning the comparative RMSD analysis for the oligomeric *h*TLR-4 states, steadier and lower *Cα*-RMSD fluctuation were depicted for the dimeric state as compared to those of the monomer ones. On similar bases, significantly lower and more steady *Cα*-RMSD tones were depicted within models where the *h*TLR-4 target proteins were covalently bound to higher oligosaccharides. Except for limited fluctuations, the monitored *Cα*-RMSDs of the combined *h*TLR-4/epitope proteins showed rapid equilibration plateaus (11.23 ± 0.11 Å), beyond 25 ns and till the end of the simulations runs (**Figure 7C**). Notably, the least fluctuating complex *Cα*-RMSD trajectories (9.30 ± 0.57 Å) were seen for the dimeric *N*-glycosylated *h*TLR-4/vaccine system.

Monitoring the *Cα*-RMSF tones across the whole simulated trajectories (100 ns) provided interesting information regarding the residue-wise fluctuation pattern of each corresponding simulated protein. Interestingly, typical well-behaved molecular simulation profiles were depicted for each simulated protein where terminal residues and their vicinal ranges showed higher mobility patterns (high *Cα*-RMSF) as compared to those at the core regions (**Figure 8**). As a general observation, the epitope vaccine ligands showed much higher residue-wise fluctuation profiles as compared to their corresponding in-bound *h*TLR-4 target proteins. The latter observation was most recognized for the vaccines' respective carboxy end amino acids as well as their vicinal residues (high residue sequence numbering; from 301 to 380) in relation to those of the *N*-terminus. Notably, the depicted vaccine-oriented fluctuation trends were more associated with

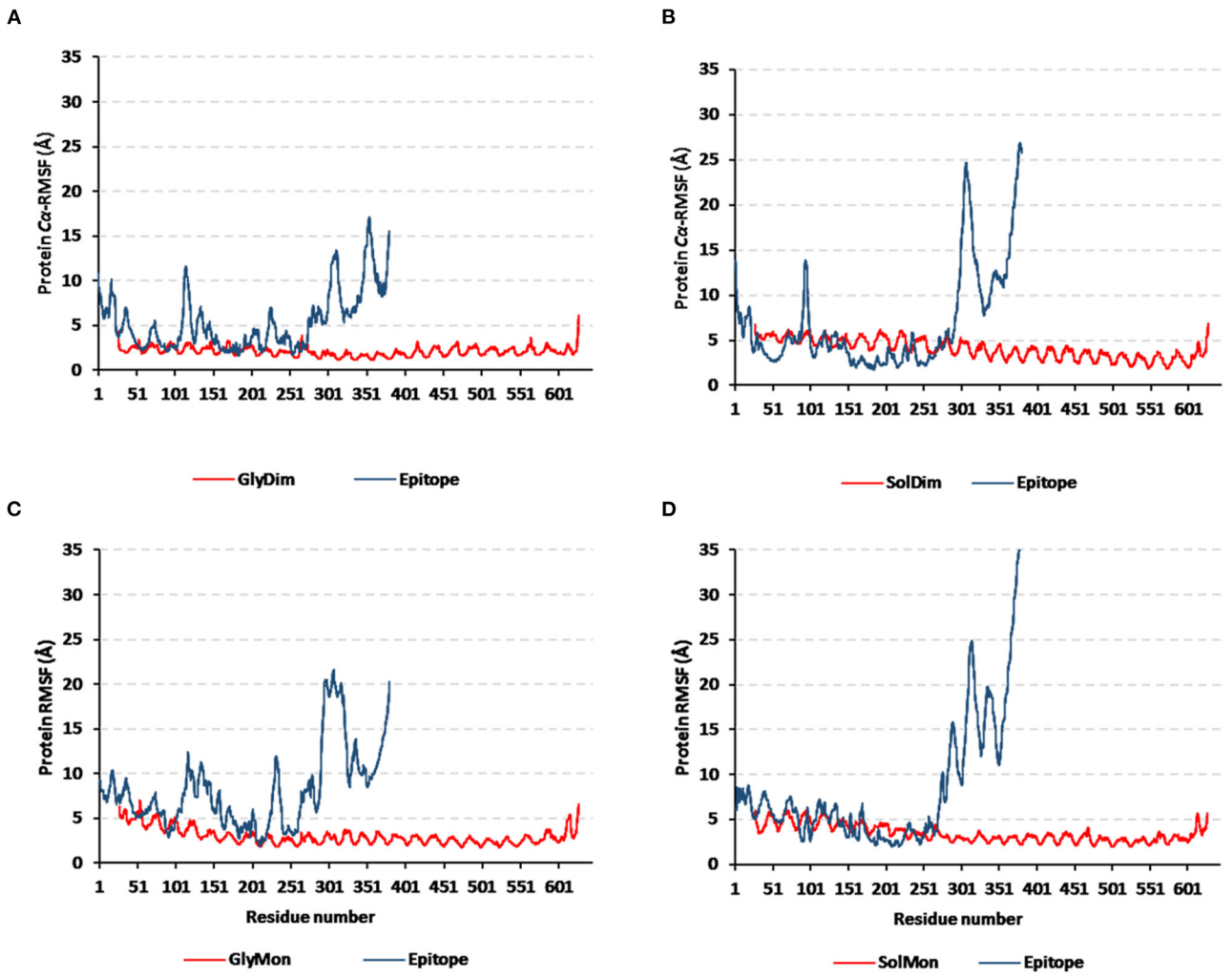


FIGURE 8 | Protein's residue-wise fluctuation patterns. Monitored *Cα*-RMSF tones across the whole 100 ns molecular dynamics runs for the investigated hTLR-4 bound epitope vaccine within different oligomeric/saccharide-linked states. **(A)** glycosylated dimeric, **(B)** non-glycosylated dimeric, **(C)** glycosylated monomeric, and **(D)** non-glycosylated monomeric states, along the sequence residue numbers (bound TLR-4 protein target = 27–627; epitope vaccine = 1–380 residue ranges).

monomeric/non-*N*-glycosylated *h*TLR-4 state (*Cα*-RMSF up to 35 Å) as compared to other oligomeric models. On similar bases, the lowest RMSF-based mobility trends were assigned to the *N*-glycosylated dimeric *h*TLR-4 model as compared to any other simulated model.

Subsequent analysis of the key alterations within the conformations of both simulated epitope vaccine and *h*TLR-4 protein was proceeded throughout examining each simulated model at the initial and last molecular dynamics timeframes. Extracted frame lines at 0 and 100 ns were subjected to 1×10^{-3} Kcal/mol.A^2 gradient minimization using MOE-Molecular Operating Environment software. The RMSDs of the overlaid conformations were 6.645, 9.855, 5.488, and 7.784 Å for the glycosylated dimeric, non-glycosylated dimeric, glycosylated monomeric, and non-glycosylated monomeric complexes. Notably, all simulated models illustrated stable binding states for the epitope vaccine at the *h*TLR-4 binding site (**Figure 9**). Limited conformational changes were assigned for the *h*TLR-4 target protein across the four simulated models. On the contrarily, the simulated epitope vaccines depicted significant conformational changes causing them to adopt a more compacted conformation/orientation at the *h*TLR-4 binding site. Within the four simulated models, the epitope vaccine showed more profound movement for its carboxy end and vicinal regions as compared to that of its *N*-terminus, the thing that allowed proximity of the vaccine *C*-terminus toward the *h*TLR-4 lateral side and near the *h*TLR-4 1:1 homodimerization interface. The latter dynamic behavior was most recognized at the monomeric *h*TLR-4 states as compared to the dimeric ones, as well as at the non-glycosylated *h*TLR-4 state in relation to that bounded to the higher oligosaccharides. Notably, applying both Rg and SASA analysis for the epitope vaccine within the four simulated models showed comparable findings (**Supplementary Figure 5**). Both analytical parameters

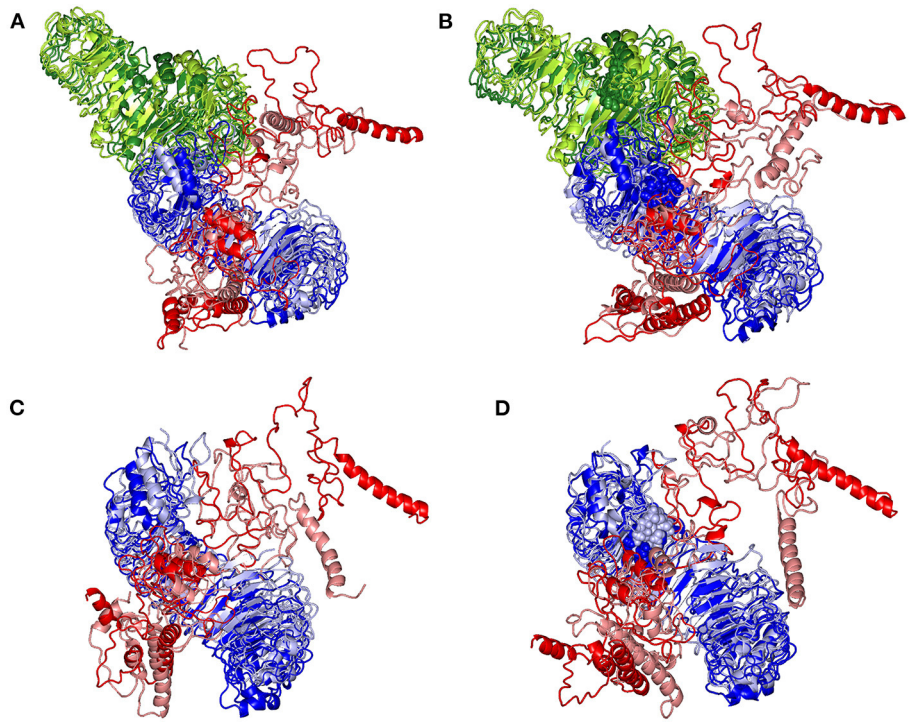


FIGURE 9 | Conformational changes for the simulated *h*TLR-4/epitope vaccine complexes. Overlaid snapshots of the 3D-cartoon representation of the; **(A)** glycosylated dimeric, **(B)** non-glycosylated dimeric, **(C)** glycosylated monomeric, and **(D)** non-glycosylated monomeric states at 0 ns and 100 ns of the molecular dynamics runs. The *h*TLR-4 target proteins (blue/green) and epitope vaccine ligands (red) are colored in respective to the initial or last extracted frames (dark and light colors are for 0 ns and 100 ns extracted frames, respectively). Moieties of *N*-glycosylation are presented as spheres and colored in respective to their linked *h*TLR-4 protomer and extracted frame.

TABLE 4 | Free binding energies ($\Delta G_{Total\ binding}$) and individual energy terms regarding the designed multitope vaccine at *h*TLR-4 target protein binding site.

Energy (kJ/mol ± SD)	Ligand-receptor complex			
	Glycosylated dimer	**Non-glycosylated dimer**	**Glycosylated monomer**	**Non-glycosylated monomer**
$\Delta G_{van\ der\ Waal}$	−732.671 ± 72.004	−780.821 ± 82.502	−782.349 ± 115.500	−932.685 ± 23.538
$\Delta G_{Electrostatic}$	−7954.091 ± 79.714	−7798.818 ± 11.828	−7845.389 ± 86.071	−7710.021 ± 97.634
$\Delta G_{Solvation;\ Polar}$	1529.300 ± 11.244	1650.603 ± 18.770	1497.129 ± 39.595	1743.513 ± 12.268
$\Delta G_{Solvation;\ SASA}$	−98.206 ± 13.476	−106.847 ± 9.345	−97.202 ± 16.784	−111.119 ± 24.432
$\Delta G_{Total\ binding}$	−7255.668 ± 73.478	−7035.883 ± 38.444	−7227.811 ± 20.760	−7010.312 ± 11.335

showed high values at the beginning of the MD simulation runs, yet as the simulation proceeded, the epitope vaccine attained lower as well as much steady plateaued trajectories till the end of the MD timeframes. It is worth mentioning that higher Rg fluctuations tones (around 15-65 ns), as well as late equilibration plateau (not before 70 ns), were achieved for the non-glucosylated/monomeric states as compared to the glycosylated/dimeric ones (**Supplementary Figure 5A**). Similar findings were also illustrated with SASA analysis where at non-glucosylated/monomeric states, the epitope vaccine exhibited higher fluctuations around 40 ns and till the end of the MD runs (**Supplementary Figure 5B**).

The designed epitope vaccine across the simulated models showed significant free binding energies toward the *h*TLR-4 binding sites being estimated as kJ/mol ± SD (**Table 4**). The highest negative values of binding-free energies were assigned to the dimeric states of *h*TLR-4 rather than their respective monomeric ones. Additionally, the glycosyl-bound target proteins depicted a higher affinity toward the anchored epitope vaccine than those being non-glycosylated. Dissecting the obtained total free binding energies $\Delta G_{Total\ binding}$ showed a preferential energy contribution for the electrostatic non-bonded interactions over the van der Waal potentials. Across the four simulated models, the comparable pattern of energy term contributions was depicted depending on the glycosylation status of the target *h*TLR-4 protein. Differentially higher electrostatic and lower van der Waal values were depicted for the glycosylated target protein states over the monomeric ones. Regarding the

solvation energy term contributions, both non-polar and polar solvation energies were of higher negative and positive values, respectively, for the non-glycosylated target proteins over their respective glycosylated forms. The non-polar solvation energy of interaction was monitored across the 100 ns of MD simulation using the SASA-only model calculation for the individual *h*TLR-4, epitope vaccine, as well as their respective combination at each oligomeric/glucosylic state (**Supplementary Figure 6**). Almost steady solvation energy terms were depicted for the four simulated proteins across the whole MD simulation run. On the other hand, the ligand's solvation energy of interaction showed a significant drop to lower values starting from around 40 ns and till the end of the MD runs achieving the average energy plateau for more than half the simulation timeframes. Steadier solvation energy tones were assigned for the dimeric/glycosylated targets rather than those of the lower simpler states. Notably, the complex solvation energy patterns were significantly impacted *via* the ligand's values and in turn its respective dynamic behavior rather with those of the simulated *h*TLR-4 proteins.

The $\Delta G_{Total\ binding}$ was further decomposed identifying the residue-wise energy contribution for both the vaccine and *h*TLR-4 where the more negative is the better (**Figure 10** and **Supplementary Figure 7**). Findings within the latter figure showed higher positive-valued energy contributions for the vaccines in bound to non-glycosylated *h*TLR-4 states, while the vaccines of glycosylated forms depicted significant negative-valued energy contributions with a wider range of contributing residues. Similar residue-wise energy contribution patterns were assigned for the *h*TLR-4 residues.

Immune Simulation of the Designed Vaccine

The immune response regarding antibody titer, cytokines level, B and T cells population is shown in **Figure 11**. The current study potential vaccine was estimated to stimulate a high level of IgM+IgG which increases with consecutive doses of the vaccine. Regarding cytokines level, several classes were stimulated and INF-γ exhibited the highest level of induced cytokine. Moreover, the count of stimulated T and B cells demonstrated an increase with the doses of the vaccine, and the highest level was obtained as a result of the second booster dose of the vaccine.

DISCUSSION

Many microorganisms face difficulty in cultivation or attenuation leading to undesirable immune response, proving that the classical approaches for vaccine development against these pathogens require a technical revolution (68). Therefore, the last few years witnessed a large turn in the employed approaches for vaccine development where the multi-omics approaches stepped forward and preceded the traditional ones (69). Recent studies that utilized bioinformatics and structural biology tools for the generation of epitope-based vaccines, that included the antigenic parts only and demonstrated a promising ability for fighting against pathogenic microorganisms, represented a large percentage of the whole studies directed to the vaccine development (70).

The development of an effective vaccine against the several pathotypes of *E. coli* has faced many obstacles. The complex nature of this bacteria and its genetic plasticity hinders the trials for vaccine development. Additionally, lack of broadly applicable testing to assess disease burden, particularly in remote areas where incidence may be quite high besides the occurrence of several bacterial pathotypes are another barriers (71). The current study tried to address these difficulties by analyzing the complete genome of *E. coli* to find common conserved vaccine candidates that cover different pathotypes then assess the characteristics of this potential vaccine computationally before moving to the future phase of wet lab validation. In our analysis, the reverse vaccinology approach was employed to generate a shortlist of potential vaccine candidates after analysis of *E. coli* complete proteome then immunoinformatics computational tools were applied for designing a multitpe vaccine based on filtered vaccine candidates. Only two outer membrane proteins (LptD and BamA) were chosen after applying the filtration steps on 5,155 proteins. The same approach has been reported to be successful in protein filtration and potential vaccine candidature with many pathogens, for instance; *Klebsiella Pneumoniae* (72), *Staphylococcus aureus* (19), *Mycobacterium tuberculosis* (73), *Shigella flexneri* (74), *Pseudomonas aeruginosa* (75), *Moraxella catarrhalis* (76), and Nipah virus (18). In addition to that, *E. coli* was one of the main pathogens that were analyzed for designing a vaccine through a bioinformatics approach in many previous studies. We can divide these studies into two main categories, firstly, studies that just predicted single B and T cell epitopes of protein candidates of certain pathotypes of *E. coli* such as Khan and Kumar (77) and Mehla and Ramana (78). Secondly, more advanced studies that designed and validated a multitope vaccine through a computational approach such as (79), where the vaccine was designed based on Intimin, Stx, Lt, and Cfa proteins and directed against ETEC and EHEC, Another study (80), the vaccine was designed based on IutA and FimH proteins and directed against UPEC, a third study (81), the vaccine was designed based on the bacterial type-3 secretion system and directed against extraintestinal pathogenic *E. coli*. In the current study, we followed the approach of the second group of studies but at the same time, we introduced two unique points that greatly affect the design and the potential application. First, of all, unlike mentioned studies that selected their vaccine candidates based on literature mentioned virulent proteins of specific *E. coli* pathotype, we nominated the vaccine candidates of the current study after a complete filtration of a reference pathogenic *E. coli* strain where the filtration steps represented a basic portion of the study to select proteins that are antigenic, virulent, with a low similarity percentage with human proteins, essential, and conserved throughout various pathotypes of *E. coli*. Consequently, we came up with two novel vaccine candidates (BamA and LptD), and therefore the designed multitope vaccine based on these proteins would have different sequences and characteristics from previously discussed ones. The criteria of protein conservation moved us to the second unique point in the current study which is cross-reactivity. As

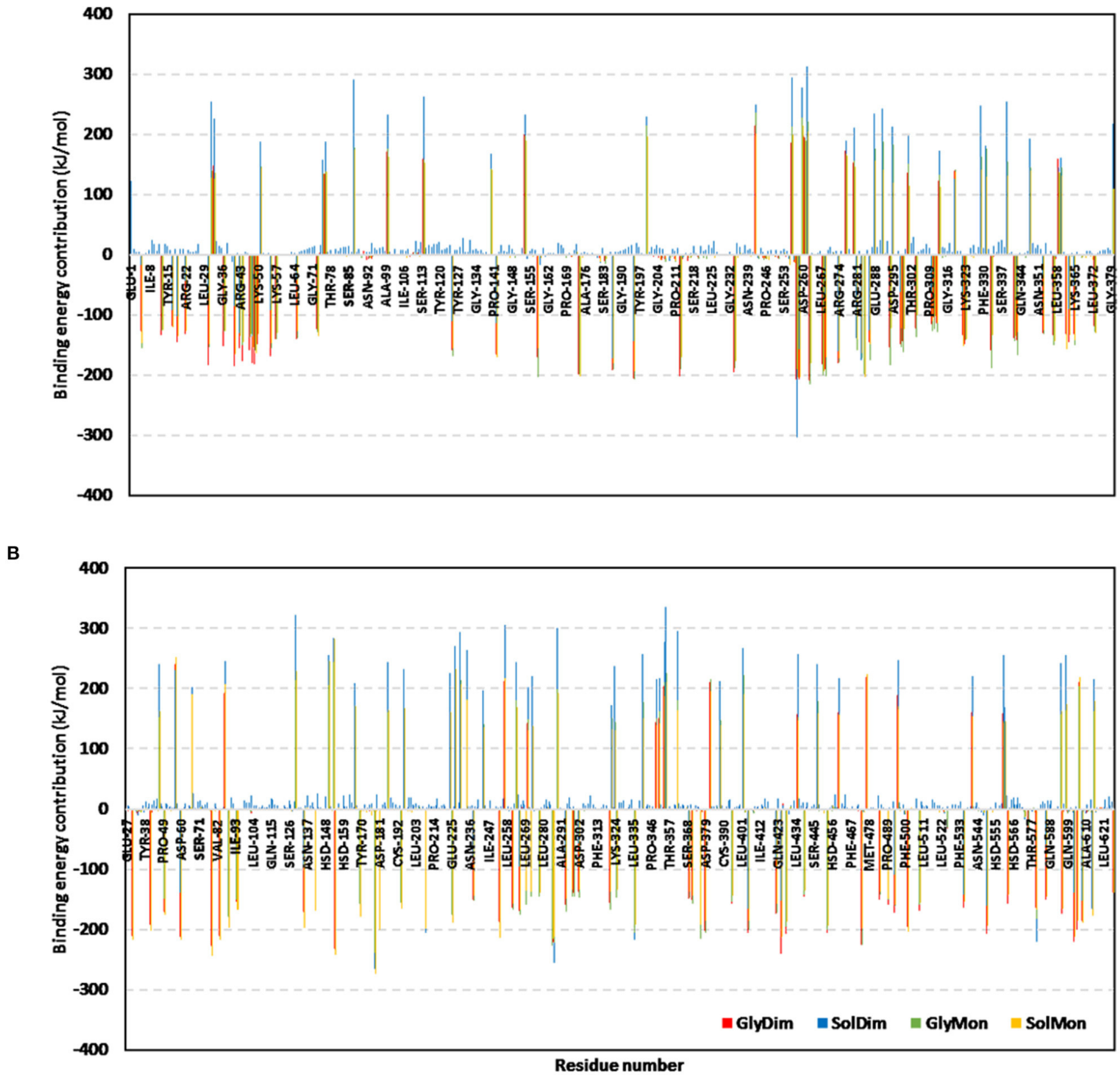


FIGURE 10 | Residue-wise binding-free energy decomposition for the simulated epitope vaccine-*h*TLR-4 complexes. The energy contributions in kJ/mol within the total free binding energies for the residues comprising the simulated **(A)** epitope vaccine; **(B)** *h*TLR-4 target proteins. Energy contributions are represented against the residue number.

mentioned, previous trials targeted only specific *E. coli* pathotype in their vaccine design while in this study we targeted essential proteins with a high percentage of conservation in basic *E. coli* pathotypes thus the predicted vaccine would have a potential activity against different *E. coli* pathotypes. Based on what mentioned above, we can choose the steps of common conserved protein candidates selection for several *E. coli* pathotypes and defining the prominent epitopes of these candidates as the most important steps that shaped the novelty of the current potential vaccine construct.

As mentioned above, the process of vaccine design through the immunoinformatics approach has been applied against several pathogens with a common methodology. On the other hand, with the continuous development in computational tools, a recent study (82) proposed a multitope vaccine against COVID-19 through the integration of a deep learning approach for

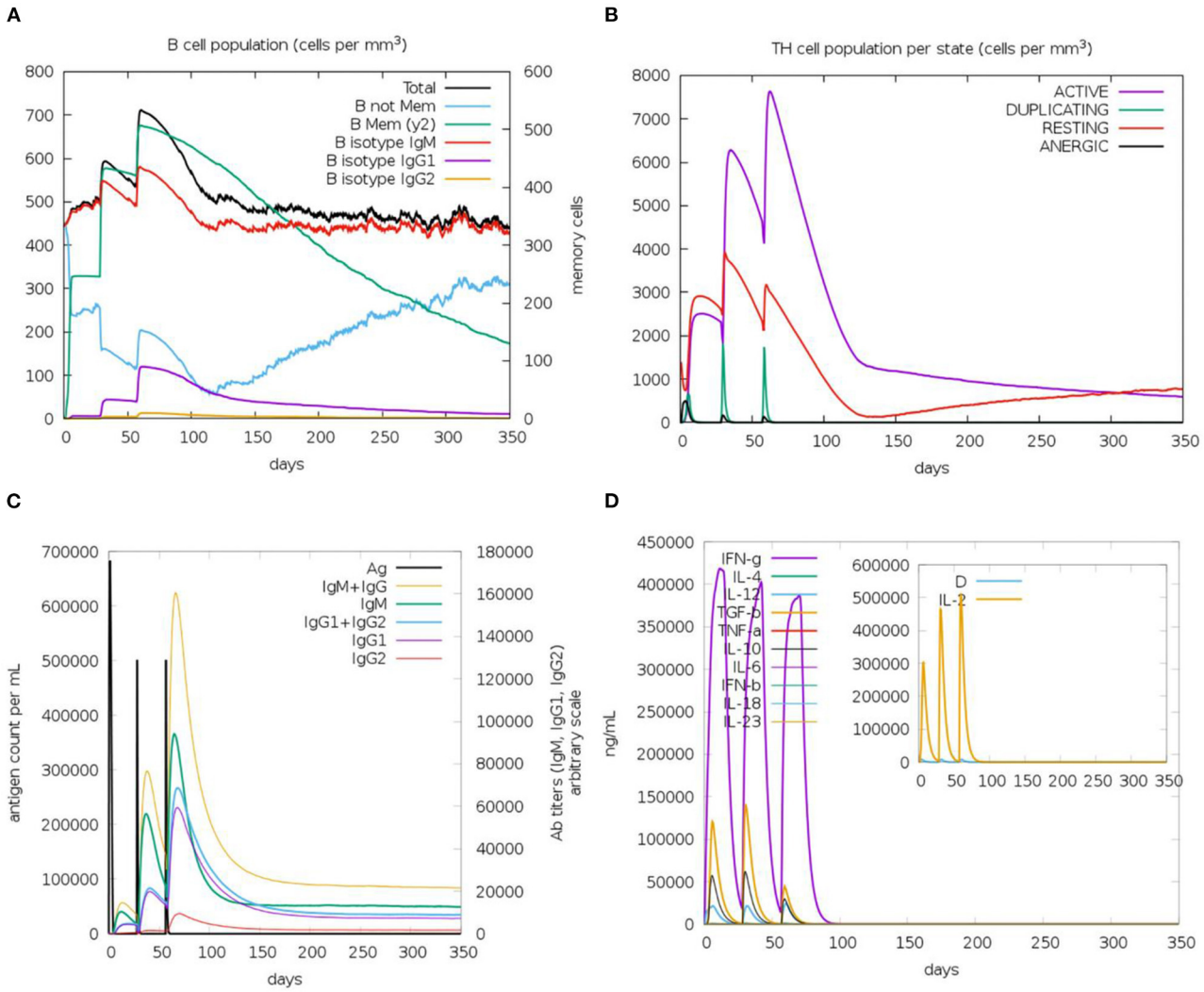


FIGURE 11 | The predicted immune response after designed vaccine injection. **(A,B)** shows the population of B and T cells, respectively, while **(C,D)** demonstrate antibody and cytokine count in a response to vaccine administration.

prediction and design. In that study, DeepVacPred was employed to conduct multiple predictions in a time-saving manner. The current study followed the former methodology, especially that the designed vaccines by this common methodology showed promising results during the *in-vivo* validation (80, 83) and as mentioned above the main focus of this study was the development of a potential vaccine that covers most *E. coli* pathotypes and deeply investigate the characteristics of this construct more than developing novel prediction methodologies.

In the current study, several online servers were used to identify potential vaccine candidates against *E. coli* pathotypes where the nominated proteins were found to be highly conserved in the majority of *E. coli* pathotypes. In addition to that, they were outer membrane proteins, essential for bacterial survival, had a high antigenicity score, and were non-homologous to human proteins to ensure their safety in clinical trials. Bacterial outer membrane proteins play an important role in molecule transporting, membrane integrity maintenance, in addition to pathogenesis (84). Moreover, their easy accession to the host immune system gives them the advantage to be highly potential candidates for vaccine development (85). Lipopolysaccharide (LPS) is a major structural component in most Gram-negative bacteria and it is essential for bacterial growth, LptD is one of eight proteins involved in the proper assembly of LPS after its biosynthesis in *E. coli* (86). Regarding BamA, it belongs to the Omp85 family, which is characterized as a major antigenic and immunogenic protein expressed by most Gram-negative pathogenic bacteria (87).

Vaccine construction using mapped epitopes is a sophisticated strategy to trigger an immune response against infectious agents (88). On the other hand, reliance on peptide vaccines for human usage has faced some limitations as single peptide epitopes were

found to be not strong enough to stimulate a suitable and prolonged immune response and exhibited low immunogenicity in comparison with live attenuated vaccines. Additionally, there are some questions about the stability of these peptides to induce the immune response before getting lysed by human proteolytic enzymes (89). To overcome these limitations, the current study proposed vaccine was constructed based on a combination of peptides with suitable linkers where those peptides were filtered through several criteria, and only conserved, highly antigenic, non-allergen, and non-toxic epitopes were selected to design the multitope vaccine. multi-epitope vaccines are considered a better choice than monovalent ones as they can stimulate efficient humoral as well as cellular immune responses (90). In addition to these epitopes, other components were added to the final vaccine construct to improve the immune response for this potential vaccine. β-defensin adjuvant was incorporated to create a deposit of the antigenic compound at the site of the vaccine that is steadily released over time, elongating the robust immune response and overcoming one of the main peptide vaccine limitations (91). Moreover, suitable linkers were used to join the selected epitopes from each candidate protein where they provide effective separation between the epitopes (92). Firstly, EAAAK was employed to improve the bi-functional catalytic activity, give stiffness in addition to enhancing fusion protein stability (93). The second linker, GPGPG, was selected for its ability to induce HTL immune response and the ability to break the junctional immunogenicity, resulting in individual epitopes' restoration of immunogenicity (94). The final linker, KK, was employed because of its ability to bring the pH value close to the physiological range (95). In addition to the epitopes, linkers, and adjuvant, the PADRE sequence was also added as it has been revealed that this sequence could minimize the polymorphism of HLA molecules in the population (96). Analyzing the results of the multitope vaccine assessment showed that, the current potential vaccine construct is antigenic with an antigenicity score of 1.07 which was more than the corresponding scores of the designed multitope vaccines against severe acute respiratory syndrome coronavirus 2 (97), Marburg virus (94), and *Leptospira* (98). Regarding the population coverage, the world coverage showed 100% for the constructed vaccine and that was similar to the corresponding coverage reported in (99). For the binding affinity of the single CTLs to the representative alleles, the scores ranged between −7.6 and −8.4, which was smaller than the range reported in (100) but at the same time was large enough when compared to the control binding score of the current study. Moving to the binding energy of the whole multitope vaccine, we reported here a score of −1420.9 kcal/mol, which was smaller than corresponding scores of vaccines predicted with a similar approach (18, 69, 76), implying that a strong binding with TLR would occur. Collectively, The final construct exhibited promising physicochemical, immunological, and chemical characteristics when assessed computationally where molecular dynamics simulation studies were adopted to give a close view of the behavior of this potential vaccine with the receptors of the immune system.

Finally, the designed epitope vaccine showed significant confinement and stability within the hTLR-4 binding site throughout the conducted 100 ns explicit molecular dynamic runs. With preferentiality for the dimeric glycosylated target protein, the designed vaccine exhibited steady conventional thermodynamic behavior with $C\alpha$-RMSDs leveling up for more than 70 ns. The adopted $C\alpha$-RMSD analytical tool allows the estimation of molecular deviation from the designated original/reference structure, the thing that can be used for ensuring ligand-target stability/confinement as well as the validity of the MD protocol (101). Obtaining $C\alpha$-RMSD at low values as well as being rapidly equilibrated has been correlated with the strong affinity of the designed vaccine ligand against target protein as well as the successful convergence of the simulated models requiring no further molecular dynamics simulation runs (102). This $C\alpha$-RMSD-based vaccine-*h*TLR-4 stability was comparable to the stability of several reported proteinaceous multitope vaccines targeting different microorganisms TLRs (103, 104). Regarding the obtained residue-wise fluctuation analysis, the higher $C\alpha$-RMSF values of the epitope vaccine as compared to *h*TLR-4 can be reasonably correlated to their differential tertiary structure folding and/or packing. Generally, the $C\alpha$-RMSF flexibility analysis tool estimates the averaged deviations of protein's residues in relation to their reference positions, the thing that would provide a valuable evaluation of protein's residues regarding their respective dynamic behavior being represented through flexibility and fluctuation (105). In these regards, the incorporation of long α-helices with flexible β-loop connections within the vaccine's designed structures would rationalize the initial relaxation and significant convergence into more stable compacted conformations. On the other hand, *h*TLR-4 exhibited densely packed shoe-like architecture with plenty of highly ordered parallel β-sheets. This differential inherited flexibility was also highlighted through the 3D-conformational analysis between initial and last frames as well as the adopted dihedral coordinate-based normal mode analysis. Having non-uniform stiffness/immobility profiles, as well as high B-factor, deformability, and mobility indices, conferred the profound flexibility being assigned for the vaccine (50). The assigned high immobility profiles for the epitope vaccine were most recognized through its carboxy-terminal amino acids and vicinal residues as being clearly demonstrated through the conformational analysis as well as RMSF, Rg, and SASA findings.

It is worth mentioning that the high RMSF flexibility of the 301–380 residue range was highly reasoned since these residues started the MD simulation being apart from the *h*TLR-4 interface and then ended being near the hTLR-4 lateral side. Nevertheless, the epitope vaccine rapidly attained a more stable compacted conformation/orientation in relation to the bound *h*TLR-4 as the MD runs proceeded. The latter more profoundly stable conformation/orientation of the vaccine was mostly related to the movement of this residue range toward *h*TLR-4 lateral side representing the dimerization interface. This was confirmed through the Rg and SASA analysis where values significantly dropped as the MD simulation proceeded reaching to lower steady trajectories for more than half of the MD simulation timeframes. Generally, Rg accounts for the global stability of either ligand/protein ternary structures, where such stability

parameter is defined as the mass-weighted RMSD for atom groups in relation to their respective common center of mass (106). Thus, the depicted dynamic behavior of simulated vaccine to exhbit low Rgs maintaining a plateau around an average value conferred significant stability/compactness at the *h*TLR-4 binding. Notably, the Rg finding further highlighted the preferential vaccine anchoring at the dimeric and/or glycosylated *h*TLR-4 states since the monomeric and/or non-glycosylated ones achieved higher fluctuations and late Rg equilibration trajectories (not before 70ns) suggesting non-optimal compactness and intermolecular binding around these timeframes. Findings from the ligand's SASA analysis came in good agreement with the above preferential complex stability since the simulated vaccine showed steadier SASA tone along the average equilibration plateau around 40 ns and till the simulation end. Since SASA is a quantitative measurement about the extent of protein/solvent interaction correlating for molecular surface area assessable to solvent, thus, low SASA tones imply relative structural shrinkage under the impact of the solvent surface charges yielding more compact and stable conformations (107). The latter findings were also consistent with the vaccine's non-polar solvation energy (only SASA-model) across the MD simulation runs. It is worth mentioning that all above epitope-*h*TLR-4 flexibility patterns were also similarly depicted within several reported studies investigating the potential binding affinity of peptide-based vaccines toward microbial TLRs (103, 104).

The presented study further highlights the impact of *h*TLR-4 oligomerization as well as oligosaccharide states on vaccine binding. Depicting lower *Cα*-RMSD and RMSF values with more steady tones at glycosylated dimeric model raised the suggestion that *N*-glycosylation and *h*TLR-4 dimerization were beneficial for vaccine anchoring at the target binding site. Accumulated evidence has illustrated the importance of TLR ectodomains' *N*-glycosylation for orchestrating the localization and signaling capacity (108). Additionally, *N*-linked glycosylation (sialylation) of *h*TLR-4 and its coreceptor, MD-2, enhances the lipoprotein-driven nuclear factor kappa-B activation, cytokine expressions, and, as well as regulates *h*TLR-2 and *h*TLR-3 signaling pathway (109–111). Furthermore, sialylated residues are important for mediating the association between *h*TLR-4 and MD-2, enhancing *h*TLR-4 dimerization, as well as the assembly of complete TLRs signaling complexes (108, 112). Thus, having the importance in enhancing ligand anchoring at TLRs' binding sites as well as facilitating TLRs dimerization it was highly reasoned why highly stabilized/steady thermodynamic behaviors, as well as less fluctuating/, mobilized residues were assigned for both glycosylated dimeric *h*TLR-4/vaccine complex. The MM/PBSA-driven free binding energy calculations also highlighted the higher affinity of the simulated vaccine toward the glycosylated *h*TLR-4 in relation to those being non-glycosylated. Depicting higher negative total free binding energies as well as more pronounced Coloumb's electrostatic energy contributions were highlighted for the more favored vaccine anchoring/affinity toward the glycosylated target proteins. Binding to the *N*-glycosylation chains was also found satisfactory to counterbalance the predicted electrostatic penalties and solvation energies during epitope vaccine ligand binding since lower polar solvation energy terms ($\Delta G_{\text{Solvation}}$) were depicted at the glycosylated models. This was also confirmed through monitoring the non-polar solvation energy *via* the only SASA-model across the MD simulation runs.

Generally, solvation energy terms represent significant repulsive forces against the ligand-binding since binding is a solvent displacement process. It was depicted that these large repulsive forces were mediated majorly by the *h*TLR-4 residues rather than by the vaccine amino acids as being depicted within the residue-wise energy contributions the thing that could be related to the high ordered water molecules at the hydrophobic surface of the TLR-2 ligand-binding site. Thus, the presence of *N*-linked glycosylation chains would minimize such repulsive penalties against the vaccine binding. On the other hand, the total non-polar interactions ($\Delta G_{\text{van der Waal}}$ plus ΔG_{SASA}) were shown to be higher at the non-glycosylated models conferring their respective larger surface area as well as higher hydrophobic potentialities toward vaccine binding. The latter was rationalized since accumulated evidence has considered the general TLRs binding site to be extended and more hydrophobic in nature (52–55, 113–115). Based on the above evidence, it was satisfactory that the designed epitope vaccine depicted significant binding affinity toward the *h*TLR-4 binding pocket with higher preferentiality toward the glycosylated dimeric state.

CONCLUSION

The current study aimed to demonstrate the role of modern approaches for vaccine development as a potential solution to fight resistant pathogens. Here, we reported two proteins namely BamA and LptD, after the filtration of the whole proteome of *E. coli* reference strain, to act as a base for multitope vaccine construct against *E. coli* pathotypes. The multitope construct included top-ranked epitopes of the filtered proteins in addition to beta-defensin and PADRE peptide. The molecular modeling simulation studies illustrated relevant affinity/binding of the designed epitope vaccine toward the *h*TLR-4 binding pocket, yet with higher preferentiality toward the glycosylated dimeric state. Finally, The predicted physicochemical and immunological characteristics of the constructed vaccine nominated it as a potential solution against several *E. coli* pathotypes and recommended its movement to wet lab validation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MS, MB, MSA, MA, and KD: conceptualization, methodology, and original draft preparation. SA, EF, SE, AL, and RE: writing—review and editing. MS and MA: supervision and

project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Taif University Researchers Supporting Project number (TURSP-2020/202), Taif University, Taif, Saudi Arabia.

ACKNOWLEDGMENTS

We would like to express our great appreciation to the Deanship of Scientific Research, Taif University, Taif, Kingdom of Saudi Arabia for funding our project through Taif University Researchers Supporting Project number (TURSP-2020/202). The simulations in this work were performed at King Abdulaziz University's High-Performance Computing Center (Aziz Supercomputer) (http://hpc.kau.edu.sa), the authors, therefore, acknowledge with thanks the center for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.829467/full#supplementary-material

Supplementary Figure 1 | Predicted positions of MHC-I peptides (yellow color) in the 3D structure of HLA-B*44:03 receptor (red color), structures **(A–F)** are for epitopes number 1, 2, 3, 4, 5, and 6, respectively, from **Supplementary Table 3**.

Supplementary Figure 2 | Predicted positions of MHC-II peptides (yellow color) in the 3D structure of HLA-DRB1*04:01 receptor (red color), structures **(A–F)** are for epitopes number 1, 2, 3, 4, 5, and 6, respectively, from **Supplementary Table 3**.

Supplementary Figure 3 | Graphical map for the designed multitope vaccine.

Supplementary Figure 4 | The 3D model of the 9 predicted conformational B-cell epitopes in the refined final vaccine construct where the yellow residues represent the epitopes and the gray ones are for the rest of the predicted vaccine. Letters from A to I represents the predicted epitopes in **Supplementary Table 7** from 1 to 9, respectively.

Supplementary Figure 5 | Global stability of the epitope vaccine at the four simulated models across 100 ns explicit molecular dynamics simulation runs. Time-evolution of **(A)** Rg, **(B)** SASA, along molecular dynamics timeframes (ns).

Supplementary Figure 6 | Non-polar solvation energy terms using the SASA-only model calculation for the four investigated epitope vaccine ligand-bound hTLR-4 target models across 100 ns explicit molecular dynamics simulation runs. Time-evolution of **(A)** Proteins, **(B)** Ligands, and **(C)** Complexes, along molecular dynamics timeframes (ns).

Supplementary Figure 7 | Residue-wise binding-free energy decomposition for the simulated epitope vaccine-hTLR-4 complexes. 3D-representation (Cartoon) for the regions implying favored binding interactions on the initial/reference vaccine/hTLR-4 complex. Glycosylated dimeric, non-glycosylated dimeric, glycosylated monomeric, and non-glycosylated monomeric states are at upper-left, upper-right, lower-left, and lower right quadrant, respectively (Only one protomer is shown for clarity). Regions of the proteins are in spectrum colors from dark blue (negative-valued ΔG kJ/mol conferring highly favored attractive binding forces) down to dark red (positive-valued ΔG kJ/mol conferring unfavored high repulsive binding forces).

Supplementary Table 1 | Characteristics of Potential Vaccine Candidates for *E. coli* O157:H7 str. Sakai.

Supplementary Table 2 | Protein conservancy of the selected candidates among *E. coli* pathotypes.

Supplementary Table 3 | Selected epitopes' binding energies with representative MHC-I and MHC-II alleles.

Supplementary Table 4 | Reacting alleles and population coverage of filtered CLTs and HTLs.

Supplementary Table 5 | Population coverage percentage of the multitope vaccine and its constructing peptides (CTLs and HTLs).

Supplementary Table 6 | Predicted physicochemical properties of the chimeric vaccine.

Supplementary Table 7 | The predicted conformational B cell epitopes from the refined vaccine 3D construct.

REFERENCES

1. Clements A, Young JC, Constantinou N, Frankel G. Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes.* (2012) 3:71–87. doi: 10.4161/gmic.19182
2. Gould D. Causes, prevention and treatment of *Escherichia coli* infections. *Nurs Stand.* (2010) 24:50–6. doi: 10.7748/ns.24.31.50.s46
3. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in molecular serotyping and subtyping of *Escherichia coli*. *Front Microbiol.* (2016) 7:644. doi: 10.3389/fmicb.2016.00644
4. Suardana IW, Artama WT, Widiasih DA, Ngurah IG, Mahardika K. Genetic Diversity of *Escherichia coli* O157:H7 strains using random amplified polymorphic DNA (RAPD). *Int Res J Microbiol.* (2013) 55:671–4.
5. Puzari M, Sharma M, Chetia P. Emergence of antibiotic resistant Shigella species: A matter of concern. *J Infect Public Health.* (2018) 11:451–4. doi: 10.1016/j.jiph.2017.09.025
6. Yang SC, Lin CH, Aljuffali IA, Fang JY. Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. *Arch Microbiol.* (2017) 199:811–25. doi: 10.1007/s00203-017-1393-y
7. Rasheed MU, Thajuddin N, Ahamed P, Teklemariam Z, Jamil K. Antimicrobial drug resistance in strains of *Escherichia coli* isolated from food sources. *Rev Inst Med Trop São Paulo.* (2014) 56:341–6. doi: 10.1590/S0036-46652014000400012
8. Collignon P. Resistant *Escherichia coli*—we are what we eat. *Clin Infect Dis.* (2009) 49:202–4. doi: 10.1086/599831
9. Bourgeois AL, Wierzba TF, Walker RI. Status of vaccine research and development for enterotoxigenic *Escherichia coli*. *Vaccine.* (2016) 34:2880–6. doi: 10.1016/j.vaccine.2016.02.076
10. Mejías MP, Hiriart Y, Lauché C, Fernández-Brando RJ, Pardo R, Bruballa A, et al. Development of camelid single chain antibodies against Shiga toxin type 2 (Stx2) with therapeutic potential against Hemolytic Uremic Syndrome (HUS). *Sci Rep.* (2016) 6:24913. doi: 10.1038/srep24913
11. Ferreira Oliveira A, Almeida Cardoso S. Bruno dos Reis Almeida F, Licursi de Oliveira L, Pitondo-Silva A, Gomes Soares S, Seixas Hanna E. Oral immunization with attenuated Salmonella vaccine expressing Escherichia coli O157: H7 intimin gamma triggers both systemic and mucosal humoral immunity in mice. *Microbiol Immunol.* (2012) 56:513–22. doi: 10.1111/j.1348-0421.2012.00477.x
12. Ahmed A, Li J, Shiloach Y, Bobbins JB, Szu SC. Safety and immunogenicity of *Escherichia coli* O157 O-specific polysaccharide conjugate vaccine in 2-5-year-old children. *J Infect Dis.* (2006) 193:515–21. doi: 10.1086/499821
13. Harris JA, Roy K, Woo-Rasberry V, Hamilton DJ, Kansal R, Qadri F, et al. Directed evaluation of enterotoxigenic *Escherichia coli* autotransporter proteins as putative vaccine candidates. *PLoS Negl Trop Dis.* (2011) 5:e1428. doi: 10.1371/journal.pntd.0001428

14. Zhang C, Iqbal J, Gómez-Duarte OG. Murine immunization with CS21 pili or LngA major subunit of enterotoxigenic *Escherichia coli* (ETEC) elicits systemic and mucosal immune responses and inhibits ETEC gut colonization. *Vet Microbiol.* (2017) 202:90–100. doi: 10.1016/j.vetmic.2016.02.001
15. Huttner A, Hatz C, van den Dobbelsteen G, Abbanat D, Hornacek A, Frölich R, et al. Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic *Escherichia coli* in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. *Lancet Infect Dis.* (2017) 17:528–37. doi: 10.1016/S1473-3099(17)30108-1
16. Habibi M, AsadiKaram MR, Bouzari S. Evaluation of prevalence, immunogenicity and efficacy of FyuA iron receptor in uropathogenic *Escherichia coli* isolates as a vaccine target against urinary tract infection. *Microb Pathog.* (2017) 110:477–83. doi: 10.1016/j.micpath.2017.07.037
17. Rappuoli R, Bottomley MJ, D'Oro U, Finco O. De Gregorio E. Reverse vaccinology 20: Human immunology instructs vaccine antigen design. *J Exp Med.* (2016) 213:469–81. doi: 10.1084/jem.20151960
18. Soltan MA, Eldeen MA, Elbassiouny N, Mohamed I, El-damasy DA, Fayad E, Ali OAA, Raafat N, Eid RA, Al-karmalawy AA. Proteome based approach defines candidates for designing a multitope vaccine against the nipah virus. *Int J Mol Sci*. (2021) 22:9330. doi: 10.3390/ijms22179330
19. Soltan MA, Magdy D, Solyman SM, Hanora A. Design of staphylococcus aureus new vaccine candidates with B and T cell epitope mapping, reverse vaccinology, and immunoinformatics. *Omi A J Integr Biol.* (2020) 24:195–204. doi: 10.1089/omi.2019.0183
20. Soltan MA, Eldeen MA, Elbassiouny N, Kamel HL, Abdelraheem KM, El-Gayyed HA, et al. In silico designing of a multitope vaccine against rhizopus microsporus with potential activity against other mucormycosis causing fungi. *Cells.* (2021) 10:3014. doi: 10.3390/cells10113014
21. Lu T, Seo H, Moxley RA, Zhang W. Mapping the neutralizing epitopes of F18 fimbrial adhesin subunit FedF of enterotoxigenic *Escherichia coli* (ETEC). *Vet Microbiol.* (2019) 230:171–7. doi: 10.1016/j.vetmic.2019.02.015
22. Carvalho TF, Haddad JPA, Paixão TA, Santos RL. Meta-analysis and advancement of brucellosis vaccinology. *PLoS ONE.* (2016) 11:e0166582. doi: 10.1371/journal.pone.0166582
23. Martínez-Carranza E, Barajas H, Alcaraz LD, Servín-González L, Ponce-Soto GY, Soberón-Chávez G. Variability of bacterial essential genes among closely related bacteria: the case of *Escherichia coli*. *Front Microbiol.* (2018) 9:1059. doi: 10.3389/fmicb.2018.01059
24. Wen QF, Liu S, Dong C, Guo HX, Gao YZ, Guo FB. Geptop 20: an updated, more precise, and faster Geptop server for identification of prokaryotic essential genes. *Front Microbiol.* (2019) 10:1236. doi: 10.3389/fmicb.2019.01236
25. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 30: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* (2010) 26:1608–15. doi: 10.1093/bioinformatics/btq249
26. Garg A, Gupta D. VirulentPred: A SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinform.* (2008) 9:62. doi: 10.1186/1471-2105-9-62
27. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* (2001) 305:567–80. doi: 10.1006/jmbi.2000.4315
28. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* (1999) 112:531–52. doi: 10.1385/1-59259-584-7:531
29. Vishnu US, Sankarasubramanian J, Gunasekaran P, Rajendhran J. Novel vaccine candidates against brucella melitensis identified through reverse vaccinology approach. *Omi A J Integr Biol.* (2015) 19:722–9. doi: 10.1089/omi.2015.0105
30. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* (2007) 8:4. doi: 10.1186/1471-2105-8-4
31. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol.* (2017) 8:278. doi: 10.3389/fimmu.2017.00278
32. Weiskopf D, Angelo MA, De Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci USA.* (2013) 110:E2046–53. doi: 10.1073/pnas.1305227110
33. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics.* (2011) 63:325–35. doi: 10.1007/s00251-011-0513-0
34. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J Chem Theory Comput.* (2014) 10:4745–58. doi: 10.1021/ct500592m
35. Oleg T, Arthur JO. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* (2012) 31:455–61.
36. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform.* (2007) 8:361. doi: 10.1186/1471-2105-8-361
37. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. *Infect Genet Evol.* (2017) 48:83–94. doi: 10.1016/j.meegid.2016.12.010
38. Magnan CN, Randall A, Baldi P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics.* (2009) 25:2200–7. doi: 10.1093/bioinformatics/btp386
39. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* (1999) 292:195–202. doi: 10.1006/jmbi.1999.3091
40. Cheng J, Randall AZ, Sweredoski MJ, Baldi P, SCRATCH. A protein structure and structural feature prediction server. *Nucleic Acids Res.* (2005) 33:W72–6. doi: 10.1093/nar/gki396
41. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* (2013) 41:W384–8. doi: 10.1093/nar/gkt458
42. Hasan M, Hasan R, Hazrat Ali M, Fazle Rabbee M, Al-Hakim B, Mohammad Rejwan H, et al. In-silico characterization and homology modeling of catechol 1,2 dioxygenase involved in processing of catechol-an intermediate of aromatic compound degradation pathway. *Glob J Sci Front Res*. (2015)
43. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* (2007) 35:W407–10. doi: 10.1093/nar/gkm290
44. Ponomarenko J, Bui HH Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* (2008) 9:514. doi: 10.1186/1471-2105-9-514
45. Craig DB. Dombkowski AA. Disulfide by Design 20: A web-based tool for disulfide engineering in proteins. *BMC Bioinform.* (2013) 14:346. doi: 10.1186/1471-2105-14-346
46. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Sci Rep.* (2018) 8:9044. doi: 10.1038/s41598-018-26689-7
47. Finamore A, Roselli M, Imbinto A, Seeboth J, Oswald IP, Mengheri E. Lactobacillus amylovorus inhibits the TLR4 inflammatory signaling triggered by enterotoxigenic *Escherichia coli* via modulation of the negative regulators and involvement of TLR2 in intestinal caco-2 cells and pig explants. *PLoS ONE.* (2014) 9:e94891. doi: 10.1371/journal.pone.0094891
48. 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. doi: 10.1038/nprot.2016.169
49. Awan FM, Obaid A, Ikram A, Janjua HA. Mutation-structure-function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): a comprehensive informatics approach. *Int J Mol Sci.* (2017) 18:139. doi: 10.3390/ijms18010139

50. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P, IMODS. Internal coordinates normal mode analysis server. *Nucleic Acids Res.* (2014) 42:W271–6. doi: 10.1093/nar/gku339
51. Pandey RK, Verma P, Sharma D, Bhatt TK, Sundar S, Prajapati VK. High-throughput virtual screening and quantum mechanics approach to develop imipramine analogues as leads against trypanothione reductase of leishmania. *Biomed Pharmacother.* (2016) 83:141–52. doi: 10.1016/j.biopha.2016.06.010
52. Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. *Biochem Biophys Res Commun.* (2009) 388:621–5. doi: 10.1016/j.bbrc.2009.08.062
53. Jin MS, Lee JO. Structures of the toll-like receptor family and its ligand complexes. *Immunity.* (2008) 29:182–91. doi: 10.1016/j.immuni.2008.07.007
54. Ohto U, Yamakawa N, Akashi-Takamura S, Miyake K, Shimizu T. Structural analyses of human toll-like receptor 4 polymorphisms D299G and T399I. *J Biol Chem.* (2012) 287:40611–7. doi: 10.1074/jbc.M112.404608
55. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature.* (2009) 458:1191–5. doi: 10.1038/nature07830
56. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).*
57. Al-Karmalawy AA, Dahab MA, Metwaly AM, Elhady SS, Elkaeed EB, Eissa IH, et al. Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the hACE2 receptor. *Front Chem.* (2021) 9:661230. doi: 10.3389/fchem.2021.661230
58. Elmaaty AA, Darwish KM, Khattab M, Elhady SS, Salah M, Hamed MIA, et al. In a search for potential drug candidates for combating COVID-19: computational study revealed salvianolic acid B as a potential therapeutic targeting 3CLpro and spike proteins. *J Biomol Struct Dyn.* (2021) 30:1–28. doi: 10.1080/07391102.2021.1918256
59. Helal MA, Shouman S, Abdelwaly A, Elmehrath AO, Essawy M, Sayed SM, et al. Molecular basis of the potential interaction of SARS-CoV-2 spike protein to CD147 in COVID-19 associated-lymphopenia. *J Biomol Struct Dyn.* (2020) 16:1–11. doi: 10.1080/07391102.2020.1822208
60. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat Methods.* (2016) 14:71–3. doi: 10.1038/nmeth.4067
61. Saleh AH, Abdelwaly A, Darwish KM, Eissa AAHM, Chittiboyina A, Helal MA. Deciphering the molecular basis of the kappa opioid receptor selectivity: a molecular dynamics study. *J Mol Graph Model.* (2021) 106:107940. doi: 10.1016/j.jmgm.2021.107940
62. Ross GA, Rustenburg AS, Grinaway PB, Fass J, Chodera JD. Biomolecular Simulations under Realistic Macroscopic Salt Conditions. *J Phys Chem B.* (2018) 122:5466–86. doi: 10.1021/acs.jpcb.7b11734
63. Zaki AA, Ashour A, Elhady SS, Darwish KM, Al-Karmalawy AA. Calendulaglycoside A showing potential activity against SARS-CoV-2 main protease: molecular docking, molecular dynamics, and SAR studies. *J Tradit Complement Med.* (2021) (Online ahead of print) doi: 10.1016/j.jtcme.2021.05.001
64. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem.* (1997) 18:1463–72. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
65. Páll S, Hess B, A. flexible algorithm for calculating pair interactions on SIMD architectures. *Comput Phys Commun.* (2013) 184:2641–50. doi: 10.1016/j.cpc.2013.06.003
66. Kumari R, Kumar R, Consortium OSDD, Lynn A. g _ mmpbsa—A GROMACS tool for MM-PBSA and its optimization for high-throughput binding energy calculations. *J Chem Inf Model.* (2014) 54:1951–62. doi: 10.1021/ci500020m
67. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE.* (2010) 5:e9862. doi: 10.1371/journal.pone.0009862
68. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics.* (2009) 10:837–52. doi: 10.2217/pgs.09.25
69. Hasan M, Azim KF, Begum A, Khan NA, Shammi TS, Imran AS, et al. Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infect Genet Evol.* (2019) 70:140–57. doi: 10.1016/j.meegid.2019.03.003
70. Jiang P, Cai Y, Chen J, Ye X, Mao S, Zhu S, et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitopes vaccine in BALB/c mice model. *Vaccine.* (2017) 35:3096–103. doi: 10.1016/j.vaccine.2017.04.031
71. Fleckenstein JM. Confronting challenges to enterotoxigenic *Escherichia coli* vaccine development. *Front Trop Dis.* (2021) 20. doi: 10.3389/fitd.2021.709907
72. Dar HA, Zaheer T, Shehroz M, Ullah N, Naz K, Muhammad SA, et al. Immunoinformatics-aided design and evaluation of a potential multi-epitope vaccine against klebsiella pneumoniae. *Vaccines.* (2019) 7:88. doi: 10.3390/vaccines7030088
73. Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, et al. In silico predicted mycobacterial epitope elicits in vitro T-cell responses. *Mol Immunol.* (2014) 61:16–22. doi: 10.1016/j.molimm.2014.04.009
74. Leow CY, Kazi A, Ismail CMKH, Chuah C, Lim BH, Leow CH, et al. Reverse vaccinology approach for the identification and characterization of outer membrane proteins of shigella flexneri as potential cellular-and antibody-dependent vaccine candidates. *Clin Exp Vaccine Res.* (2020) 9:15–25. doi: 10.7774/cevr.2020.9.1.15
75. Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. *Sci Rep.* (2019) 9:5240. doi: 10.1038/s41598-019-41496-4
76. Soltan MA, Elbassiouny N, Gamal H, Elkaeed EB, Eid RA, Eldeen MA, et al. In Silico Prediction of a Multitope Vaccine against Moraxella catarrhalis: Reverse Vaccinology and Immunoinformatics. *Vaccines.* (2021) 9:669. doi: 10.3390/vaccines9060669
77. Khan F, Kumar A. An integrative docking and simulation-based approach towards the development of epitope-based vaccine against enterotoxigenic *Escherichia coli*. *Netw Model Anal Heal Informatics Bioinforma.* (2021) 10:11. doi: 10.1007/s13721-021-00287-6
78. Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a comparative genomics and immunoinformatics approach. *Mol Biosyst.* (2016) 12:890–901. doi: 10.1039/C5MB00745C
79. S. jeshvaghani F, K. Rahjerdi A, Amani J, Rad I, Jafari M, H. Salmanian A. Designing and structure evaluation of multi-epitope vaccine against ETEC and EHEC, an in silico approach. *Protein Pept Lett.* (2015) 23:33–42. doi: 10.2174/0929866522666151026122116
80. Hasanzadeh S, Habibi M, Shokrgozar MA, Ahangari Cohan R, Ahmadi K, Asadi Karam MR, et al. In silico analysis and *in vivo* assessment of a novel epitope-based vaccine candidate against uropathogenic *Escherichia coli*. *Sci Rep.* (2020) 10:16258. doi: 10.1038/s41598-020-73179-w
81. Wieser A, Magistro G, Nörenberg D, Hoffmann C, Schubert S. First multi-epitope subunit vaccine against extraintestinal pathogenic *Escherichia coli* delivered by a bacterial type-3 secretion system (T3SS). *Int J Med Microbiol.* (2012) 302:10–8. doi: 10.1016/j.ijmm.2011.09.012
82. Yang Z, Bogdan P, Nazarian S. An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Sci Rep.* (2021) 11:3238. doi: 10.1038/s41598-021-81749-9
83. Muhammad SA, Zafar S, Rizvi SZ, Imran I, Munir F, Jamshed MB, et al. Experimental analysis of T cell epitopes for designing liver cancer vaccine predicted by system-level immunoinformatics approach. *Am J Physiol Gastrointest Liver Physiol.* (2020) 318:G1055–69. doi: 10.1152/ajpgi.00068.2020
84. Lin J, Huang S, Zhang Q. Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes Infect.* (2002) 4:325–31. doi: 10.1016/S1286-4579(02)01545-9
85. Rizwan M, Naz A, Ahmad J, Naz K, Obaid A, Parveen T, et al. VacSol: a high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinform.* (2017) 18:106. doi: 10.1186/s12859-017-1540-0

86. Okuda S, Sherman DJ, Silhavy TJ, Ruiz N, Kahne D. Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nat Rev Microbiol.* (2016) 14:337–45. doi: 10.1038/nrmicro.2016.25
87. Su YC, Wan KL, Mohamed R, Nathan S. Immunization with the recombinant Burkholderia pseudomallei outer membrane protein Omp85 induces protective immunity in mice. *Vaccine.* (2010) 28:5005–11. doi: 10.1016/j.vaccine.2010.05.022
88. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, et al. Proof of principle for epitope-focused vaccine design. *Nature.* (2014) 507:201–6. doi: 10.1038/nature12966
89. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. *Chem Sci.* (2016) 7:842–54. doi: 10.1039/C5SC03892H
90. Amanna IJ, Slifka MK. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. *Virology.* (2011) 411:206–15. doi: 10.1016/j.virol.2010.12.016
91. Gupta A, Rosato AJ, Cui F. Vaccine candidate designed against carcinoembryonic antigen-related cell adhesion molecules using immunoinformatics tools. *J Biomol Struct Dyn.* (2021) doi: 10.1080/07391102.2020.1797539
92. Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, et al. In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum Vaccines Immunother.* (2015) 39:6084–98. doi: 10.1080/21645515.2015.1012017
93. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. *Adv Drug Deliv Rev.* (2013) 65:1357–69. doi: 10.1016/j.addr.2012.09.039
94. Sami SA, Marma KKS, Mahmud S, Khan MAN, Albogami S, El-Shehawi AM, et al. Designing of a multi-epitope vaccine against the structural proteins of marburg virus exploiting the immunoinformatics approach. *ACS Omega.* (2021) 6:32043–71. doi: 10.1021/acsomega.1c04817
95. Gu Y, Sun X, Li B, Huang J, Zhan B, Zhu X. Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against Trichinella spiralis infection in mice. *Front Microbiol.* (2017) 8:1475. doi: 10.3389/fmicb.2017.01475
96. 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. doi: 10.1371/journal.pone.0142563
97. Obaidullah AJ, Alanazi MM, Alsaif NA, Albassam H, Almehizia AA, Alqahtani AM, et al. Bin. Immunoinformatics-guided design of a multi-epitope vaccine based on the structural proteins of severe acute respiratory syndrome coronavirus 2. *RSC Adv.* (2021) 11:18103–21. doi: 10.1039/D1RA02885E
98. Kumar P, Lata S, Shankar UN, Akif M. Immunoinformatics-based designing of a multi-Epitope chimeric vaccine from multi-domain outer surface antigens of leptospira. *Front Immunol.* (2021) 12:735373. doi: 10.3389/fimmu.2021.735373
99. Mahmud S, Rafi MO, Paul GK, Promi MM, Shimu MSS, Biswas S, et al. Bin, Dhama K, Alyami SA, Moni MA, et al. Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Sci Rep.* (2021) 11:15431. doi: 10.1038/s41598-021-92176-1
100. Rakib A, Sami SA, Mimi NJ, Chowdhury MM, Eva TA, Nainu F, et al. Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Comput Biol Med.* (2020) 124:103967. doi: 10.1016/j.compbiomed.2020.103967
101. Arnittali M, Rissanou AN, Harmandaris V. Structure of biomolecules through molecular dynamics simulations. *Procedia Comput Sci.* (2019) 156:69–78. doi: 10.1016/j.procs.2019.08.181
102. Liu K, Watanabe E, Kokubo H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J Comput Aided Mol Des.* (2017) 31:201–11. doi: 10.1007/s10822-016-0005-2
103. Chauhan V, Rungta T, Goyal K, Singh MP. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci Rep.* (2019) 9:2517. doi: 10.1038/s41598-019-39299-8
104. Khatoon N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep.* (2017) 7:8285. doi: 10.1038/s41598-017-08842-w
105. Benson NC, Daggett V, A. comparison of multiscale methods for the analysis of molecular dynamics simulations. *J Phys Chem B.* (2012) 116:8722–31. doi: 10.1021/jp302103t
106. Likić VA, Gooley PR, Speed TP, Strehler EE, A. statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. *Protein Sci.* (2005) 14:2955–63. doi: 10.1110/ps.051681605
107. Pirolli D, Sciandra F, Bozzi M, Giardina B, Brancaccio A, De Rosa MC. Insights from molecular dynamics simulations: Structural basis for the V567D mutation-induced instability of zebrafish alpha-dystroglycan and comparison with the murine model. *PLoS ONE.* (2014) 9:e103866. doi: 10.1371/journal.pone.0103866
108. Ferrao R, Li J, Bergamin E, Wu H. Structural insights into the assembly of large oligomeric signalosomes in the toll-like receptor-interleukin-1 receptor superfamily. *Sci Signal.* (2012) 5:re3. doi: 10.1126/scisignal.2003124
109. Abdulkhalek S, Amith SR, Franchuk SL, Jayanth P, Guo M, Finlay T, et al. Neu1 sialidase and matrix metalloproteinase-9 cross-talk is essential for toll-like receptor activation and cellular signaling. *J Biol Chem.* (2011) 286:36532–49. doi: 10.1074/jbc.M111.237578
110. Feng C, Stamatos NM, Dragan AI, Medvedev A, Whitford M, Zhang L, et al. Sialyl residues modulate LPS-mediated signaling through the toll-like receptor 4 complex. *PLoS ONE.* (2012) 7:e32359. doi: 10.1371/journal.pone.0032359
111. Ohnishi T, Muroi M, Tanamoto K, N. -Linked Glycosylations at Asn 26 and Asn 114 of Human MD-2 Are Required for Toll-Like Receptor 4-Mediated Activation of NF-κB by Lipopolysaccharide. *J Immunol.* (2001) 167:3354–9. doi: 10.4049/jimmunol.167.6.3354
112. Amith SR, Jayanth P, Franchuk S, Finlay T, Seyrantepe V, Beyaert R, et al. Neu1 desialylation of sialyl α-2,3-linked β-galactosyl residues of TOLL-like receptor 4 is essential for receptor activation and cellular signaling. *Cell Signal.* (2010) 22:314–24. doi: 10.1016/j.cellsig.2009.09.038
113. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell.* (2007) 130:1071–82. doi: 10.1016/j.cell.2007.09.008
114. Ho MK, Se CO Ki JL, Kasamatsu J, Jin YH, Beom SP, Lee H, et al. Structural diversity of the hagfish variable lymphocyte receptors. *J Biol Chem.* (2007) 282:6726–32. doi: 10.1074/jbc.M608471200
115. Pancer Z, Cooper MD. The evolution of adaptive immunity. *Annu Rev Immunol.* (2006) 24:497–518. doi: 10.1146/annurev.immunol.24.021605.090542

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Immunology

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fimmu.2022.872676

In Vitro and *In Vivo* Antigen Presentation and Diagnosis Development of Recombinant Overlapping Peptides Corresponding to *Mtb* ESAT-6/CFP-10

Qing Zhang[1,2], Xiong Lu[2], Liang Gao[3], Siyu Tao[4], Yinghua Ge[2], Daocheng Cui[2], Renying Zhu[2], Wenshu Lu[2,5], Jian Wang[1]* and Shisong Jiang[6]**

[1] College of Veterinary Medicine, Faculty of Animal Science, Southwest University, Chongqing, China, [2] R & D Department, Oxford Vacmedix (Changzhou) Co. Ltd., Changzhou, China, [3] Department of Tuberculosis, Changzhou Third People's Hospital, Changzhou, China, [4] Department of Clinical Laboratory, Dehong Prefectural Hospital, Dehong Prefecture, China, [5] R & D Department, Shanghai JW Inflinhix Co. Ltd., Shanghai, China, [6] Department of Oncology, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
Adriana Harbuzariu, Emory University, United States

Reviewed by:
Alice Sijts, Utrecht University, Netherlands
Makoto Tsuiji, Hoshi University School of Pharmacy and Pharmaceutical Science, Japan

***Correspondence:**
Shisong Jiang
shisong.jiang@oncology.ox.ac.uk
Wenshu Lu
wenshu.lu@oncology.ox.ac.uk
Jian Wang
wj0931@swu.edu.cn

Specialty section:
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

Received: *09 February 2022*
Accepted: *16 May 2022*
Published: *16 June 2022*

Citation:
Zhang Q, Lu X, Gao L, Tao S, Ge Y, Cui D, Zhu R, Lu W, Wang J and Jiang S (2022) In Vitro and In Vivo Antigen Presentation and Diagnosis Development of Recombinant Overlapping Peptides Corresponding to Mtb ESAT-6/CFP-10. Front. Immunol. 13:872676. doi: 10.3389/fimmu.2022.872676

Cellular immunity in *Mycobacteria tuberculosis* (*Mtb*) infection is important for the pathogenesis and final clearance of intracellular *Mtb* infection. In addition, it is valuable for the diagnosis of tuberculosis. In this pioneering work, we tested *in vitro* and *in vivo* antigen presentation and diagnostic application of a recombinant overlapping peptide-protein derived from two *Mtb* RD1 antigens ESAT-6 and CFP-10 (ROP-TB). The overlapping peptide sequence of ROP-TB is cleaved by the cathepsin S enzyme and covers the entire length of the two proteins. ROP-TB can be expressed and purified from *E. coli*. Once taken in by antigen-presenting cells, ROP-TB can be cleaved into a peptide pool by cathepsin S within the cells. We found that in dendritic cells, ROP-TB can be processed in 6 hours of co-culture, while the ESAT-6/CFP-10 fusion protein remained in the endosomal compartment. In *Mtb*-infected mice, ROP-TB stimulated stronger specific T cell responses than pooled synthetic peptides derived from ESAT-6 and CFP-10. With regard to the presentation of *in vivo* antigens, in a guinea pig model infected with *Mtb*, ROP-TB induced delayed type hypersensitivity (DTH) responses comparable to those of the tuberculin purified protein derivative (PPD) and ESAT-6/CFP-10 fusion protein. In *Mycobacterium bovis (Bovine TB)*-infected cattle, ROP-TB elicited DTH responses. Finally, in *Mtb* infected patients, ROP-TB stimulated cellular immune responses in majority of patients (16/18) of different HLA phenotypes while a single peptide derived from the same proteins did not elicit the immune responses in all patients. In summary, *in vitro* and *in vivo* data suggest that ROP-TB stimulates a strong cellular immune response irrespective of HLA phenotypes and is therefore suitable for use *in vitro* and *in vivo* diagnostics.

Keywords: recombinant overlapping peptides, mycobacteria tuberculosis, Bovine TB, antigen presentation, immune responses, diagnostic

INTRODUCTION

Cell-mediated immunity requires antigen presentation by class I or II MHC molecules on the surface of antigen-presenting cells (APC). CD8+ T cells typically recognize epitopes presented with MHC class I derived from endogenous antigen degradation in the cytoplasm of any nucleic cell, including APCs (Class I pathway). In contrast, exogenous antigens are processed in the endosome/lysosome in APC into peptides which can be presented by MHC class II (only expressed in APCs) to the cell surface and stimulate CD4+ T cells (Class II pathway). In some situations, exogenous antigens can be cross-presented in APCs to prime CD8+ T cells (Cross-presentation). However, the efficiency of antigen cross-presentation is low (1–3).

Cellular immunity is very important in the pathogenesis and elimination of *Mtb* infection. Furthermore, since the antibody is not suitable for the diagnosis (4–6), cellular immunity is a useful surrogate diagnostic marker for *Mtb* infection. In fact, the World Health Organization (WHO) has recommended two methods based on cellular immune response for latent TB infection (LTBI) (7). The methods have enabled the identification of LTBI and treatment with anti-TB drugs. This is particularly important for stopping the transmission of *Mtb*. In the TB guidelines of the European Union, the interferon-gamma (IFN-γ) release assay (IGRA) has also been suggested as an accessory diagnostic method for active TB (8).

The first cellular immune response-based method is the IGRA, in which peripheral blood mononuclear cells (PBMCs) of an LTBI-suspected individual are exposed to *Mtb* antigens. If the individual has LTBI, memory T cells in the PBMCs will be activated *in vitro* by *Mtb* antigens to release cytokines, including IFN-γ, which can be measured (9). The key reagent for the currently approved IGRA-based latent TB diagnosis is a pool of short peptides derived from two *Mtb* RD1 antigens: early secreted antigen target 6 (ESAT-6) and cell filtrate protein 10 (CFP-10). The advantages of using the ESAT-6 and CFP-10 derived peptide antigens are: 1) distinguish from BCG (bacille Calmette-Guérin) vaccination; 2) stimulate strong cellular immune responses; and 3) overcome MHC restriction as the pooled peptides contain multiple epitopes suitable for different HLA phenotypes. However, quality control in the industrial manufacturing process of multiple peptides requires multiple GMP processes and is labor-intensive and time-consuming. It has posed a limitation to cost-effective manufacturing of diagnostic kits. Consequently, the use of the IGRA assay as a screening test for latent TB for the general population has not become a reality.

Another method of detecting LTBI is the Mantoux skin test based on a purified protein derivative (PPD). This is an *in vivo* method based a delayed-type hypersensitivity (DTH) response in which the antigen presentation of the exogenous PPD antigen leads to a cellular immune response as a consequence of a collective cytokine release. The sign of an immune response is a skin reaction at the PPD injection site. The advantages of this method include the ability to measure cellular immune responses not only based on one cytokine (i.e., IFN-γ) but also on the collective stimulation of other factors (cytokines/chemokines). Furthermore, it a simple and economical test, with no need for *in vitro* cell culture equipment or expensive reagents (tissue culture plates and antibodies); therefore, it is suitable as a screening test in the general population. However, the PPD-based skin test cannot distinguish *Mtb* infection from BCG vaccination.

The goal of this study is to develop an antigen that can serve as a key reagent for both the IGRA assay and the skin test. The reagent should be able to distinguish *Mtb* infection from BCG vaccination and be suitable for industrial manufacture. We have previously reported using recombinant overlapping peptide (ROP) proteins as stimulants for cellular immunity (10, 11). An ROP is composed of a chained series of overlapping peptide sequences derived from a target protein (e.g., ESAT-6) and covers the whole sequence of the target protein (**Figure 1A**). The overlapping sequences of an ROP are linked by a substrate of protease (e.g., Cathepsin) in the antigen-presenting cells. The ROP can be easily cleaved inside APCs and cross-present to induce potent CD4+ and CD8+ T-cell responses (11). There are several advantages of using ROP: 1) the ease of industrial manufacture, due to only a single GMP processing step and the requirement for one quality control; 2) ROPs are linear sequences containing all possible T cell epitopes which can promiscuously stimulate T cells irrespective of MHC phenotypes; and 3) ROPs do not have tertiary structure, therefore, they do not possess any toxic protein functions of the native microbial/tumor proteins.

Consequently, ROP technology may be a solution to overcome the drawbacks of manufacturing and quality control of the key reagents in the current IGRA assay. In addition, it may also serve as a skin test agent. In this study, our goal was to design and produce an ROP-TB diagnostic agent based on the two *Mtb* antigens, ESAT-6 and CFP-10. We then used *in vitro, ex vivo*, and *in vivo* experiments to demonstrate antigen cross-presentation and induction of cellular immune responses.

RESULTS

Design, Expression, Purification, and Enzymatic Digestion of ROP-TB

ROP-TB was derived from ESAT-6 and CFP-10. It was designed to include a sequence of 8 peptides consisting of 35 amino acids per peptide. Each peptide overlapped with its adjacent peptides by 10 amino acids. The eight sequential overlapping peptides covered the complete sequences of ESAT-6 and CFP-10. The cathepsin S substrate sequence (LRMK) was interspersed between each peptide. Cathepsin S is known to exist in professional APCs (10). A schematic presentation of ROP-TB is shown in **Figure 1A**. ROP-TB was expressed by *E. coli* BL21 cells and then purified by Ni NTA affinity chromatography. (**Figure 1B**). The purity of the protein was confirmed by HPLC to be more than 95% (**Figure 1C**).

To remove any potential endotoxins that may have potentially contaminated the product, ROP-TB was further

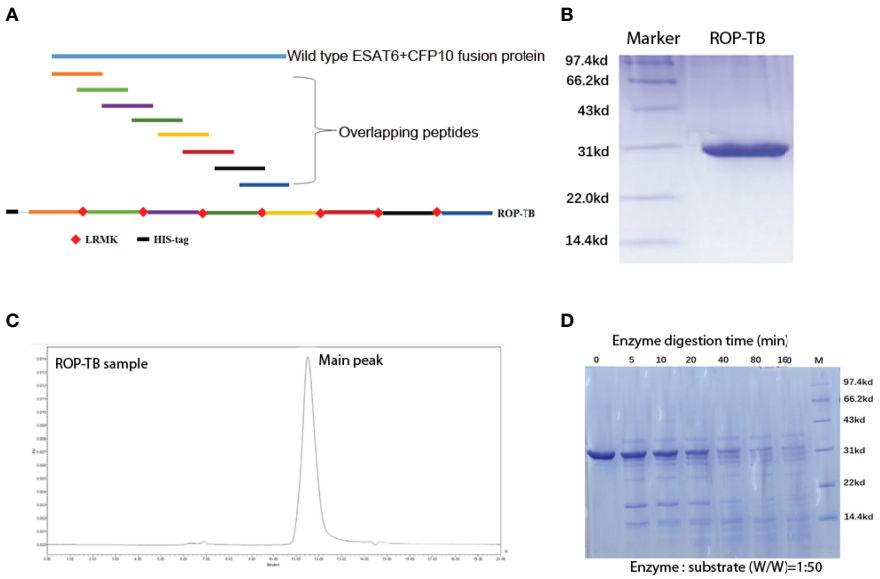


FIGURE 1 | Purification and digestion of ROP-TB. **(A)** Schematic presentation of an ROP-TB antigen. The blue line on top represents the sequence of the fusion protein of ESAT-6 and CFP-10; the short colorful lines are the overlapping peptides covering the full sequence length of ESAT-6-CFP-10. The bottom line represents ROP-TB which is a chain of the overlapping peptides linked by LRMK, the substrate of the enzyme Cathepsin S. **(B)** ROP-TB was expressed by *E. coli* BL-21 and purified using a Ni-NTA affinity column. **(C)** After purification, ROP-TB was analyzed by HPLC to further confirm purity. **(D)** ROP-TB was digested at 37°C by Cathepsin S at 1:50. After digestion, samples were collected at different times and analyzed by SDS-PAGE gel.

purified by endotoxin removal chromatography. Subsequent endotoxin levels in ROP-TB before application in cell cultures was tested and were <1 EU/mL. To test the effects of endotoxin on the ROP-TB-based IGRA assay, an enzyme-linked immunosorbent spot (ELISPOT) assay study using ROP-TB spiked with endotoxin was carried out. The doses of endotoxin used (up to 33 EU/mL) did not influence the results of the ROP-TB-related ELISPOT (**Supplementary Figure 1**). This is consistent with another report showing that only more than 100 EU/mL of endotoxin induces an ELISPOT response (12).

To test the enzymatic effects of cathepsin S on the protein, ROP-TB was incubated *in vitro* with cathepsin S. After digestion, samples were analyzed using SDS-PAGE. As **Figure 1D** shows, ROP-TB can be digested *in vitro* into peptides of different sizes by cathepsin S (**Figure 1D**).

Comparing the Metabolic Rate of ROP-TB, TB Peptides, and TB Fusion Protein in APCs

To understand the cross-presentation efficiency of ROP-TB in APCs, it is important to investigate its metabolic rate in the endolysosome. We, therefore, compared ROP-TB with TB peptides and TB wild-type protein (fusion of ESAT-6 and CFP-10) by tracing their intracellular localization at different time points in the mouse dendritic cell line DC2.4 cells. ROP-TB, TB peptides, and TB protein were labeled with a green fluorescent dye (WV=530/30nm) and were incubated with DC2.4 cells for 30 min and 6 h. DC2.4 cells were co-incubated with Lysotracker to emit red fluorescence (WV=630/30nm). The uptake of ROP-TB, TB peptides, or TB protein is shown in green, and endolysosomes are shown in red under a confocal microscope. The co-localization between red and green colors represents the approximate location of the two objects. As shown in **Figures 2A–C**, ROP-TB, TB peptides, and TB fusion protein are internalized by DC2.4 within 30 min of incubation and are largely co-localized within endolysosomes. This indicates that these antigens are located within the endolysosomal compartment. After 6 hours, the green fluorescence in **Figures 2D, E** (ROP-TB and TB peptides) but not in **Figure 2F** (TB fusion protein), disappeared (**Figures 2D–F**), suggesting that ROP-TB and TB peptides have been processed more rapidly (**Figure 2**).

Development of T Cell-Based Diagnosis in *Mtb*-Infected Mouse Model: ROP-TB and Pooled Synthetic TB Peptides Induce Specific T Cell Responses

To test whether ROP-TB can be developed as a diagnostic agent for T cell function, we assessed specific T cell responses induced by synthetic ROP-TB or *Mtb* peptides *ex vivo* in a mouse model. We first established the test model by infecting BALB/c mice with *Mtb* (H37R1) intranasally (**Figure 3**). Infectivity was validated by acid-fast staining and Hemotoxin-Eosin (H&E) staining of lung tissue (**Figure 3A** upper panel). The lungs of infected mice clearly showed infection with H37R1 and

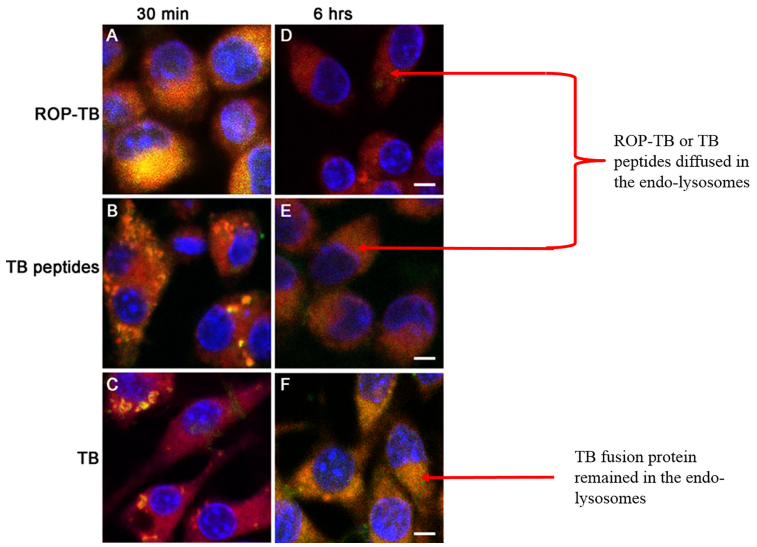


FIGURE 2 | Intracellular localization of ROP-TB, TB peptides, and TB protein in DC2.4 cells. DC2.4 cells were incubated with green fluorophore labeled ROPTB, TB peptides, or TB protein for 30 min and 6 hours followed by confocal microscopy analysis (63x objective). DC2.4 cells were co-stained with lysotracker red for visualizing endo-lysosomes. The nucleus was counterstained with DAPI. **(A–C)** ROP-TB, TB peptides, and TB fusion protein are internalized by DC2.4 within 30 min of incubation and are largely co-localized within endolysosomes compartment. **(D, E)** ROP-TB and TB peptides diffused in the endo-lysosomes after 6 hours. **(F)** TB fusion protein remained in the endo-lysosomes after 6 hours. Scale bar: 10 μm.

presented lesions when compared to non-infected mice (**Figure 3A**, lower panel).

Spleen cells from *Mtb-infected* mice were then isolated and stimulated with different concentrations of synthetic ROP-TB and *Mtb* peptides. T cell responses to these antigens were measured by the IGRA assay (ELISPOT). As shown in **Figure 3B**, ROP-TB stimulates stronger specific T cell responses than synthetic peptides, indicating that ROP-TB is a more sensitive antigen in the mouse model than synthetic peptides in detecting T cell responses.

Delayed-Type Hypersensitivity Responses to ROP-TB, ESAT-6/CFP-10 and PPD in Guinea Pigs Infected With Mtb and Cattle Infected With *M. bovis*

Delayed-type hypersensitivity (DTH) reflects T cell responses to antigen stimulation *in vivo*. To investigate whether ROP-TB can be processed and presented *in vivo* to specific T cells, guinea pigs were infected with *Mtb* H37R1, and DTH-related skin reactions to ROP-TB, ESAT-6/CFP-10, and PPD (purified protein derivative) were investigated. Both ROP-TB and ESAT-6/CFP-10 trigger

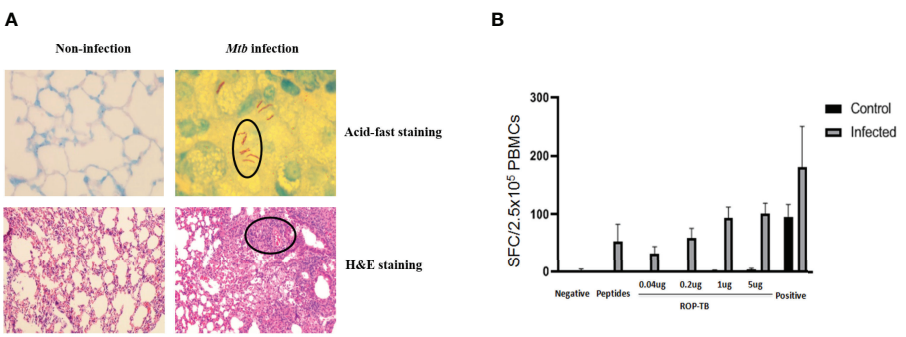


FIGURE 3 | ROP-TB stimulates specific T cell immune response in the *Mtb*-infected mouse model **(A)** Establish *Mtb*-infected mouse model by intranasal infection of *Mtb*. Upper panels: acid-fast staining of lung alveolar in noninfected (left) and infected (right) mice. *Mtb* are observed in the lung (circled) only in infected mice (right). Lower panels: H&E staining of histopathological changes of the lungs in non-infected mice (left) and infected mice (right). Marked lymphocyte infiltration in the interstitial and alveolar space is present only in infected mice (right) together with pulmonary edema, alveolar distortion, and thickening of the alveolar wall. **(B)** Specific T cell responses of infected mice. Responses of spleen cells from noninfected (control) and infected mice to different doses of ROP-TB are only seen in infected mice. Negative control: culture medium (R10); stimulant: peptides (5 μg/well); ROP-TB with 4 different doses (0.04, 0.2, 1 and 5 μg/well); Positive control: PHA (5 μg/well).

positive responses to DTH, as well as PPD (**Figure 4A**). DTH elicited by ROP-TB was stronger than that elicited by ESAT-6/CFP-10 and PPD in 100 μg and 10 μg doses ($p<0.01$) (**Figure 4A**).

Bovine TB (*M. Bovis*) epidemics have been reported in many parts of the world. This has posed a risk for humans drinking infected, unpasteurized milk. Practically all cattle on a farm need to be slaughtered even if only one head of cattle is found to be infected with *M. bovis*. This causes great economic loss. Thus, as a preventative measure, a tuberculin-based skin test using a PPD is usually performed to screen for infected cattle (13). BCG may also be used as a vaccine for *M. bovis* infection (14) if there is another available diagnostic reagent which will not be interfered by the vaccination.

Since *M. bovis* and *Mtb* share the same ESAT-6 and CFP-10 sequences (15), we wondered whether our ROP-TB reagent could be used as another choice of *M. bovis* skin test reagent. Thus, the application ROP-TB as a diagnostic reagent for *M. bovis* infection was tested in cattle from two farms (Farm 1, n=12 and Farm 2, n=40, respectively) recruited for ROP-TB and PPD- based skin tests. As shown in **Figure 4B**, both ROP-TB and PPD stimulates skin reactions in the cattle from Farm 1. Among the 12 cattle, 7 show positive skin reactions to ROP-TB stimulation (**Figure 4C**, left panel) and 7 to PPD stimulation (**Figure 4C**, right panel). **Figure 4D** shows the interrelationship of ROP-TB-stimulated and PPD-stimulated skin reactions in these cattle. None of the cattle in Farm 2 produced a positive skin reaction to antigens. **Table 1** summarizes the skin reactions of both farms. The sensitivity and specificity of ROP-TB with PPD are 85.7% and 97.7%, respectively (**Table 1**).

ROP-TB Overcomes HLA Restriction in Stimulating T Cells in TB Patients

To observe whether ROP-TB can promiscuously stimulate the memory T cell response in humans, we tested ROP-TB for a

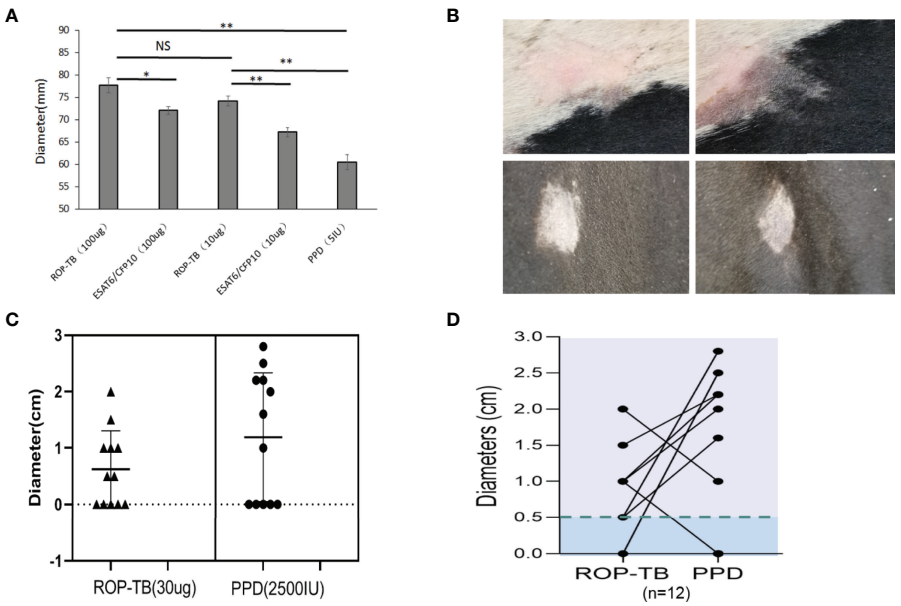


FIGURE 4 | Delayed-type hypersensitivity to ROP-TB, ESAT-6/CFP10, and PPD in *Mtb* infected guinea pigs and *M. bovis*-infected cattle. **(A)** Skin reaction in guinea pigs: the bars show the size of skin reaction to two doses (100 μg and 10 μg) of ROP-TB and ESAT-6/CFP-10, respectively in guinea pigs (n=4 per group). PPD at 5 IU was used as a comparison. Delayed-type hypersensitivity to ROP-TB and PPD in cattle infected with tuberculosis (n=12). **(B)** Skin reaction to cattle: Representative positive skin reactions to ROP-TB (upper left panel) and PPD (upper right panel); and negative reactions to ROP-TB (lower left panel) and PPD (lower right panel). **(C)** The size of the skin reactions of all 12 cattle to ROP-TB (left panel) and PPD (right panel). **(D)** The relationship of DTH of each head of cattle to ROP-TB and PPD. "*" ($P<0.05$): significant difference, "**" ($P<0.01$): extremely significant difference. NS; no statistically significant difference.

TABLE 1 | Delayed-type hypersensitivity skin test of ROP-TB and PPD cattle.

Antigen	Farm 1 N=12		Farm 2 N=40	Sensitivity (%)[a]	Specificity (%)[b]
	Positive	Negative	positive		
PPD	7	5	0	85.7	97.7
ROP-TB	6	6	0		

[a]*Percentage of positive responders with ROP-TB in positive responders with PPD.*
[b]*Percentage of negative responders with ROP-TB in positive responders with PPD.*

specific T cell immune response measured by ELISPOT in a population of diverse HLA phenotypes. We recruited 18 patients with tuberculosis who were diagnosed by sputum test, acid-fast staining, and lung computed tomography (CT) scan. The demographic data for these patients are summarized in **Table 2**. Their HLA phenotypes are shown in **Table 3**. We then compared the ELISPOT response of the PBMC of the 18 patients with the following antigens: 8 individual overlapping synthetic peptides derived from ESAT-6 and CFP-10, pooled peptides of the 8 overlapping synthetic peptides derived from ESAT-6 and CFP-10, and ROP-TB (which is composed of linked and chained peptides of the 8 peptide sequences). As shown in **Table 4**, the 18 patients having different HLA types (**Table 3**) respond to different individual peptides. However, most patients, 16/18 in ROP-TB and 15/18 in pooled peptides, respond to the pooled peptides and ROP-TB (**Table 4**). This experiment shows that ROP-TB, like the pooled peptides, is able to stimulate memory T cell responses in different HLA types of TB patients.

DISCUSSION

Monitoring and detecting specific T cell responses is important for many diseases, such as tumors, infectious diseases, and inflammatory diseases. However, currently, TB is the only disease indication for commercially available, T cell-based diagnostic assays, which consist of a skin test and IGRA assays. The difficulties of developing a specific T-cell-based test include the identification of specific antigens that can promiscuously stimulate T cells from various MHC backgrounds; the manufacture of such antigens without labor intensive processes such as multiple GMP manufacturers and/or quality controls; and the requirement of cell culture facilities for their execution, which may hamper the use of the assays in a hospital setting, especially in remote areas in developing countries.

For the two TB diagnostic methods currently used clinically, the key reagents/antigens are derived from different origins. The antigen used in the skin test assay is a PPD, while the antigen used in the IGRA assays consists of a peptide pool derived from

TABLE 2 | Clinical data of the 18 *Mtb* patients.

Patient	Sex	Age	White blood cell count (10^9/L)	Tuberculin Test	Acid-fast staining	X-pert	Sputum culture	CT
WS-118	Female	31	5.03	++	—	+	—	Plaque shadows in the left lung
WS-119	Female	12	4.4	++	—	+	—	Plaque shadows in both lungs,
WS-120	Female	40	4.28	++	+	+		Plaque shadows in the right lower middle lung
WS-121	Male	77	9.29	+	+	—		Multiple cavities in both lungs
WS-127	Male	75	8.18	+	+		+	Both lung lesions; damage to the left lung
WS-129	Male	63	2.76	++	+	+	—	Partial damage to the right lung
WS-130	Male	39	5.34	++	+	+	—	Plaque shadows in the left lung
WS-139	Male	30	8.56	++	+	+	—	Multiple high-density shadows in both lungs
WS-140	Male	64	4.54	++	—	+	—	Nodules in the upper lobe of the right lung with a few surrounding patchy shadows
WS-141	Female	22	4.45	++	—	—	—	Patchy shadows in the upper right lobe of the lung
WS-142	Female	39	7.08	++	—	+		Patchy shadows in the upper lobe of the left lung,
WS-143	Female	48	2.49	+	—			Plaque shadows and pleural effusion in the left lung
WS-157	Female	27	5.99	++	—			High-density patchy shadows in both lungs
WS-158	Female	77	5.3	++	–			Density patchy shadows in both lungs
WS-159	Male	58	7.96	+		+		Multiple cavities in both lungs
WS-160	Female	51	7	++	—	—	—	Density patchy shadows in both lung
WS-161	Male	59	3.58	+	+	+		Density patchy shadows in both lungs; pleural effusion in the left lung
WS-162	Female	72	6.23	++	+			Plaque, shadow/cavities in the left lung; plaque shadows in the right lung

"++" strong positive, "+"positive, "—"negative.

TABLE 3 | HLA phenotypes of the 18 *Mtb* patients.

Patient	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1
WS-118	A*11:02 A*33:03	B*40:01 B*58:01	C*03:02 C*07:02	DRB1*03:01 DRB1*08:03	DQB1*02:01 DQB1*06:01
WS-119	A*24:08 A*24:02	B*15:01 B*54:01	C*01:02 C*01:02	DRB1*11:01 DRB1*12:01	DQB1*03:01 DQB1*03:01
WS-120	A*02:07 A*24:02	B*46:01 B*52:01	C*01:02 C*07:02	DRB1*08:03 DRB1*14:10	DQB1*05:03 DQB1*06:01
WS-121	A*02:07 A*11:01	B*13:01 B*54:01	C*01:02 C*07:02	DRB1*04:05 DRB1*12:02	DQB1*03:01 DQB1*04:01
WS-127	A*24:02 A*31:01	B*46:01 B*55:02	C*01:02 C*01:02	DRB1*14:05 DRB1*14:01	DQB1*05:02 DQB1*05:03
WS-129	A*02:10 A*30:01	B*13:02 B*40:06	C*06:02 C*08:01	DRB1*07:01 DRB1*12:01	DQB1*02:03 DQB1*03:04
WS-130	A*02:03 A*02:01	B*15:02 B*35:01	C*03:03 C*08:01	DRB1*12:02 DRB1*15:01	DQB1*03:01 DQB1*06:02
WS-139	A*30:01 A*33:03	B*13:02 B*58:01	C*03:02 C*06:02	DRB1*07:01 DRB1*13:02	DQB1*02:02 DQB1*06:09
WS-140	A*11:01 A*24:02	B*15:01 B*35:01	C*03:03 C*04:01	DRB1*09:01 DRB1*15:01	DQB1*03:03 DQB1*06:02
WS-141	A*02:07 A*11:01	B*15:58 B*38:02	C*01:02 C*07:02	DRB1*04:03 DRB1*09:01	DQB1*03:02 DQB1*03:03
WS-142	A*11:01 A*33:03	B*15:02 B*58:01	C*03:02 C*08:01	DRB1*12:02 DRB1*13:02	DQB1*03:01 DQB1*06:09
WS-143	A*11:01 A*24:02	B*15:27 B*35:01	C*03:03 C*04:01	DRB1*11:01 DRB1*15:01	DQB1*03:01 DQB1*06:02
WS-157	A*02:01 A*31:01	B*15:11 B*15:01	C*01:02 C*03:03	DRB1*07:01 DRB1*09:01	DQB1*02:02 DQB1*03:03
WS-158	A*24:02 A*24:02	B*15:27 B*40:02	C*03:04 C*04:01	DRB1*04:06 DRB1*09:01	DQB1*03:02 DQB1*03:03
WS-159	A*11:01 A*24:02	B*40:06 B*40:01	C*07:02 C*08:01	DRB1*08:03 DRB1*09:01	DQB1*03:03 DQB1*06:01
WS-160	A*03:01 A*33:03	B*35:01 B*52:01	C*07:02 C*12:02	DRB1*15:01 DRB1*15:02	DQB1*06:01 DQB1*06:02
WS-161	A*11:02 A*11:01	B*15:02 B*27:04	C*08:01 C*12:02	DRB1*08:03 DRB1*12:02	DQB1*03:01 DQB1*06:01
WS-162	A*11:01 A*11:02	B*15:02 B*40:01	C*07:02 C*08:01	DRB1*09:01 DRB1*12:02	DQB1*03:01 DQB1*03:03

ESAT-6 and CFP-10 (16, 17). PPD does not distinguish TB infection from BCG vaccination. Although, pooled ESAT-6/CFP-10 peptides can distinguish TB infection from BCG vaccination; they may not be suitable for industrial manufacture due to complicated multiple quality controls and regulation issues, especially in the case of skin tests in which the antigen will be injected into a human body.

The aim of this study was to develop an antigen suitable for both IGRA and a diagnostic skin test. We designed a chain of overlapping peptides (ROP) linked by the enzyme substrate cathepsin S. This antigen is cleavable in APCs where the enzyme cathepsin S is located. After degradation by the enzyme, the antigen will undergo presentation and/or cross-presentation to stimulate T cells. Our data from this study showed that ROP-TB could be produced and purified recombinantly using the *E. coli* system (**Figure 1**). The ROP-TB can be cross presented in APC (**Figure 2**) to stimulate T cells (**Figure 3**) in mice. It has also been shown to induce skin reactions in guinea pigs and cattle (**Figure 4**). Finally, ROP-TB promiscuously stimulates human T cell responses (**Table 4**).

There are several advantages of this artificial ROP antigen. First, ROP-TB is derived from ESAT-6/CFP-10 which is different from BCG and hence it is specific for TB infection but not for BCG vaccination. Second, since the antigen is produced by *E. coli*, which can be adapted to large-scale manufacturing processes, the costs are reasonably low. Moreover, ROP-TB is produced as one entity; therefore, only one quality control step is required in the downstream process of manufacture. Furthermore, unlike pooled peptides, ROP-TB is convenient (as it is applied as a single substrate), which is useful when applying for regulatory approval. Third, like pooled peptides, the antigen is promiscuous and stimulates T cells from individuals of

TABLE 4 | ELISPOT responses of the 18 *Mtb* patients to various forms of TB antigens.

Patient	P1	P2	P3	P4	P5	P6	P7	P8	Pool of P1-P8	ROP	CFP-10	ESAT-6
WS-118	—	—	—	+	—	—	+	+	+	+	+	+
WS-119	—	+	—	—	+	—	+	+	+	+	+	+
WS-120	+	—	+	+	+	+	+	+	+	+	+	+
WS-121	—	—	—	—	+	—	+	—	+	+	+	+
WS-127	—	—	—	—	—	—	—	—	—	—	—	—
WS-129	—	—	—	—	—	—	—	+	—	—	—	+
WS-130	—	—	+	—	—	—	—	—	+	+	+	+
WS-139	—	—	—	—	—	—	+	—	—	+	—	+
WS-140	—	—	+	—	—	—	+	—	+	+	+	+
WS-141	—	+	+	+	—	—	+	+	+	+	+	+
WS-142	—	—	+	—	+	—	+	+	+	+	+	+
WS-143	—	—	—	—	—	—	—	+	+	+	—	+
WS-157	—	—	—	—	+	—	—	—	+	+	+	
WS-158	+	+	+	+	—	—	—	—	+	+	+	—
WS-159	+	—	—	—	+	—	—	—	+	+	+	—
WS-160	+	—	+	—	—	—	+	+	+	+	+	+
WS-161	—	—	+	+	+	+	—	—	+	+	+	+
WS-162	+	—	+	—	—	—	—	—	+	+	+	—

"+" Represent positive, "—" Represent negative.

different HLA phenotypes. This has been demonstrated by this study (**Tables 3**, **4**). Fourth, ROP-TB does not have share the same conformation structure as native antigens (ESAT-6/CFP-10), therefore, it lacks the potential pathogenic functions of native ESAT-6/CFP-10. This is especially important if it is intended to be used as a skin test reagent. Fifth, ROP-TB can also be produced under other forms such as DNA, mRNA, and viral/bacterial vectors, which represents an added advantage to using ROP-based antigens.

In summary, we have demonstrated that ROP-TB can be used as a new diagnostic platform technology for tuberculosis diagnostics in both IGRA assays and skin tests.

METHODS

Molecular Cloning, Protein Expression, and Purification

The cDNA genes encoding native ESAT-6-CFP-10 and the ROP-TB respectively were designed and synthesized as previously described (6).

For purification, the corresponding plasmids were transformed into *E. coil* BL21 (Solarbio, China). A single colony was harvested in LB medium (50 μg/mL Kanamycin) and cultured overnight. Bacteria from the overnight culture were then inoculated into the new LB medium at a 1:100 ratio and cultured until the OD600 reached 0.6. Isopropyl β- d-1-thiogalactopyranoside (IPTG, Solarbio, China) was added at a final concentration of 0.5 mM and the bacteria were collected by centrifugation after 16 h of incubation. Lysis buffer (20mM Tris-HCl, 0.5 M NaCl, 2 mM EDTA, pH 8.0) was used to resuspend the bacteria, which were then lysed by sonication. The insoluble fractions were separated by centrifugation at 15,000 g for 1 hour.

The protein was further purified by nickel affinity chromatography using Ni-NTA (GE, USA). To detect expressed soluble TB protein (ESAT-6-CFP-10), Ni-NTA resin was added to the soluble fractions for 1 h. The protein was washed with lysis buffer containing 10 mM imidazole (washing buffer) and eluted with lysis buffer containing 300 mM imidazole (elution buffer). The insoluble ROP-TB in the pellet was resuspended in 8 M urea buffer and centrifuged at 15,000 g for 1 h. The rest procedures were similar to those of ESAT-6-CFP-10 except that the washing and elution buffer contained 8 M urea. For refolding, the eluted proteins (ESAT-6-CFP-10 and ROP-TB) were exchanged by PBS for dialysis.

Endotoxin Removal and Limulus Amebocyte Lysate (LAL) Assay for Endotoxin

Ni-NTA purified ROP-TB was further purified by endotoxin removal chromatography (BIOENDO, China). Briefly, the column (1.5 mL) was washed with balanced buffer, then ROP-TB was flowed through the column at 0.5 mL/min and the effluent collected.

The endotoxin present in ROP-TB was detected using the Limulus amebocyte lysate assay (BIOENDO, China) following the manufacturer's instructions. Briefly, ROP-TB was diluted with endotoxin-free water by serial two-fold dilutions, then diluted ROP-TB was added, respectively, to the Limulus amebocyte lysates, which were dissolved in 100 μL endotoxin-free water. The mixtures were incubated in ampoule bottles at 37°C for 1 hour. If a gelatinous substance formed at a certain dilution, it was considered positive: Endotoxin concentration was calculated as maximum positive × 0.125 EU/mL, which indicated the sensitivity of Limulus amebocyte lysate.

Antigen Digestion *In Vitro* and Trafficking in DCs by Confocal Microscopy

Purified protein (ROP-TB) in PBS and 4mM DTT was incubated with Cathepsin S (SinoBiological, China) at 37°C for 0.5, 1, 2, 4, 8, and 22 h. The mass ratios of enzyme and protein in digestion were 1:50 and 1:100, respectively. Samples with different digestion times were run through SDS-PAGE for measuring enzyme activity.

To observe antigen uptake, 1 μM ROP-TB, TB peptides, and TB protein (ESAT-6-CFP-10) were incubated with DC2.4 cells (a gift from K Rock of Massachusetts University, United States) in culture plates ($3x10^5$ per dish) at 37°C for 30 minutes or 6 hours. ROP-TB, TB peptides, and TB protein marked with the LIVE/DEAD cell stain kit (Invitrogen, USA). To stain endolysosomes, 10 μL of 1:200 diluted (red) Lysotracker (Invitrogen, USA) was added to the cell culture one hour before incubation. After incubation, cells were washed and fixed with 4% paraformaldehyde for 10 min. The samples were further stained with DAPI (Solarbio, USA) prior to observation under a confocal microscope (ZEISS, LSM800).

In Vitro of IGRA-Based T Cell Diagnosis in Mice

The mouse experiment was outsourced to the Shanghai Public Health Center (Shanghai, China). All procedures were approved by the Committee of the Shanghai Public Health Center and followed the national guidelines for the use of experimental animals. Briefly, all mice (BALA/c, age 6–8 weeks, SPF) were infected intranasally with *Mtb* (H37R1) at $5x10^5$ CFU in 40 μL. After infection, the animals were housed in cages kept in laminar flow safety enclosures in a Level III biosafety facility.

To verify the establishment of the *Mtb* model, the lung tissues of infected and uninfected mice were sectioned for fast acid staining. The sections were placed in a staining dish and stained with Carbol Fuschin. After heating to steam and incubation for 5 min, the lung sections were washed with slow running water, decolorized with acid alcohol, counterstained with Malachite green for 1 min, and then air-dried. The final sections were positive when red and negative otherwise. The procedure of H&E staining (Sigma)for the evaluation of lymphocyte infiltration is the following: samples were first fixed overnight with 4% paraformaldehyde. After embedding in paraffin, the samples were cut into 2-μm thick sections and subjected to H&E staining. Histopathological analysis was performed using a Zeiss microscope.

The IGRA-based T cell function assay was performed using ELISpot kits (Mebtech, Sweden) (6). Briefly, spleen cells were separated and incubated at $2.5x10^5$ cells/well with 5 μg/well ROP-TB or peptides at 37°C for 24 h. Then, 96-well culture plates were precoated with IFN-γ antibodies. After the incubation, the cells were discarded, and the plates were washed with PBS followed by incubation with biotinylated anti-IFN-γ antibodies for 2 h at room temperature. The plates were washed with PBS and an enzyme-labeled anti-biotin antibody was added for 1 h. Finally, BCIP/NBT was used for color development, and the reaction was stopped by washing plates with tap water. Spots were counted with an Elispot reader (Autoimmun Diagnostike, Strasburg, Germany).

Ex Vivo IGRA-Based T Cell Diagnosis in Human

The human experiments were reviewed and approved by the Ethics Committee of Changzhou No. 3 People's Hospital. Eighteen patients with tuberculosis infection were recruited from the hospital and provided informed written consent prior to the collection of their detailed clinical data (**Table 2**). HLA typing of the patients was performed by Weihe Biotechnology INC (Jiangsu, China).

IGRA-based T cell testing was performed using ELISpot kits (Mebtech, Sweden). Briefly, $2.5x10^5$/150 μL of PBMCs were stimulated overnight with 5 μg/well of ROP-TB, either pooled or as singular peptides in anti-IFN-γ-Ab precoated plates (Millipore, Bedford, MA). Cells were discarded and biotinylated anti-IFN-γ antibodies were added for a 2-h incubation at room temperature followed by another 1-h incubation at room temperature with an enzyme-labeled anti-biotin antibody. After the color development, the reaction was stopped by washing the plates with tap water, and the plates were air dried. Spots were counted with an Elispot reader (Autoimmun Diagnostike, Strasburg, Germany).

In Vivo Diagnosis of Tuberculosis (Skin Test) in Guinea Pigs

The experiment was outsourced to the Shanghai Public Health Center (Shanghai, China). All animals were infected with *Mtb* (H37R1) of 0.5 mL ($5x10^4$ CFU) through nasal drop. After infection, the animals were housed in cages contained within laminar flow safety enclosures in a Level III biosecurity facility.

The DTH-based skin test was performed as follows: guinea pigs were divided into six groups of five animals per group. They were injected intradermally on the flank with 0.1 mL of ESAT-6-CFP-10 (1 mg/mL, 0.1 mg/mL), ROP-TB (1 mg/mL, 0.1 mg/mL), national standard and negative control (physiological saline). The diameters of the erythema were read after 24–72 hours.

In Vivo Skin Test of TB in Cattle

The experiments were approved by the Ethics Committee of Southwest University (Ethics approval number: IACUC-20210320-03). Cattle from two farms (n=12 and n=40, respectively) were recruited. To evaluate the potential of a DTH-based skin test with ROP-TB, we selected commercially available PPD, which is the standard reagent for the diagnosis of *M. bovis* TB, to be compared with ROP-TB in the experiments. The PPD skin test was performed following the standard Chinese diagnostic technique for tuberculosis in animals. ROP-TB antigen (30 μg/animal) was injected into a volume of 0.1 mL on one side of the cow neck, while PPD (2500 IU/animal) on the other side. Skin thicknesses at injection sites were measured between 48 and 72 h of injection. For the readout of the skin test, if the skin thickness of the cattle was ≥ 4 mm, the animal was considered skin test-positive and negative if the thickness was <2 mm. In the case that the difference was nearly or equal to 2 mm, the result was considered inconclusive.

Statistical Analysis

Statistical analysis was performed with Student's t test. P values <0.05 were considered statistically significant. **Figure 4A** was generated and analyzed using GraphPad Prism version 8 (GraphPad Software Inc.).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The human experiments were reviewed and approved by the Ethics Committee of Changzhou No 3 People's Hospital. The patients/participants provided their written informed consent to participate in this study. The mouse and guinea pig experiments were approved by the Committee of Shanghai Public Health Centre and the cattle experiments were approved by Ethic Committee of Southwest University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

SJ, WL, and JW designed the study and wrote the manuscript; QZ performed most of the experiments and carried out data analysis; XL performed the cow skin test and ELISPOT assays; ST helped to carry out the HLA analysis; DC performed the purification of ROP-TB; YG and LG helped to collect patient samples; RZ participated in the experimental design and management. All authors contributed to the article and approved the submitted version.

FUNDING

The authors declare that this study received partial funding from Oxford Vacmedix UK Ltd and CBI. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

This research was also supported by National Key Research and Development Program of China (2018YFE0192500) and the Key Research and Development Program of Changzhou (CE20135023, CE20185016). SJ is funded by Innovate UK.

ACKNOWLEDGMENTS

We thank all the study participants.

REFERENCES

1. Burgdorf S, Kautz A, Bohnert V, Knolle PA, Kurts C. Distinct Pathways of Antigen Uptake and Intracellular Routing in CD4 and CD8 T Cell Activation. *Science* (2007) 316:612–6. doi: 10.1126/science.1137971
2. Burgdorf S, Scholz C, Kautz A, Tampe R, Kurts C. Spatial and Mechanistic Separation of Cross-Presentation and Endogenous Antigen Presentation. *Nat Immunol* (2008) 9:558–66. doi: 10.1038/ni.1601
3. Ackerman AL, Cresswell P. Cellular Mechanisms Governing Cross-Presentation of Exogenous Antigens. *Nat Immunol* (2004) 5:678–84. doi: 10.1038/ni1082
4. World Health Organization (WHO). Commercial Serodiagnostic Tests for Diagnosis of Tuberculosis: Policy Statement. Geneva: World Health Organization. (2011).
5. Organisation WH. WHO Warns Against the Use of Inaccurate Blood Tests for Active Tuberculosis. *J Mater Sci* (2011) 29(5):1341–7. doi: 10.1007/BF00975086
6. Steingart KR, Henry M, Laal S, Hopewell PC, Ramsay A, Menzies D, et al. A Systematic Review of Commercial Serological Antibody Detection Tests for the Diagnosis of Extrapulmonary Tuberculosis. *Postgrad Med J* (2007) 83:705–12. doi: 10.1136/thx.2006.075754
7. World Health Organization (WHO)Guidelines on the Management of Latent Tuberculosis Infection. Geneva: World Health Organization (2015).
8. Vesenbeckh SM, Schonfeld N, Mauch H, Bergmann T, Wagner S, Bauer TT, et al. The Use of Interferon Gamma Release Assays in the Diagnosis of Active Tuberculosis. *Tuberc Res Treat* (2012) 2012:768723. doi: 10.1155/2012/768723
9. Campbell JR, Krot J, Elwood K, Cook V, Marra F. A Systematic Review on TST and IGRA Tests Used for Diagnosis of LTBI in Immigrants. *Mol Diagn Ther* (2015) 19:9–24. doi: 10.1007/s40291-014-0125-0
10. Cai L, Zhang J, Zhu R, Shi W, Xia X, Edwards M, et al. Protective Cellular Immunity Generated by Cross-Presenting Recombinant Overlapping Peptide Proteins. *Oncotarget* (2017) 8:76516–24. doi: 10.18632/oncotarget.20407
11. Zhang H, Hong H, Li D, Ma S, Di Y, Stoten A, et al. Comparing Pooled Peptides With Intact Protein for Accessing Cross-Presentation Pathways for Protective CD8+ and CD4+ T Cells. *J Biol Chem* (2009) 284:9184–91. doi: 10.1074/jbc.M809456200
12. Simhadri VL, McGill JR, Sauna ZE. Endotoxin Contamination in Commercially Available Cas9 Proteins Potentially Induces T-Cell Mediated Responses. *Gene Ther* (2021). doi: 10.1038/s41434-021-00301-6
13. Praud A, Bourely C, Boschiroli ML, Dufour B. Assessment of the Specificity of a Gamma-Interferon Test Performed With Specific Antigens to Detect Bovine Tuberculosis, After non-Negative Results to Intradermal Tuberculin Testing. *Vet Rec Open* (2019) 6:e000335. doi: 10.1136/vetreco-2019-000335
14. Buddle BM, Keen D, Thomson A, Jowett G, McCarthy AR, Heslop J, et al. Protection of Cattle From Bovine Tuberculosis by Vaccination With BCG by the Respiratory or Subcutaneous Route, But Not by Vaccination With Killed Mycobacterium Vaccae. *Res Vet Sci* (1995) 59:10–6. doi: 10.1016/0034-5288(95)90023-3
15. Flores-Villalva S, Suarez-Guemes F, Espitia C, Whelan AO, Vordermeier M, Gutierrez-Pabello JA. Specificity of the Tuberculin Skin Test is Modified by Use of a Protein Cocktail Containing ESAT-6 and CFP-10 in Cattle Naturally Infected With Mycobacterium Bovis. *Clin Vaccine Immunol* (2012) 19:797–803. doi: 10.1128/CVI.05668-11
16. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma Interferon Release Assays for Detection of Mycobacterium Tuberculosis Infection. *Clin Microbiol Rev* (2014) 27:3–20. doi: 10.1128/CMR.00034-13
17. Pai M, Zwerling A, Menzies D. Systematic Review: T-Cell-Based Assays for the Diagnosis of Latent Tuberculosis Infection: An Update. *Ann Intern Med* (2008) 149:177–84. doi: 10.7326/0003-4819-149-3-200808050-00241

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.872676/full#supplementary-material

Supplementary Figure 1 | ELISPOT responses to ROP-TB spiked with different concentrations of endotoxin in two *Mtb*-infected patients and two healthy donors. Negative control: culture medium (R10); stimulant: ROP-TB (5 μg/well); endotoxin with 3 different doses (33, 6.6 and 3.3 EU/mL); Positive control: PHA (5 μg/well).

Conflict of Interest: QZ, XL, YG, DC, RZ, WL and SJ are employees and/or shareholders of Oxford Vacmedix (Changzhou) Ltd which develops TB diagnostic kits. WL is employed by Shanghai JW Inflinhix Ltd (a subsidy company of Oxford Vacmedix(Changzhou)).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

OPINION
published: 20 June 2022
doi: 10.3389/fmed.2022.903876

Depending on Epitope Profile of COVID-19 mRNA Vaccine Recipients: Are They More Efficient Against the Arising Viral Variants? An Opinion Article

Nawal Abd El-Baky and Amro Abd Al Fattah Amara**

Department of Protein Research, Genetic Engineering and Biotechnology Research Institute (GEBRI), City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt

OPEN ACCESS

Edited by:
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

Reviewed by:
Waleed Mahallawi,
Taibah University, Saudi Arabia
Xingui Tian,
First Affiliated Hospital of Guangzhou Medical University, China
Utpal Kumar Adhikari,
Western Sydney University, Australia

***Correspondence:**
Nawal Abd El-Baky
nelbaky@srtacity.sci.eg;
nawalabdelbaky83@gmail.com
Amro Abd Al Fattah Amara
aamara@srtacity.sci.eg;
amroamara@web.de

Specialty section:
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *24 March 2022*
Accepted: *01 June 2022*
Published: *20 June 2022*

Citation:
Abd El-Baky N and Amara AAAF (2022) Depending on Epitope Profile of COVID-19 mRNA Vaccine Recipients: Are They More Efficient Against the Arising Viral Variants? An Opinion Article. Front. Med. 9:903876. doi: 10.3389/fmed.2022.903876

Keywords: COVID-19, epitopes, mRNA vaccine, neutralizing antibodies, RBD, spike, subunit vaccine, viral variant

INTRODUCTION

Since the emergence of COVID-19 pandemic caused by SARS-CoV-2 in December 2019 in China and its prompt spread by March 2020 globally, the world is suffering from a vast public health crisis (1–4). As stated by the World Health Organization (WHO) and real-time data from Worldometer, the reported cases of COVID-19 exceeded 418 million cases worldwide, of which, 5.85 million deaths as of 17th February 2022 (https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/). Vaccines of different types, immunization procedures and routes, and from different countries and companies unexpectedly started to appear by the end of 2020 and were approved for marketing worldwide (5–8). Nevertheless, understanding the antigen epitopes in all of these vaccines is still poor.

The public debate began, which vaccine to choose, which one is more effective and with less adverse reactions, besides the new type of vaccines [messenger RNA (mRNA)-based vaccines] added more factors to this debate. Millions worldwide still do not accept receiving the COVID-19 mRNA vaccines and fear the risk of their adverse events following immunization (AEFI) especially with the unavailability of convincingly enough data for clinical experimentation of these vaccines and long-term surveillance of their safety on recipients of different medical conditions, age, and gender. Furthermore, reports indicated that even mRNA vaccines offer high short-term protection against SARS-CoV-2 (9). On the other hand, both scientific and health communities urgently suggest that mRNA vaccination is incontestably the most effective way to provide population immunity (5–8, 10) and should be spread worldwide to put an end to the pandemic and return to normal life. The next debate is about the coverage of the continuously arising and spreading SARS-CoV-2 variants.

Neutralizing antibodies with high potency are an effective line of defense in preventing and inhibiting viral infections via interrupting viral attachment to its target cell. This ability of neutralizing antibodies depends on their specific recognition of epitopes. Therefore, knowing the interaction between antibodies and epitopes provides a basis for the rational design of vaccines. After the introduction of recombinant DNA technologies, the concept of vaccination changed from immunization with inactivated intact pathogens (full spectrum of pathogen antigens) to a new concept in which isolated antigens are used for immunization. The vaccines prepared based on the last concept are designated subunit vaccines (11). Identifying an effective subunit vaccine provides the advantage of identifying the specific epitopes responsible for the protection, then immunize

with the precise epitope, corresponding to an effective neutralizing antibody. Theoretically, to identify a desirable epitope, the simplest way is to isolate this epitope-binding monoclonal antibody. Then, identifying a highly potent neutralizing monoclonal antibody, and immunizing an individual with the epitope of that monoclonal antibody. Additionally, if the monoclonal antibody can significantly cross react, it can bind a variety of genetic variants of the pathogen it neutralizes, thus immunized individuals with the epitope of this antibody will have the ability to broadly cross neutralize different variants (12).

In the current article, we aimed to outline the available data regarding the epitope profile of mRNA-based vaccine recipients in comparison to subunit vaccines and naturally infected individuals. Besides, concluding if these data confirm the proposed superior efficacy of mRNA-based vaccine-induced immunity over subunit vaccines- and natural infection-induced immunity against SARS-CoV-2 variants.

SARS-CoV-2 AND ITS VARIANTS

SARS-CoV-2 and SARS-CoV share the genome structure, which is a single-stranded positive-sense (+ss) RNA genome with a length of 29.8–29.9 kb, comprising two large ORFs that encode the polyproteins, four structural protein genes that encode envelope, membrane, spike, and nucleocapsid proteins, as well as some accessary protein genes (13, 14). For SARS-CoV-2 to enter its host cells and initiate infection, the viral spike glycoprotein interacts with the angiotensin-converting enzyme 2 (ACE2) receptor on host cells and fuses with cellular membranes (15). The surface-exposed location of the spike protein not only permits it to achieve fusion with cellular membranes but also makes it a direct and major target for host immune responses (host neutralizing antibodies) (15). Thus, the spike protein or its structural domains [e.g., receptor-binding domain (RBD)] is the primary target for eliciting potent neutralizing antibodies, and design of vaccines along with antiviral therapeutics (16–19).

The SARS-CoV-2 virus carries a high mutation rate and transmission capacity when compared to other coronaviruses. Variants of SARS-CoV-2 are classified as variant being monitored (VBM) including Alpha (B.1.1.7 and Q lineages), Beta (B.1.351 and descendent lineages), Epsilon (B.1.427 and B.1.429), Eta (B.1.525), Gamma (P.1 and descendent lineages), Kappa (B.1.617.1), Iota (B.1.526), Mu (B.1.621, B.1.621.1), Zeta (P.2), and 1.617.3, variant of interest (VOI), and variant of concern (VOC) including Delta (B.1.617.2 and AY lineages) and Omicron (B.1.1.529 and BA lineages) (20–25).

Some SARS-CoV-2 variants have mutations in their spike protein, which improve the affinity between virus and the ACE2 receptor on host cells, leading to accelerated viral transmission and substantial or complete escape from neutralizing antibodies (26–28). Furthermore, single amino acid substitutions (e.g., D429A, R441A or D454A) or deletions of numerous amino acids at the RBD C-terminal or N-terminal region were reported to completely disrupt most neutralizing epitopes in spike RBD of SARS-CoV (29). In case of SARS-CoV-2, mutations at sites 473, 475, and 476 in spike RBD are already found in a small number of viral variants, and they may have the potential to decrease the binding affinity and effectiveness of neutralizing antibodies. Meanwhile, the mutations in epitope $_{473}$YQAGSTP$_{479}$ found in RBM are beneficial for the virus in some unrevealed way (30).

EPITOPE PROFILING IN RBD-BASED PROTEIN SUBUNIT VACCINE

At present, vaccine-induced protection against SARS-CoV-2 in an individual is estimated by humoral responses, for example total antibodies titers against the viral spike RBD and half-maximal neutralization titers (NT50s) using pseudotyped or live viruses (5, 6, 8, 10). Although humoral immune responses in clinical trials for various COVID-19 vaccine candidates have been described, the absence of standardized neutralization assays makes comparing different vaccine candidates a difficult task. Yang et al. (31) carried out two randomized, double-blind, placebo-controlled, phase 1 and phase 2 trials for a recombinant tandem-repeat dimeric RBD-based protein subunit vaccine (ZF2001) against SARS-CoV-2. Their study was the first to report phase 1 and phase 2 clinical data for RBD-based protein subunit COVID-19 vaccine that comprises an alum adjuvant. In phase 1, they measured vaccine safety by the occurrence of adverse events and serious adverse events. In phase 2, they evaluated vaccine safety as well as immunogenicity by the seroconversion rate and geometric mean titers [GMTs] of neutralizing antibodies against SARS-CoV-2. They found that ZF2001 is both well tolerated and immunogenic. Phase 3 clinical trials are continuing to further explore ZF2001 safety and protective efficacy (31).

Mapping the epitope profile of antibodies elicited by both vaccines and natural infection is the key step in assessing vaccine functionality and advantage of vaccine-induced immunity. Besides, this mapping can elucidate the molecular basis of humoral immune responses. Furthermore, the evolution of vaccine-induced immune escape if considered in conjunction with viral virulence will help achieve herd immunity against COVID-19. Yang et al. (32) reported protective immunity induced by a vaccine targeting SARS-CoV-2 spike protein RBD in three animal species. They stated that antibodies induced by the vaccine and those found in the sera of patients with COVID-19 share common binding epitopes. They also observed that aluminum adjuvant can further enhance the immune response induced by the vaccine, a single vaccine dose might elicit a high level of virus-neutralizing activity, the toxicology studies in the non-human primates confirmed vaccine safety, and that vaccination protected non-human primates against an *in vivo* challenge with SARS-CoV-2 (32).

Epitope profiling in RBD-based antigens of SARS-CoV-2 revealed the critical antigenic determinants, which include three immunodominant epitopes, a highly conserved epitope ($_{350}$VYAWN$_{354}$) located exposed on the surface of viral spike protein trimer, a variable epitope among different viral strains ($_{473}$YQAGSTP$_{479}$) found in the receptor binding motif (RBM), and a highly conserved cryptic cross-reactive epitope ($_{407}$VRQIAP$_{412}$) shared between RBD of SARS-CoV-2 and

SARS-CoV (30). These data can elucidate the humoral immune responses to the viral spike protein RBD and may enable design of new anti-SARS-CoV-2 vaccines.

EPITOPE PROFILING IN COVID-19 mRNA VACCINES

With the appearance of SARS-CoV-2 variants that harbor mutations in key epitopes, the risk of eroding adaptive immunity elicited by either vaccination or prior infection as a result of this antigenic evolution increased. Since the breadth of epitopes targeted by vaccine-induced antibodies or natural infection-induced antibodies partially contributes to susceptibility to erosion by viral evolution, the specificity of polyclonal antibodies induced by natural infection or Moderna mRNA-1273 COVID-19 vaccine (two doses) was compared by deep mutational scanning (33). The results showed that vaccine-elicited antibodies have more targeted neutralizing activity to SARS-CoV-2 spike protein RBD than natural infection-elicited antibodies. Additionally, vaccine-elicited antibodies demonstrated greater binding breadth across multiple RBD epitopes compared to infection-elicited antibodies. Thus, single mutations in RBD seemed to have less impact on antibody immunity acquired by Moderna mRNA-1273 COVID-19 vaccine compared to prior infection (33).

Wisnewski et al. (34) mapped immunogenic amino acid motifs and linear epitopes of primary sequence of SARS-CoV-2 spike protein that induce IgG in recipients of Pfizer-BioNTech COVID-19 mRNA vaccine. The obtained data identified various distinctive amino acid motifs recognized by vaccine-elicited IgG, a subset of those recognized by IgG from natural infection (hospitalized COVID-19 patients), which can mimic 3-dimensional conformation (mimotopes). The identified dominant linear epitopes in the C-terminal region of the spike protein subunits include amino acids 558–569, 627–638, and 1,148–1,159 (34). These epitopes of COVID-19 mRNA vaccine are identical with those of SARS-CoV, bat coronavirus, and epitopes that trigger IgG during natural infection, but have limited homology to spike protein of non-pathogenic human coronavirus. The identified epitopes in COVID-19 mRNA vaccine may form the basis for further research of immune escape, viral variants, and design of vaccine and therapy.

Amanat et al. (35) revealed that polyclonal antibody responses after SARS-CoV-2 mRNA vaccination target three epitopes within viral spike protein; RBD, the N-terminal domain (NTD), and S2 subunit. They also highlighted that antibody responses after vaccination were comparable to or exceeded those after natural infection, mRNA vaccination induced a high rate of antibodies with no neutralizing activity, vaccination induced cross-reactive antibodies to seasonal human coronaviruses HKU1 and OC43, and that a proportion of vaccine-induced antibodies with binding ability to RBD can offer substantial protection against viral variants carrying single E484K RBD mutations.

Nitahara et al. (36) performed high-resolution linear epitope profiling of recipients of Pfizer-BioNTech COVID-19 mRNA vaccine and COVID-19 patients and found that vaccine-induced antibodies targeting viral spike RBD have a broader distribution across RBD than natural infection-induced antibodies. Moreover, mutation panel assays targeting the viral variants of concern demonstrated that the epitope variety induced by mRNA vaccine is rich in breadth, thus can grant resistance against viral evolutionary escapes in future, which represents an advantage of vaccine-induced immunity.

Collectively, these studies suggested that mRNA vaccination against SARS-CoV-2 elicited antibodies targeting viral spike RBD that have a broader distribution across RBD than natural infection-induced antibodies, which seem to offer more resistance against future SARS-CoV-2 evolutionary escapes. **Table 1** summarizes the identified epitopes used in mRNA and subunit COVID-19 vaccines and those in natural infection.

IMPACT OF VIRAL VARIANTS ON mRNA AND SUBUNIT VACCINES-INDUCED IMMUNITY

BNT162b2 mRNA vaccine was reported to elicit neutralization of spike glycoproteins of B.1.617.1, B.1.617.2, B.1.618 Indian SARS-CoV-2 variants or B.1.525 (Nigeria lineage) after 2–4 weeks after second dose (38). However, neutralization titres against the variants (especially the B.1.617.1 variant) were lower than that against the original virus. In addition, the durability of neutralization titres against the variants was not examined. These variants with mutations in spike glycoproteins were reported to have the potential to change neutralization by influencing spike function instead of antigenicity, despite the fact that these variants revealed similar specific infectivity and infectious titres to the original virus. In a more recent study, it was found that neutralizing titers were reduced more than 22-fold against variant of concern Omicron (B.1.1.529) compared with Wuhan-neutralizing titers after two doses of BNT162b2 (39). One month after the third dose, neutralizing titers to Omicron were increased 23-fold in relation to those after two doses and were comparable to Wuhan-neutralizing titers after two doses. These findings suggested the need for three doses (vaccine boosters) of BNT162b2 mRNA vaccine to protect against Omicron variant.

The development of humoral and T cell responses against ancesteral SARS-CoV-2 and B.1.351, B.1.617.2, and P.1 variants that carry specific mutations in the spike gene was analyzed in previously infected (recovered) or uninfected (naive) individuals who received mRNA SARS-CoV-2 vaccines (40). Previously infected individuals sustained higher antibody titres against both the original virus and variants than uninfected ones, yet uninfected post-vaccination individuals reached equivalent neutralization against the original virus after second dose. Overall, the neutralization capacity was reduced against the variants, but B.1.351 and P.1 had the greatest reduction. These data pointed to vaccine boosters as an approach to alleviate the effect of continously emerging new SARS-CoV-2 variants with mutations in major neutralizing antibody-binding sites on the longevity of vaccine-induced immune protection. *In vitro* stimulation analysis using spike or nucleocapsid peptide

TABLE 1 | Comparison between the identified epitopes in mRNA and subunit COVID-19 vaccines and those in natural infection.

Subjects	Identified epitopes	Epitopes position	Reference
Naturally infected individuals with disease severity of mild non-hospitalized, moderate hospitalized, and severe admitted to the intensive care unit	155 AAIVLQLPQGTTLPKGF 171	Epitope 155–171 from N protein	(37)
RBD-based antigens (subunit vaccine)	350 VYAWN 354	Exposed on the surface of viral spike protein trimer	(30)
	473 YQAGSTP 479	In the receptor binding motif	
	407 VRQIAP 412	Shared between RBD of SARS-CoV-2 and SARS-CoV	
mRNA vaccine and natural infection	Amino acids 558–569, 627–638, and 1148–1159	C-terminal region of the spike protein subunits	(34)
BNT162b2 mRNA vaccine and natural infection	N394 to A411, T415 to F429, V433 to N450, R457 to S477, recognized by both vaccine recipient and patient sera N334 to A348, S373 to L390, S514 to F541, recognized only by vaccine-elicited sera Substituting amino acids K417N, K417T, E484K, and N501Y, recognized only by vaccine-elicited sera	Within the RBD Single amino acid mutations of RBD in B.1.1.7, B.1.351, and P.1 variants	(36)

revealed that T cell activation markers progressively increased after vaccination.

Extensive neutralization against Delta and Omicron variants elicited by Omicron-specific subunit vaccine booster is another example for the importance of booster doses to face the surge of variants and ensure that SARS-CoV-2 variants carrying a great number of mutations will not escape vaccine-mediated protection (41). A third dose of mRNA vaccine could provide strong (91%) protection against Delta SARS-CoV-2 variant in frontline workers (42). ZF2001 subunit vaccine booster following two doses of inactivated vaccines was found to increase the neutralizing antibody titer against SARS-CoV-2 and its variants in mice, particularly against Delta variant (43). This heterologous booster vaccine was recommended for future immunization program. Heterologous booster of recombinant protein subunit vaccine after two doses of inactivated whole-virion vaccines was evidenced to be highly immunogenic and safe for healthy adults and could significantly increase anti-RBD responses and neutralizing titers against SARS-CoV-2 and its variants comprising B.1.617.2 (Delta variant), B.1.351, B.1.1.7, and P.1 (44).

While COVID-19 vaccines are ideally designed to elicit neutralizing antibodies against viral spike protein, the mRNA vaccines approved so far provided efficient protection even after only one dose, when just non-neutralizing antibodies and moderate responses of T helper 1 cell are detectable, whereas almost no neutralizing antibodies found (45). These data recommended that vaccine-mediated protection probably needs low neutralizing antibodies levels and might involve other effectors such as T cells, non-neutralizing antibodies, and innate immune mechanisms. T cell responses prime early and contribute to protection but are comparatively impaired in severe disease leading to intense activation and lymphopenia (46). T cell memory incorporates extensive recognition of SARS-CoV-2 proteins, determined at about 30 epitopes/individual, and is well-sustained to date. This extensive recognition may limit the impact of individual viral mutations and can offer protection against severe disease caused by viral variants, even Omicron (46). The cellular immune response induced by BNT162b2 mRNA vaccine among healthcare workers in Bulgaria was investigated in the course of 16 weeks after first dose (47). One month after completing immunization, the number of virus-specific T cells producing IFNγ significantly correlated with virus-neutralizing activity and RBD-specific IgA levels induced after first dose. However, detection of T cells producing IFNγ may need longer stimulation beyond the first month after completing immunization (47).

FINAL CONSIDERATIONS

In this opinion article, we discussed the epitope profile of COVID-19 mRNA-based vaccine recipients in comparison to subunit vaccines and naturally infected individuals and subsequent superior efficacy of mRNA-based vaccine-induced immunity against SARS-CoV-2 variants. To the best of our knowledge, there are currently no clear and sufficient published data concerning the antigen epitopes in each type of marketed COVID-19 vaccines, or available studies that compare the efficiency and functionality of inactivated vaccine-, subunit vaccine-, viral-vectored vaccine-, and mRNA vaccine-induced immunity against either SARS-CoV-2 or its variants based on epitope profile. However, some preliminary evidences point that mRNA vaccination against SARS-CoV-2 has induced antibodies targeting viral spike RBD that have a broader distribution across RBD than natural infection-induced antibodies, which seem to offer more resistance against future SARS-CoV-2 evolutionary escapes. Studies to comprehensively assess the epitope variety induced by each COVID-19 vaccination have yet to be conducted to raise public confidence in COVID-19 vaccines, promote their

functionality and advantage of vaccine-induced immunity in spite of associated AEFI. This research if conducted would have tremendous impact on vaccination strategies to face the pandemic and viral variants of concern. Moreover, the method applied to comprehensively assess the epitope variety for vaccination against viral variants should be carefully considered.

AUTHOR CONTRIBUTIONS

NA and AA drafted, critically revised the manuscript, and agree to be accountable for the content of the work. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China. *N Engl J Med*. (2019) 382:727–33. doi: 10.1056/NEJMoa2001017
2. Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. *Viruses*. (2020) 12:135 doi: 10.3390/v12020135
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with. (2019 novel coronavirus in Wuhan, China. *Lancet*. (2020) 395:497–506. doi: 10.1016/S0140-6736(20) 30183-5
4. Mahase E. Covid-19: WHO declares pandemic because of "Alarming Levels" of spread, severity, and inaction. *BMJ*. (2020) 368:m1036. doi: 10.1136/bmj.m1036
5. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*. (2020) 586:589–93. doi: 10.1038/s41586-020-2639-4
6. Walsh EE, Frenck RW, Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. Safety and immunogenicity of two RNA-based COVID-19 vaccine candidates. *N Engl J Med*. (2020) 383:2439–50. doi: 10.1056/NEJMoa202 7906
7. Chen M, Yuan Y, Zhou Y, Deng Z, Zhao J, Feng F. et al. Safety of SARS-CoV-2 vaccines: A systematic review and meta-analysis of randomized controlled trials. *Infect Dis Poverty*. (2021) 10:94. doi: 10.1186/s40249-021-00878-5
8. Muller L, Andrée M, Moskorz W, Drexler I, Walotka L, Grothmann R, et al. Age-dependent immune response to the Biontech/Pfizer BNT162b2 coronavirus disease. 2019 vaccination. *Clin Infect Dis*. (2021) 73:2065–72 doi: 10.1093/cid/ciab381
9. Hall V, Foulkes S, Insalata F, Kirwan P, Saei A, Atti A, et al. Protection against SARS-CoV-2 after COVID-19 vaccination and previous infection. *NEJM*. (2022) 386:1207–20 doi: 10.1056/NEJMoa2118691
10. Ebinger JE, Fert-Bober J, Printsev I, Wu M, Sun N, Prostko JC, et al. Antibody responses to the BNT162b2 mRNA vaccine in individuals previously infected with SARS-CoV-2. *Nat Med*. (2021) 27:981–4. doi: 10.1038/s41591-021-01325-6
11. Babai I, Samira S, Barenholz Y, Zakay-Rones Z, Kedar E. A novel influenza subunit vaccine composed of liposome-encapsulated haemagglutinin/neuraminidase and IL-2 or GM-CSF. I. Vaccine characterization and efficacy studies in mice. *Vaccine*. (1999) 17:1223–38 doi: 10.1016/S0264-410X(98)00346-6
12. Gershoni JM, Roitburd-Berman A, Siman-Tov DD, Tarnovitski Freund N, Weiss Y. Epitope mapping: the first step in developing epitope-based vaccines. *BioDrugs*. (2007) 21:145–56. doi: 10.2165/00063030-200721030-00002
13. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, et al. From SARS to MERS, Thrusting coronaviruses into the spotlight. *Viruses*. (2019) 11:59. doi: 10.3390/v11010059
14. Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep*. (2020) 19:100682. doi: 10.1016/j.genrep.2020.100682
15. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. (2020) 181:281–92.e6 doi: 10.1016/j.cell.2020.11.032
16. Huo J, Le Bas A, Ruza RR, Duyvesteyn HME, Mikolajek H, Malinauskas T, et al. Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat Struct Mol Biol*. (2020) 27:846–54. doi: 10.1038/s41594-020-0469-6
17. Ju B, Zhang Q, Ge J, Wang R, Sun J, Ge X, et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*. (2020) 584:115–9. doi: 10.1038/s41586-020-2380-z
18. Shi R, Shan C, Duan X, Chen Z, Liu P, Song J, et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature*. (2020) 584:120–24. doi: 10.1038/s41586-020-2381-y
19. Tortorici MA, Beltramello M, Lempp FA, Pinto D, Dang HV, Rosen LE, et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science*. (2020) 370:950–7. doi: 10.1126/science.abe3354
20. Edara VV, Floyd K, Lai L, Gardner M, Hudson W, Piantadosi A, et al. Infection and mRNA-1273 vaccine antibodies neutralize SARS-CoV-2 UK variant. *MedRxiv*. (2021) 2021.02.02.21250799. doi: 10.1101/2021.02.02.21250799
21. Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. (2021) 184:2372-2383.e9. doi: 10.1016/j.cell.2021.03.013
22. Madhi SA, Baillie V, Cutland CL, Voysey M, Koen AL, Fairlie L. et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B1351 Variant. *N Engl J Med*. (2021) 384:1885–98. doi: 10.1056/NEJMoa2102214
23. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*. (2021) 184:64–75.e11 doi: 10.1016/j.cell.2020.11.020
24. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. (2021) 593:130–35. doi: 10.1038/s41586-021-03398-2
25. Wang P, Casner RG, Nair MS, Wang M, Yu J, Cerutti G, et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *bioRxiv*. (2021) 2021.03.01.433466. Update in: Cell Host Microbe. (2021) 29:747–51.e4. doi: 10.1016/j.chom.2021.04.007
26. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29,. 2020-January 12,. 2021. *MMWR Morb Mortal Wkly Rep*. (2021) 70:95–9 doi: 10.15585/mmwr.mm7003e2
27. Tang JW, Toovey OTR, Harvey KN. Hui DDS. Introduction of the South African SARS-CoV-2 variant 501YV2 into the UK. *J Infect*. (2021) 82:e8–e10. doi: 10.1016/j.jinf.2021.01.007
28. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B. et al. SARS-CoV-2 501YV2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med*. (2021) 27:622–5. doi: 10.1038/s41591-021-01285-x
29. He Y, Li J, Du L, Yan X, Hu G, Zhou Y, et al. Identification and characterization of novel neutralizing epitopes in the receptor-binding domain of SARS-CoV spike protein: revealing the critical antigenic determinants in inactivated SARS-CoV vaccine. *Vaccine*. (2006) 24:5498–508. doi: 10.1016/j.vaccine.2006.04.054
30. Jiang M, Zhang G, Liu H, Ding P, Liu Y, Tian Y, et al. Epitope profiling reveals the critical antigenic determinants in SARS-CoV-2 RBD-based antigen. *Front Immunol*. (2021) 12:707977. doi: 10.3389/fimmu.2021.707977
31. Yang S, Li Y, Dai L, Wang J, He P, Li C, et al. Safety and immunogenicity of a recombinant tandem-repeat dimeric RBD-based protein subunit vaccine (ZF2001) against COVID-19 in adults: two randomised, double-blind, placebo-controlled, phase 1 and 2 trials. *Lancet Infect Dis*. (2021) 21:1107–19. doi: 10.1016/S1473-3099(21)00127-4
32. Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, et al. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature*. (2020) 586:572–77. doi: 10.1038/s41586-020-2599-8
33. Greaney AJ, Loes AN, Gentles LE, Crawford KHD, Starr TN, Malone KD, et al. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. *Sci Transl Med*. (2021) 13:eabi9915. doi: 10.1126/scitranslmed.abi9915

34. Wisnewski AV, Redlich CA, Liu J, Kamath K, Abad QA, Smith RF, et al. Immunogenic amino acid motifs and linear epitopes of COVID-19 mRNA vaccines. *PLoS ONE.* (2021) 16:e0252849. doi: 10.1371/journal.pone.0252849
35. Amanat F, Thapa M, Lei T, Ahmed SMS, Adelsberg DC, Carreño JM, et al. SARS-CoV-2 mRNA vaccination induces functionally diverse antibodies to NTD, RBD, and S2. *Cell.* (2021) 184:3936–48.e10 doi: 10.1016/j.cell.2021.06.005
36. Nitahara Y, Nakagama Y, Kaku N, Candray K, Michimuko Y, Tshibangu-Kabamba E, et al. High-resolution linear epitope mapping of the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 mRNA vaccine recipients. *Microbiol Spectr.* (2021) 9:e0096521. doi: 10.1128/Spectrum.00965-21
37. Musicò A, Frigerio R, Mussida A, Barzon L, Sinigaglia A, Riccetti S, et al. SARS-CoV-2 epitope mapping on microarrays highlights strong immune-response to N protein region. *Vaccines (Basel).* (2021) 9:35. doi: 10.3390/vaccines9010035
38. Liu J, Liu Y, Xia H, Zou J, Weaver SC, Swanson KA, et al. BNT162b2-elicited neutralization of B.1.617 and other SARS-CoV-2 variants. *Nature.* (2021) 596: 273–275 doi: 10.1038/s41586-021-03693-y
39. Muik A, Lui BG, Wallisch AK, Bacher M, Mühl J, Reinholz J, et al. Neutralization of SARS-CoV-2 Omicron by BNT162b2 mRNA vaccine-elicited human sera. *Science.* (2022) 375:678–80. doi: 10.1126/science.abn7591
40. Lucas C, Vogels CBF, Yildirim I, Rothman JE, Lu P, Monteiro V, et al. Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature.* (2021) 600:523–529 doi: 10.1038/s41586-021-04085-y
41. Peng P, Feng C, Hu J, He C-L, Deng H-J, Fan Q, et al. Extensive neutralization against SARS-CoV-2 variants elicited by Omicron-specific subunit vaccine booster. *bioRxiv.* (2022). 2022.03.07.483373. doi: 10.1101/2022.03.07.483373
42. Yoon SK, Hegmann KT, Thiese MS, Burgess JL, Ellingson K, Lutrick K, et al. HEROES-RECOVER Network Investigators. Protection with a third dose of mRNA vaccine against SARS-CoV-2 variants in frontline workers. *N Engl J Med.* (2022) 386:1855–7. doi: 10.1056/NEJMc2201821
43. Zhang R, Li D, Xu K, Yang C, Luo T, Zhao X, et al. A protein subunit vaccine booster following two doses of inactivated SARS-CoV-2 vaccine provides high neutralisation of SARS-CoV-2 and its variants in mice. *Lancet Microbe.* (2021) 3:e165-6. doi: 10.1016/S2666-5247(21)00331-1
44. Ai J, Zhang H, Zhang Q, Zhang Y, Lin K, Fu Z, et al. Recombinant protein subunit vaccine booster following two-dose inactivated vaccines dramatically enhanced anti-RBD responses and neutralizing titers against SARS-CoV-2 and Variants of Concern. *Cell Res.* (2022) 32:103–6. doi: 10.1038/s41422-021-00590-x
45. Sadarangani M, Marchant A, Kollmann TR. Immunological mechanisms of vaccine-induced protection against COVID-19 in humans. *Nat Rev Immunol.* (2021) 21: 475–484 doi: 10.1038/s41577-021-00578-z
46. Moss P. The T cell immune response against SARS-CoV-2. *Nat Immunol.* (2022) 23: 186–93 doi: 10.1038/s41590-021-01122-w
47. Nikolova M, Todorova Y, Emilova R, Trifonova I, Gladnishka T, Petrova-Yancheva N, et al. Induction of humoral and cellular immune responses to COVID-19 mRNA and vector vaccines: a prospective cohort study in Bulgarian healthcare workers. *J Med Virol.* (2022) 94:2008–18. doi: 10.1002/jmv.27572

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Immunology

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fimmu.2022.865180

Immunoinformatics Aided Design and *In-Vivo* Validation of a Cross-Reactive Peptide Based Multi-Epitope Vaccine Targeting Multiple Serotypes of Dengue Virus

Vikas Kaushik[1], Sunil Krishnan G[1], Lovi Raj Gupta[1], Utkarsh Kalra[2,3], Abdul Rajjak Shaikh[2], Luigi Cavallo[4*] and Mohit Chawla[4*]*

[1] Domain of Bioinformatics, School of Bio-Engineering and Bio-Sciences, Lovely Professional University, Punjab, India, [2] Department of Research and Innovation, STEMskills Research and Education Lab Private Limited, Faridabad, India, [3] Department of Data Science, Innopolis University, Innopolis, Russia, [4] Kaust Catalysis Center, Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

OPEN ACCESS

Edited by:
Adriana Harbuzariu, Emory University, United States

Reviewed by:
Manas R Dikhit, Rajendra Memorial Research Institute of Medical Sciences, India
Philippe Auguste Robert, University of Oslo, Norway

***Correspondence:**
Abdul Rajjak Shaikh
razzaqsk@gmail.com
Luigi Cavallo
luigi.cavallo@kaust.edu.sa
Mohit Chawla
mohitchawla.bt@gmail.com;
Mohit.chawla@kaust.edu.sa

Specialty section:
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

Received: *29 January 2022*
Accepted: *05 May 2022*
Published: *21 June 2022*

Citation:
Kaushik V, G SK, Gupta LR, Kalra U, Shaikh AR, Cavallo L and Chawla M (2022) Immunoinformatics Aided Design and In-Vivo Validation of a Cross-Reactive Peptide Based Multi-Epitope Vaccine Targeting Multiple Serotypes of Dengue Virus. Front. Immunol. 13:865180. doi: 10.3389/fimmu.2022.865180

Dengue virus (DENV) is an arboviral disease affecting more than 400 million people annually. Only a single vaccine formulation is available commercially and many others are still under clinical trials. Despite all the efforts in vaccine designing, the improvement in vaccine formulation against DENV is very much needed. In this study, we used a roboust immunoinformatics approach, targeting all the four serotypes of DENV to design a multi-epitope vaccine. A total of 13501 MHC II binding CD4+ epitope peptides were predicted from polyprotein sequences of four dengue virus serotypes. Among them, ten conserved epitope peptides that were interferon-inducing were selected and found to be conserved among all the four dengue serotypes. The vaccine was formulated using antigenic, non-toxic and conserved multi epitopes discovered in the in-silico study. Further, the molecular docking and molecular dynamics predicted stable interactions between predicted vaccine and immune receptor, TLR-5. Finally, one of the mapped epitope peptides was synthesized for the validation of antigenicity and antibody production ability where the *in-vivo* tests on rabbit model was conducted. Our *in-vivo* analysis clearly indicate that the imunogen designed in this study could stimulate the production of antibodies which further suggest that the vaccine designed possesses good immunogenicity.

Keywords: immunoinformatic analysis, dengue (DENV), molecular docking & molecular dynamics (MD) simulation, vaccine design for emerging infections, multi epitope peptide vaccine, *in vivo* study

INTRODUCTION

Dengue is one of the most frequent arboviral diseases and presents a major health concern for people living in tropical and sub-tropical climate regions of the world and is present in over 125 countries (1). Mosquitoes of the genus *Aedes* are principally responsible for transmitting dengue virus (DENV) to human beings which is affecting 400 million people worldwide annually (2).

The DENV genome is composed of nearly ~ 11 kb of single and positive sense RNA, which encodes for a protein comprising 3391 amino acids (3, 4). DENV possesses four closely related serotypes (DENV 1-4) (5). However, a fifth serotype has also been reported recently (6). Remarkably four closely related dengue serotypes viz. DENV-1, DENV-2, DENV-3, and DENV-4 share 65-70% genome similarity (7).

Dengue disease ranges from mild self-limiting dengue fever (DF) to severe dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Unfortunately no drugs, vaccines or targeted therapies specifically for controlling dengue virus have been available till date (8). However proper intensive care is the only choice available in order to treat dengue. It is well known that vaccines are an effective way to control infectious diseases and also provide immunity against pathogens. The first and only commercialized dengue vaccine available till date is Dengvaxia (CYD-TDV) (9). Dengvaxia vaccine hesitancy started due to secondary heterotypic infection, long-lasting cross-protection, side effects and safety related issues (9, 10). Besides Dengvaxia, many other vaccine formulations against DENV are still under clinical trials (11) that includes TAK-003, which is a live attenuated tetravalent vaccine, containing both structural and non-structural proteins, and which is currently under phase III clinical trial and with a reported efficacy of 80% (12). TAK-003 which was shown to be effective against all four DENV serotypes, albeit producing low levels of antibodies for DENV3 serotype (13). Next potential vaccine includes a tetravalent DNA vaccine, which is also in phase III clinical trials (14). Another vaccine formulation, V180 that is a genetic construct of a tetravalent vaccine formulation, based on DENV envelope glycoprotein is under phase I clinical trials (11).

Despite all the current developments in vaccine design, improvement in vaccine design and formulation is essential, especially due to the tetravalent nature of DENV (11). Secondary infection with a different DENV serotype from primary infection could be even more lethal (15). Therefore, a serotype-specific cross-protection is an utmost requirement for an effective immunity (16, 17).

With the recent advancement in the field of bioinformatics, different strategies have been employed to design knowledge based vaccines using the immunoinformatics approach. Using an immunoinformatics approach, multi-epitope vaccines have been designed against various pathogens including Epstein–Barr virus and Chlamydia trachomatis virus species (18–20) that show a promising cellular and humoral response and have further been shown to work both *in vitro* and *in vivo* murine model.

Further, a plethora of immunoinformatics based studies are available in literature in the context of designing DENV vaccine where various research groups have been working on identification of epitopes from various structural and non-structural proteins from DENV virus (21–24). In particular, identification of highly conserved amino acid sequences for the entire DENV proteome was performed by Khan and coworkers (25) where they identified highly conserved and immune relevant DENV sequences that were common across the four serotypes and have direct implications in design of tetravalent vaccine (26). A very similar study by same group focused on identification of highly conserved and serotype specific DENV peptides that are potentially immune-relevant which may be relevant candidates for vaccine design as such sequences minimize the issue of altered peptide ligands (APLs) that are cross-reactive between the dengue serotypes (27). In one of our previous studies, we have also applied immunoinformatics approach and discovered the CD8+ epitope peptides namely TSEIQLTDY, IGIGILLTW and IAVGMVTLY from dengue virus envelope, specifically from DENV-1 serotype (28). Further, in one of our recent studies, epitope peptides specifically HTLWSNGVL and FTTNIWLKL from dengue non-structural, NS1 protein that were able to form a stable complex with MHC I HLA allele (29).

In the present work we have utilized immunoinformatics approach in order to design a candidate multi-epitope vaccine that could work against all the four serotypes of DENV 1-4 and is likely to be immunogenic. Recent experimental studies have clearly indicated the the role of dengue-specific CD4+ T cells to control dengue infection *in vivo*, which could be the potential target for dengue virus vaccine development (30, 31). Therefore, we have particularly focused on the prediction of CD4+ T-cell epitope peptides which follow two pathways to produce both T cell immunity and B cell immunity and altogether cell mediated and humoral immunity would be invoked. Suitable adjuvants and linkers were added during the construction of multi-epitope vaccine construct. The physiochemical properties along with structuiral and antigenicity properties were carefully evaluated for predicted multi-epitope vaccine construct, along with its interaction with immune receptor, TLR-5. Finally, one of the potential epitopes that was conserved among all the four DENV serotyoes was synthesized and has been tested *in-vivo* on a rabbit model.

METHODOLOGY

Protein Sequences Identification and Antigenicity Analysis

UniProtKB/Swiss-Prot database from https://www.uniprot.org/uniprot/ provides manually annotated, accurate, and reviewed protein data. The genome polyprotein sequences of DENV serotypes DENV-1, DENV-2, DENV-3, and DENV-4 having the UniProtKB Id: P33478, P07564, Q6YMS4, and Q58HT7 respectively were selected and retrieved from the UniProtKB/Swiss-Prot protein database. The antigenicity prediction server VaxiJen 2.0 (32) was used for the assessment of the antigenic protein sequence. This tool was based on auto cross-covariance (ACC) transformation on virus protein sequence dataset derived models predicted the antigenicity of the protein sequence (33).

MHC II Binding T Cell Epitope Prediction

The MHC II molecules have an important role in T cell antigen presentation and immunity. An artificial neural network-based NetMHCIIpan version 4.0 had been reported as a powerful tool to predict MHC II binding CD4+ T cell epitope peptides (34). This prediction tool was used parameters like Binding Affinity (BA) and Eluted Ligand mass spectrometry (EL) scores to find the epitope peptide at any length.

Interferon Inducing Epitope Prediction From CD4+Epitopes

To design the most effective subunit vaccine candidate, identification of epitopes with the ability of interferon-gamma induction from MHC-II binding epitopes was done using the IFN epitope server (35).

Mapping and Analysis of CD4+ T Cell Epitopes

The predicted MHC II binding CD4+ T cell epitopes were screened and analyzed for the identification of potential epitopes for efficacious vaccine design. The CD4+ T cell epitopes were mapped and analyzed by various epitope analysis tools. The conservancy of epitope peptides was analyzed by Immune Epitope Data Base (IEDB) analysis resource tool (36). This tool facilitates the identification of the linear degree of the conservancy of an epitope and epitope-based vaccine design (36). The peptide toxicity was predicted by the ToxinPred server (37). This tool predicted nontoxic epitopes from the antigenic epitope sequence dataset (37).

Physico-Chemical Properties of T Cell Epitopes

The predicted epitope was analyzed for physical/chemical properties using the ProtParam (38) algorithm of Expasy web server https://web.expasy.org/protparam/. The raw sequence of the epitope is used to compute the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) score. The calculated properties have been reported in **Table S1**, see **Supplementary File**.

Multi Epitope Vaccine Formulation

For the design and formulation of a multi-epitope vaccine, three interferon-inducing $CD4^+$ T cell epitope peptides were used (see Results and Discussion section). The epitope peptides were selected based on immunoinformatics epitope mapping. The protocol adopted in the current study is in line with the previous studies where an *in-silico* peptide vaccine formulation approach was already employed to design vaccines for Human Papilloma Virus (39), Herpes Simplex Virus type 1 and 2 (40) Dengue virus (29).The multi-epitope vaccine construct contains three interferon-inducing $CD4^+$ T cell epitopes, three adjuvants, and a linker. The VaxiJen server (32) was used to compute antigenicity and protparam (38) was used to compute physical and chemical parameters of epitope peptides, see **Table S1** and multi-epitope vaccine. Finally the solubility was predicted using SCRATCH protein predictor (41).

Molecular Modeling, Docking and Molecular Dynamic Simulations of Designed Multi-Epitope Vaccine With TLR-5 Receptor

To predict the 3D structure of multi-epitope vaccine and Human TLR-5 receptor molecule we used Alphafoldv2.0 program (42, 43). For docking of multi-epitope vaccine and TLR-5 receptor complex, the HADDOCK server (44) was used. Further, all the Molecular dynamic (MD) simulations were performed with the GROMACS 2019 simulation program (45). TLR5-vaccine was placed in a cubic box and solvated with TIP3P water molecules creating a solvent layer at least 12 Å thick. The Amber ff99SB-ILDN (46) force field was used to model the parameters of proteins. Charge was neutralized adding appropriate number of K^+ ions and extra K^+Cl^- ions were added to achieve a bulk ionic strength of 0.15 M using the Joung-Cheatham ion model (47). The simulation box contains 224932 water molecules, 607 K^+ ions and 601 CL^- ions. The total number of atoms in the system was 690620. The system was first minimized with 50000 steps of steepest descent method with 1000 kJ/mol nm^2 position restraint on protein heavy atoms. Further, minimization was carried out without any restraint on protein. Equilibration of each system was carried out in a phased manner. First, 100 ps NVT simulation was carried out with restraint on heavy atoms of the protein. It was followed by 100 ps NPT simulation with restraint on heavy atoms of the protein. Production simulations were performed using the NPT ensemble for 100 ns. Three different trials were carried out. The temperature was maintained at 300 K using velocity rescaling with a coupling time of 0.1 ps. The pressure was maintained at 1 atm for NPT simulations using a Parrinello–Rahman barostat (48) with a coupling time of 2 ps. Equations of motion were integrated using the leapfrog algorithm with a time step of 2.0 fs. The total electrostatic interactions were evaluated using the particle mesh Ewald (PME) summation (49). Coulomb and van der Waals cut-offs of 1.0 nm were employed. Periodic boundary conditions in all directions were employed to mimic the bulk behavior. Bond lengths with hydrogen were constrained with the LINCS algorithm (50). Coordinates have been collected in trajectory files every 10 ps. Trajectory processing and most of the analysis have been performed using the GROMACS tools. The visualization and molecular graphics images were created using the PYMOL (51) and VMD software (51). Graphs were plotted using Excel 2016.

Peptide Synthesis and Conjugation

One of the potential antigenic peptide with CKREKKLGEFGKAKG sequence was synthesized using the GenScript's microwave-based PepPower™ technology merging Solid-phase peptide synthesis (SPPS) and Liquid-phase peptide synthesis (LPP) and Microwave technology. The peptide concentration was measured by using NanoDrop Spectrophotometer absorbance at 280nm.The purity was measured by SDS-PAGE and antigen with an expected purity of ≥85%.The resulting peptide was added with a cysteine residue in the N –terminal and is conjugated with KLH (Keyhole Limpet Hemocyanin) protein using MBS (m-maleimidobenzoyl-N-hydroxysuccinimide ester) method.

In-Vivo Validation for Antibody Development and Assessment

KLH conjugate peptide was inoculated in New Zealand rabbits (n=2) following the PolyExpress immunization method. This method was used in anti-KRV envelope polyclonal antibody generation (52). The antibodies were retrieved and the purity of

the polyclonal antibody was determined performing SDS-PAGE. Further, concentration of the antibody was measured by NanoDrop Spectrophotometer at 280 nm. Afterwards, the antibody titer was determined with Indirect ELISA using IgG as the control. The antigen used was a synthesized peptide with a coating concentration of 4ug/ml, and 100µl/well in ELISA plate. The anti-rabbit horseradish peroxidase(HRP) conjugated IgG was used as the secondary antibody. The OD_{450nm} value was measured using an ELISA plate reader.

RESULTS AND DISCUSSION

Retrieved Protein Sequence and Analysis

For an efficacious vaccine against DENV, the requirement of potency against all four serotypes (DENV 1-4) should be established. Four sequences corresponding to the four serotypes were selected Dengue virus type 1(DENV-1), a strain fom Singapore/S275/1990; Dengue virus type 2(DENV-2), a strain from Jamaica/1409/1983; Dengue virus type 3 (DENV-3), a strain from Sri Lanka/1266/2000; and finally Dengue virus type 4 (DENV-4), a strain from Philippines/H241/1956 that were isolated from DENV epidemic countries. The VaxiJen identified for four protein sequences were antigenic with an antigenicity score >0.4. The antigenic protein sequence could be a potential vaccine candidate. The immunoinformatics approach adopted here is fast and cheaper than the current available experimental methods used in reverse vaccinology for the discovery of potential antigenic vaccine candidates (33, 53, 54).

MHC II Binding CD4+T Cell Epitopes

The NetMHCIIpan version 4.0 (34, 55) predicted a total of 13501 CD4+ epitope peptides corresponding to DENV 1-4 serotypes which were predicted to have a good binding with MHC II receptor alleles DRB1*0101, DRB1*0401, DRB1*0701, DRB5*0101, DRB1*1501, DRB1*0901, and DRB1*1302. The seven clusters of HLA-DR alleles of MHC class II were selected due to these alleles being found in 95% of the human population. A default threshold value of 2% and 10% rank was set for the prediction of strong and weak MHC II binding peptides respectively (34, 55). Of 13501 peptides, 137 epitopes were selected using binding affinity score. Further, 51 epitopes were screened based on their antigenicity potential. Out of 51 antigenic epitopes, 31 were screened to be non-toxic, and this number was further reduced down to 10 that were IFN-gamma inducing epitopes. Finally, based on the conservancy analysis, selection of three epitope peptides was made where all the three epitope peptides were found to be conserved in all the four DENV serotypes, see below and also supplementary information.

Interferon Inducing CD4+ T Cell Epitopes

From the 13501 CD4+ epitopes interferon inducing CD4+ epitope peptides were predicted using the IFN epitope server (35). The selected 10 IFN-gamma inducing epitopes are EGKIVGLYGNGVVTT, REGKIVGLYGNGVVT, GKIVGLY GNGVVTTS,TFTMRLLSPVRVPNY,SADLSLEKAAEVSWE, ATFTMRLLSPVRVPN, SSADLSLEKAAEVSW, KREK KLGEFGKAKG, KATYETDVDLGSGTR and VLRGFKK EISNMLN were considered for further analysis.

Conservancy, Antigenicity, and Toxicity of CD4+ T Cell Epitope Peptides

The usage of conserved epitopes would be an ideal choice in order to design vaccines that would eventually provide broader protection across DENV 1-4 strains. The conservancy analysis was done through IEDB conservancy analysis resource (36). The degree of linear conservancy of 10 interferon-inducing epitope peptides within the given four DENV 1-4 protein sequences was set at 100% identity level. The analysis revealed that the three epitopes were showing 100% of protein sequence matches for DENV 1-3 serotypes. The potential epitopes 'ATFTMRLLS PVRVPN' and 'TFTMRLLSPVRVPNY' shows 80% sequence identity with DENV-4 sequence. However, a higher sequence identity of 92.86% was noted for 'KREKKLGEFGKAKG' epitope with DENV 4 protein sequence. The highly conserved CD4+ epitope peptides KREKKLGEFGKAKG, TFTMRLLSPVRVPNY and ATFTMRLLSPVRVPN were selected and summarized in the **Table 1**. All the selected epitopes were predicted to be non-toxic and antigenic. The physical and chemical properties of the epitope analyzed and the results are summarized in **Table S1**, see **Supplementary File**.

Dengue Vaccine Formulation and Analysing its Physico-Chemical Properties

The formulated polyvalent epitope peptide vaccine consists of 309 amino acid residues. This includes three immunogenic T cell epitopes; three adjuvants (RS09 having sequence APPHALS; PADRE having sequence, AKFVAAWTLKAAA; and an N-terminal and C-terminal sequence of *Salmonella typhimurium* flagellin protein that were linked to the epitopes with the help of GGS linker as an immune-adjuvant to elicit a robust immune response (56, 57). In fact it was shown that the flagellin's C terminal D0 domain is required for TLR5 receptor activation (56, 57). The sequence of formulated vaccine is shown is **Figure 1**.

The formulated vaccine was antigenic with an antigenic score of 0.463 as calculated by VaxiJen 2.0 server. The final vaccine has

TABLE 1 | Selected epitope peptide conservancy analysis.

Epitope Number	Epitope sequence	Percent of protein sequence matches DENV1	Percent of protein sequence matches DENV2	Percent of protein sequence matches DENV3	Percent of protein sequence matches DENV4	Antigenicity	Toxicity
1.	ATFTMRLLSPVRVPN	100	100	100	80	ANTIGEN	Non-Toxin
2.	KREKKLGEFGKAKG	100	100	100	92.86	ANTIGEN	Non-Toxin
3.	TFTMRLLSPVRVPNY	100	100	100	80	ANTIGEN	Non-Toxin

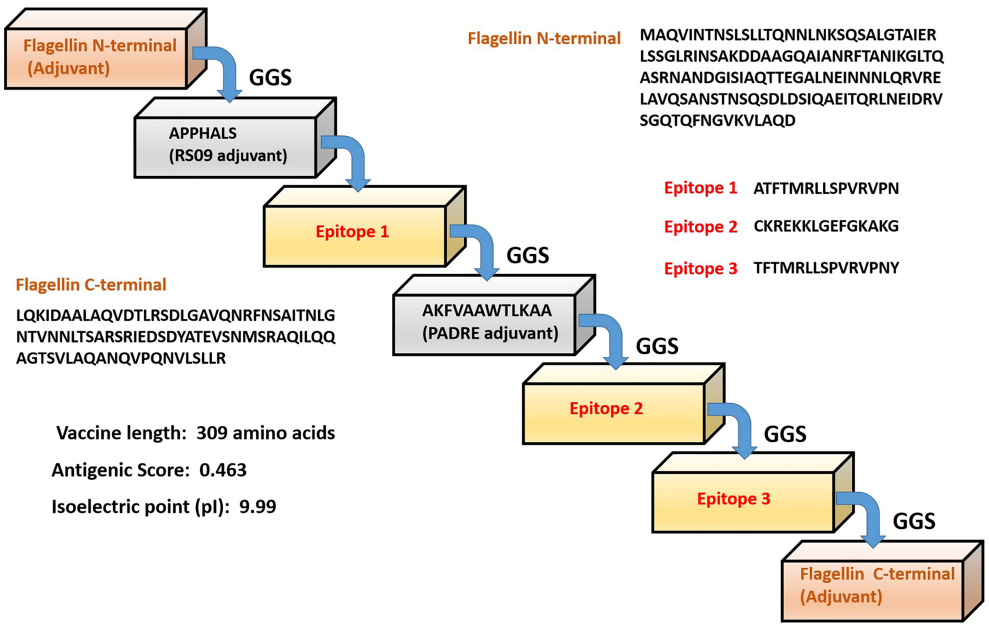


FIGURE 1 | Schematic representation of the predicted multi-epitope vaccine construct with different adjuvants, linkers and epitopes with their respective amino acid sequence.

molecular weight of 32.6 kDa, and isoelectric point (pI) of 9.99. The total number of negative and positive charged amino acids were 22 and 32 respectively. The calculated aliphatic index (88.19) indicates a stable protein in a broad range of temperatures. Our vaccine presented an estimated half-life of 30 hours in mammalian reticulocytes (*in vitro*), and with a >20 hours (in yeast, *in vivo*) and >10 hours (Escherichia coli, *in vivo*) analyzed by Expasy ProtParam server. Finally the predicted solubility of the vaccine was calculated to be 0.6909 upon overexpression in *E.coli*.

Modeling and Docking of TLR-5 and Multi-Epitope Vaccine Construct

The tertiary structure of the final multi-epitope vaccine construct and the immunogenic TLR-5 receptor was predicted using a deep learning approach of Alphafoldv2.0 program (42, 43). In order to confirm the quality of predicted structures, we calculated the Ramachandran plot of the modeled TLR-5 and Vaccine construct. For the multi-epitope vaccine construct, out of 309 residues, 85.1% fall in the core and favoured/acceptable regions, and 14.1% residues fall in additional allowed regions, and the remaining 0.8% residues fall under disallowed regions, see **Figure 2A**. However, for TLR-5 receptor, 81.2% of amino acids fall in the core acceptable region and the other 18.6% fall under the allowed region and with an additional 0.2% falling under the generously allowed region of the Ramachandran plot, see **Figure 2B**. Further, protein structure analysis (ProSA) webtool (58) was used in order to quantify the overall and local model quality of TLR-5 and multi-epitope vaccine construct, see **Figures S1**, **S2**.

The invocation of an appropriate immune response relies on the interaction between an antigenic molecule (in this case, designed multi-epitope vaccine) and a specific immune receptor (in this study, TLR-5 receptor). Molecular docking of the multi-epitope vaccine peptide with the modeled TLR-5 receptor was performed using the HADDOCK 2.2 web server (44) in order to evaluate the interaction between vaccine polypeptide and TLR-5 receptor and consequently the development of an immune response. The data-driven docking of designed multiepitope-TLR-5 complex were performed. The bacterial flagellin adjuvant is known to specifically bind to TLR-5 that stimulate and illicit response by innate immune system (56, 59). We have specifically used the information-driven docking where the knowledge about the specific residues involved in the interaction is needed to drive the docking calculations. In this regard, a recent study by Jacchieri and coworkers (59) employed a complementary hydropathy between the sequences of flagellin protein and TLR-5 receptor protein to predict the binding sites and structure of the complex. Further, their studies have clearly shown the potential binding sites in flagellin possessing "LQRVRELAVQ" sequence and for TLR5, a potential binding sites with "EILDISRNQL" sequence has been predicted. Further, their studies have shown that the side chains of Gln89, Arg92, and Glu93 on the surface of flagellin are optimally positioned to interact with the side chains of Glu552, Asp555, and Arg558 respectively on the surface of the predicted TLR-5 binding site.

The multi-epitope vaccine construct was docked with TLR-5 receptor using HADDOCK program which resulted in clustering 83 structures in 11 cluster(s), which represents 41% of the water

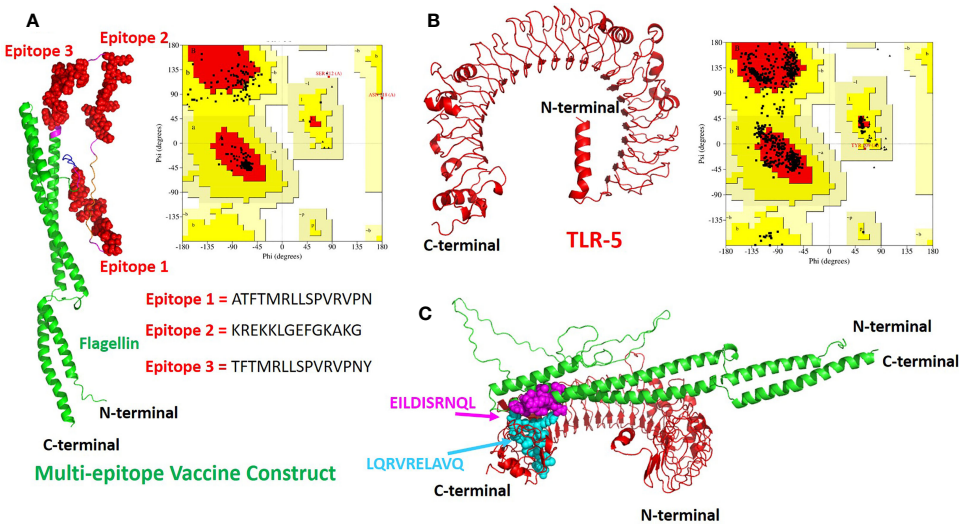


FIGURE 2 | The modeled 3D structures of the **(A)** multi-epitope vaccine construct. The three epitopes inserted are shown in red spheres and the bacterial flagellin sequence is present at the N- and C-terminal of the vaccine construct and is colored in green cartoons.; **(B)** Toll like Receptor 5(TLR-5) immunological receptors used in this study. Respective Ramachandran plots also also shown with predicted structures **(C)** The docked complex of multi-epitope vaccine and TLR-5 is shown where the hotspot residues used (see Methods section for details) for data driven docking have been shown in spheres and colored in magenta and cyan for multi-epitope vaccine construct and TLR-5 receptor respectively.

refined HADDOCK generated models. We have specifically chosen the top ranked cluster with the lowest HADDOCK score which is most significant one for docking analysis. The HADDOCK score of − 73.4± 4.1 indicates a good interaction between the multi-epitope vaccine construct and TLR-5 receptor. The score of buried surface area (BSA) is 1784.5 ± 97.8 $Å^2$ that reflected close proximity and a less water-exposed region on the protein binding surface. The **Figure 2C** show molecular docking between multi-epitope vaccine constuct and Human TLR-5 receptor. In order to study the stability of the docked complex we subjected the docked complex for Molecular Dynamic (MD) simulations, see below for detailed analysis.

MD Simulations Predicts a Stable TLR-5 and Multi-Epitope Vaccine Interaction

Molecular dynamics (MD) simulations have been proven to be very useful and effective method in order to study the stability and analysis of biological systems (60). In fact, our research group has successfully utilized and employed MD simulations to check the stability of different protein complexes (61–63) and nucleic acid systems (64–67). In this study, we used the GROMACS (45) software for molecular dynamics simulation to understand the structural properties and interaction between TLR-5 and the predicted multi-epitope vaccine. The simulation of TLR-5 immune receptor docked with multi-epitope vaccine was performed in order to analyze the interaction pattern of multi-epitope vaccine and changes that are induced in the complex after the initial docked conformation. In order to confirm the stability of TLR-5 immune receptor with the predicted multi-epitope vaccine, three MD simulations, each of 100 ns and starting with different velocities were conducted. Thus, in total, 300 ns of total simulation time was performed for TLR-5 complexes with multi-epitope vaccine construct. A 100 ns trajectory was used to analyze some essential parameters; the first one being the RMSD that reflect the stability between the two complexed moieties. The root mean square deviation (RMSD) values predict the fluctuation of Cα atoms of TLR-5 and Vaccine. In particular the RMSD value of the TLR-5 and vaccine construct complex after 10 ns and after 100ns simulations is 0.86 nm. The high RMSD can be correlated with the presence of loops region in the vaccine construct, see **Figure 3A**. The TLR-5 receptor, see **Figure 3B** shows the RMSD ranging from 0.33 nm to 0.54 nm, after 20 ns of time interval, which is considered as a mild fluctuation. In particular, the RMSD for the TLR-5 is 0.43 ± 0.03 nm. **Figure 3B** clearly indicates that TLR5 is stable throughout the simulation. However, an increased RMSD of 1.19 ± 0.08 nm was observed for multi-epitope vaccine construct. The high RMSD of multi-epitope vaccine could be correlated with the presence of loop regions where the epitope and linker sequences were inserted, that consequently resulted in the prominent fluctuations in vaccine. However, it should be noted that the RMSD of multi-epitope vaccine becomes stabilized after 20 ns of simulation time. The second analyzed parameter was Root Mean square fluctuation (RMSF) that represents amino acids side chains fluctuations. The RMSF parameter was plotted for TLR-5 and vaccine complex separately. Elevated fluctuations at the plot indicates highly flexible regions while mild ones correspond to regions that are more rigid. From **Figure 3C**, it seems evident that the calculated RMSF values for for Cα atoms of TLR-5 remained stable for the overall structure. In contrast, a high fluctuation was observed for residues ranging from 183 to

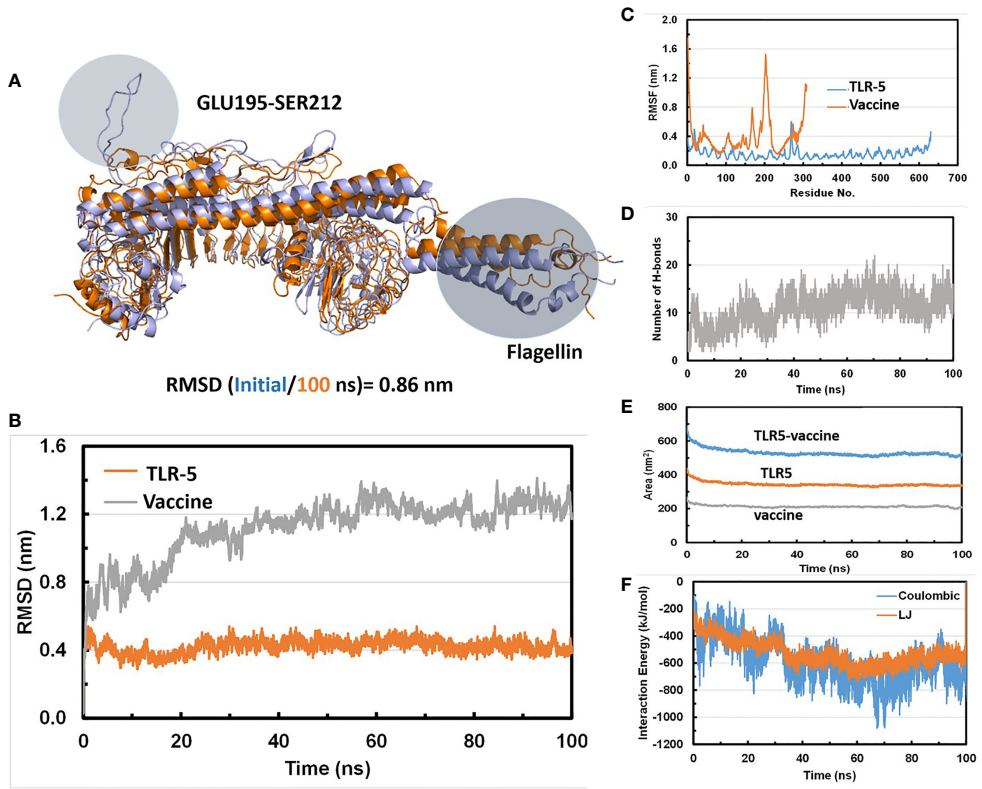


FIGURE 3 | **(A)** Superimposition of a snapshot obtained after 10 ns and 100ns of simulation time highlighting a high RMSD value due to flexible loop regions and terminal region of flagellin in a vaccine construct; **(B)** Time evolution of backbone RMSD of TLR-5 and vaccine construct during MD simulations; **(C)** Backbone RMSF plots; **(D)** Time evolution of number of hydrogen bonds Between TLR-5 and vaccione construct; **(E)** Solvent accessible surface area of TLR-5 and vaccine construct; **(F)** Variation in Interaction energy plot with segregated electrostatics and vanderwaals component obtained for entire simulation time.

220 with an average RMSF of 0.74 ± 0.4 nm, which corresponds to the loop regions in multi-epitope vaccine structure. Next, we focused on the analysis of the number of hydrogen bonds between TLR5- and vaccine construct complex, which remained constant after 20 ns of the simulations, see **Figure 3D**. From this analysis, it can further be concluded that the after initial 20 ns was required in order to maintain a stable contact between TLR5 and vaccine construct. **Figure 3E**, evidenced that the large solvent accessible surface area was observed for the vaccine, which is followed by TLR-5. Intrigued by these observations, we calculated the Interaction energies between TLR5 and vaccine complex and further decomposed the interaction energy component into coulombic interactions [E(Coul)], which represents the interaction accounting for the electrostatics between TLR-5 and vaccine construct. The other energy component is Lennard Jones [E(LJ)] potential, that accounts for the Van der waals interactions between TLR-5 and vaccine complex. From **Figure 3F**, it seems evident that the predominant interaction that stabilizes the overall TLR-5 and vaccine complex stems from electrostatic component of the interaction, which seems to play a major role in binding. This energetic analysis was in line with the number of the hydrophilic-hydrophobic residues contacts that exists between the interface region of TLR-5 and multi-epitope vaccine construct.

Detailed Analysis of Snapshots Obtained During MD Simulations and Its Comparison With Initial Modeled TLR5-Vaccine Construct

We analysed the possible non-covalent interactions between the multi-epitope vaccine and immune receptor TLR-5 using an in-house script. The docked complex is stabilized by a network of 12 hydrogen-bonding interactions, six hydrophobic interactions and five salt-bridge interactions within 5Å and 6Å distance cut-off respectively. The salt-bridge interaction exists between Arg59 of TLR-5 and ASP270 of Vaccine construct; His511 of TLR-5 and ASP227 of vaccine construct; ARG557 of TLR-5 and ASP227 of vaccine construct; LYS580 of TLR-5 and GLU94 of vaccine construct; ASP614 of TLR-5 and ARG91 of vaccine construct. In fact, the salt-bridge interaction between Arg59A of TLR-5 and ASP270B of Vaccine construct and LYS580A of TLR-5 and GLU94B of vaccine construct remained stable throughout the simulation time, see **Figure 4**. This analysis clearly revealed that the predicted vaccine shows a good interaction with TLR-5. To further assess the stability of the overall complex during the simulation time we computed RMSD of complex at different time steps and the comparative analysis of selected snapshots indicated the stability of overall TLR-5 and

vaccine complex see **Figure 5A**. Moving forward, a detailed analysis of TLR-5 and vaccine docked construct interface was performed by the COCOMAPS tool (68, 69), see **Figure 5B**. COCOMAPS enables the analysis and visualization of the interface of interaction in protein complexes by making use of intermolecular contact maps in order to identify hot spot residues (68, 69). It seems evident by looking at **Figure 5B**, that overall contacts remained stable throughout the simulation time, and the analysis is shown for the selected snapshots, where the interaction pattern with the specific contacts remained stable, as shown by black patches in the graph. The interface area is 1261.65 $Å^2$, with a percentage of polar and non-polar residues at the interface of approximately 64% and 36% respectively. Both the TLR-5 and vaccine construct interfaces involved in the

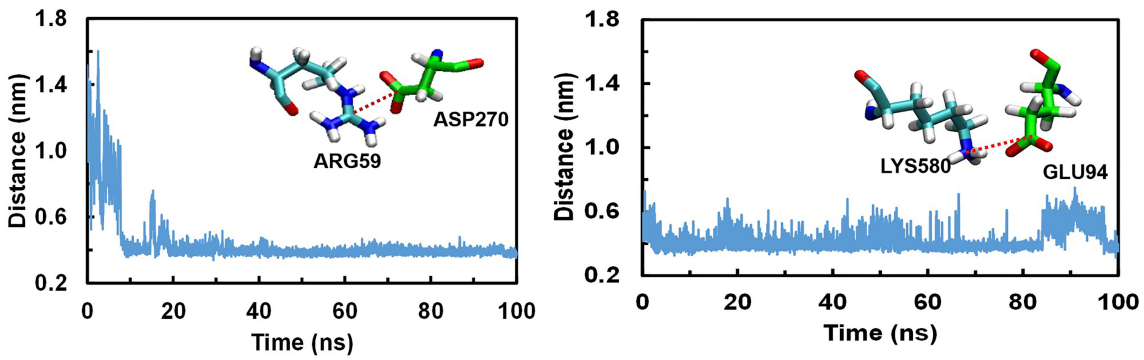


FIGURE 4 | Conserved Salt bridge interaction contacts between TLR-5 and Vaccine construct. Left) Distances over time evolution between the Arg59(CZ)-Asp270 (CB) atoms; Right) Distances over time evolution between Lys580(CE) and Glu94(CB) atoms.

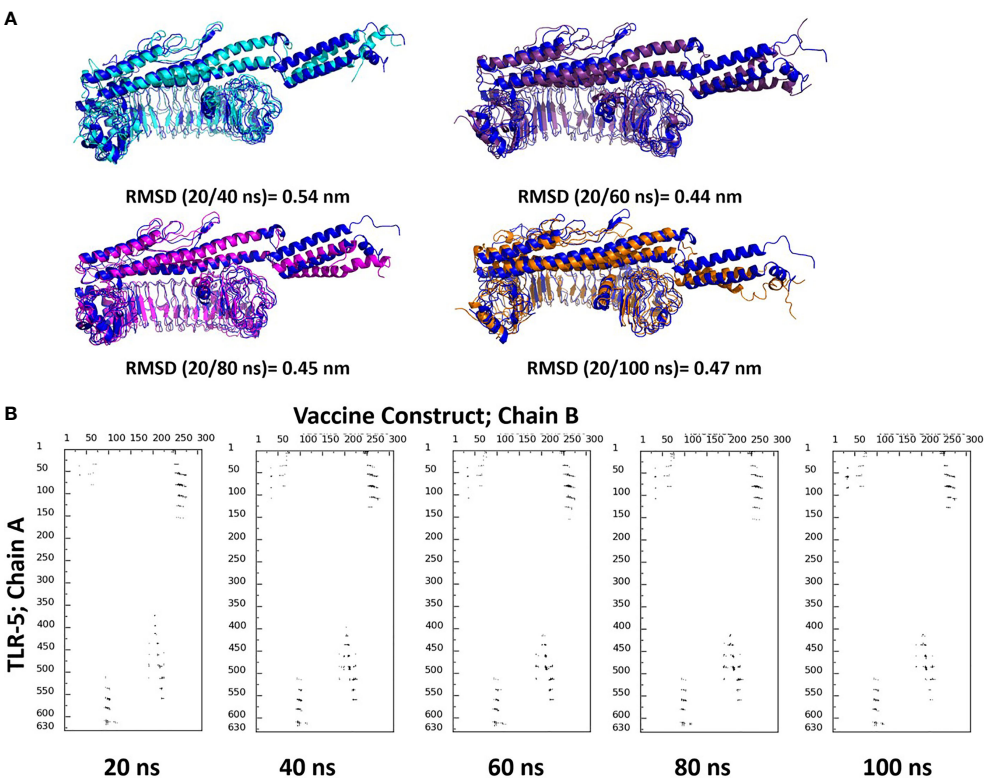


FIGURE 5 | **(A)** Superimposition of selected snapshots of TLR-5 and vaccine construct and their respective RMSD values are also reported; **(B)** Contact maps showing the conservation of contacts between residues in between TLR-5 and Vaccine construct.

interaction present a high percentage of hydrophilic residues. Using a cut-off distance of 8Å to define two atoms in contact, the dominant number of contacts exists between hydrophilic-hydrophobic residues that amounts to 230, which is followed by hydrophilic-hydrophilic and hydrophobic-hydrophobic contacts amounting to 222 and 52 respectively.

Finally, the similar analysis was performed for the docked complex by extracting a snapshot at 100ns of simulation. The TLR-5 and vaccine complex at the end of simulation was stabilized by a network of 32 hydrogen-bonding interactions, 12 hydrophobic interactions and 2 salt-bridge interactions within 5Å and 6Å distance cut-off respectively. The salt-bridge interaction exists between Arg59 of TLR-5 and ASP270 of multi-epitope vaccine construct, with an average distance between Arg59(CZ)-Asp270(CB) of 0.44 ± 0.2 nm. Next, a second salt bridge interaction between LYS580 of TLR-5 and GLU94 of vaccine construct with an average distance of Lys580 (CE) and Glu94(CB) to be 0.42 ± 0.1 nm. In fact, both these salt-bridge interactions between TLR-5 and vaccine construct remained stable throughout the simulation time, see **Figure 4**. Further, a detailed analysis of TLR-5 and vaccine docked construct interface was performed by the COCOMAPS tool (68, 69). The interface area was increased from an initial 1261.65 $Å^2$ (the initial modeled complex) to 1763.6 $Å^2$. Interestingly, the percentage of polar and non-polar residues at the interface of approximately 62% and 38% respectively, which are nearly the same as of the initial model (64% and 36% respectively for polar and non-polar residues). Similar to the initial model, where employing a cut-off distance of 8Å to define two residues in contact, the dominant number of contacts still exists between hydrophilic-hydrophobic residues that amounts to 309, which was 230 in the initial modeled structure, which is followed by hydrophilic-hydrophilic and hydrophobic-hydrophobic contacts amounting to 222 and 52 respectively, and which were 260 and 85 in the initial modeled structure. This analysis clearly indicates that the overall interaction pattern of TLR-5 and multi-epitope vaccine construct remained stable throughout the simulation time.

Peptide Synthesis and Conjugation

GenScript's microwave-based PepPower™ technology was used to synthesize the selected peptode possessing the maximum sequence identity with all the considered DENV 1-4 serotyoes, see **Table 2**. The Cysteine was added to the C-terminus of immunogenic KREKKLGEFGKAKG epitope peptide. The peptide antigen with added Cysteine at the C-terminal is now CKREKKLGEFGKAKG was subsequently conjugated with KLH (Keyhole Limpet Hemocyanin) protein to immunize two rabbits at Genscript corporation (Piscataway, NJ, USA).

In-Vivo Validation for Antibody Development and Assessment

The affinity-purified concentration of the antibody is found to be 7.16mg in 4ml of the antiserum of subcutaneously immunized New Zealand white rabbits with KLH conjugated peptide. In order to confirm whether the KLH conjugate peptide can stimulate humoral immunity and produce antibodies, the Indirect ELISA was performed to determine the antibody titer. The antibody titre is expressed as the inverse of the highest dilution of the antibody which gives positive result in ELISA. The serum collected from rabbit is serially diluted more than 8 folds, which give positive results on ELISA. On the other hand, no color is observed when IgG in the range of 1:1,000 to 1:512,000 are incubated with the antigen, see **Figure 6**. These results clearly indicate that the imunogen designed in this study could stimulate the production of antibodies *in vivo*. Interestingly, the 10 fold diluted aliquot (i.e., 1: 520,000) is also observed to give positive result on ELISA, as indicated by the titer value of 14.67 with reference signal/blank >=2.1; (see **Supplementary Information File** for more details) which further suggest that the vaccine designed possesses good immunogenicity.

CONCLUSION

In this study we have employed robust immunoinformatics approach coupled with *in vivo* studies in order to design a potential multi-epitope vaccine candidate for DENV vaccine. As a first step, we predicted 13501 MHC II binding CD4+ epitope peptides as a first step. Further we predicted and refined 10 conserved epitopes among all the four serotypes of the DENV and all the predicted epitopes are predicted to be interferon-inducing peptides. The antigenic and nontoxic epitopes 'ATFTMRLLSPVRVPN', 'KREKKLGEFGKAKG' and TFTMRLLSPVRVPNY that were conserved among all the four DENV serotypes were selected for final multi-epitope vaccine designing. The final vaccine has 309 amino residues, molecular weight 32.66 kDa, theoretical pI 9.99, and an estimated half-life of 30 hours in mammalian reticulocytes, and with a predicted antigenic score of 0.463. The molecular docking and molecular dynamic simulations predicts a stable complex between the multi-epitope vaccine construct and TLR-5 receptor. Overall, our computational analysis suggests that the constructed multi-epitope

TABLE 2 | Selected conserved epitope peptide for synthesis.

Epitope selected	Serotype	Protein name	Positions	IEDB sequence Identity
KREKKLG	1	sp|P33478|POLG_DEN1S Genome polyprotein OS=Dengue virus type 1 (strain Singapore/S275/1990)	2947-2960	100
EFGKAKG	2	sp|P07564|POLG_DEN2J Genome polyprotein OS=Dengue virus type 2 (strain Jamaica/1409/1983)	2947-2960	100
	3	sp|Q6YMS4|POLG_DEN3S Genome polyprotein OS=Dengue virus type 3 (strain Sri Lanka/1266/2000)	2946-2959	100
	4	sp|Q58HT7|POLG_DEN4P Genome polyprotein OS=Dengue virus type 4 (strain Philippines/H241/1956)	2944-2957	92.86

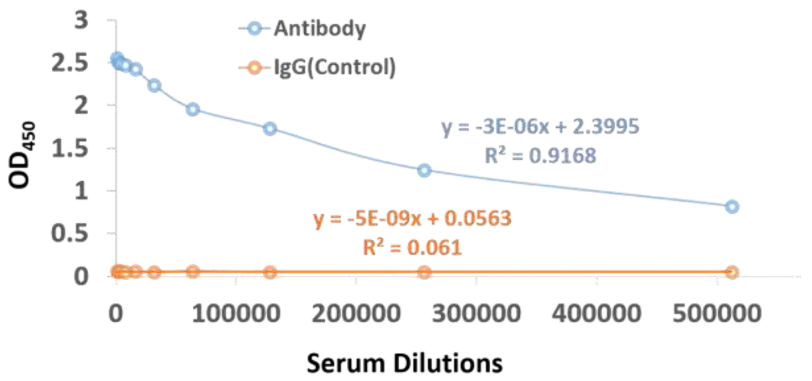


FIGURE 6 | ELISA assay scatter plot of OD_{450} values and serum dilutions from (1:1000 to 1:512,000) of dengue virus antibody and IgG control with varying concentration of antigen. The OD_{450} values can be inversely correlated to the serum dilutions.

vaccine was highly immunogenic, safe, and non-toxic in nature. As a next step, the most conserved epitope KREKKLGEFGKAKG among the four DENV serotypes was synthesized. Finally, in order to confirm whether the KLH conjugate peptide can stimulate humoral immunity and produce antibodies, the Indirect ELISA was performed to determine the antibody titer. The affinity-purified concentration of the antibody is found to be 7.16mg in 4ml of the antiserum of subcutaneously immunized New Zealand white rabbits with KLH conjugated peptide. Interestingly, the 10 fold diluted aliquot (i.e., 1: 520,000) is also observed to give positive result on ELISA, as indicated by the titer value of 14.67. These results clearly indicate that the imunogen designed in this study could stimulate the production of antibodies *in vivo*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was reviewed and approved by Genscript corporation (Piscataway, NJ, USA).

AUTHOR CONTRIBUTIONS

MC and VK designed the whole project and analysis. LRG, SKG, ARS, UK, LC, MC carried out the analysis. SG, AS, MC wrote the manuscript. MC and ARS prepared figures. LC and MC supervised the overall study. All authors contributed to the article and approved the submitted version.

FUNDING

This research received no external funding. The APC charges was funded by KAUST baseline research funding (to LC). The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia. Authors would also like to acknowledge team members from STEMskills Research and Education Lab Private Limited for critical reading of manuscript and computational support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.865180/full#supplementary-material

REFERENCES

1. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The Global Burden of Dengue: An Analysis From the Global Burden of Disease Study 2013. *Lancet Infect Dis* (2016) 16:712–23. doi: 10.1016/S1473-3099(16)00026-8
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The Global Distribution and Burden of Dengue. *Nature* (2013) 496:504–7. doi: 10.1038/nature12060
3. Perera R, Kuhn RJ. Structural Proteomics of Dengue Virus. *Curr Opin Microbiol* (2008) 11:369–77. doi: 10.1016/j.mib.2008.06.004
4. Osatomi K, Sumiyoshi H. Complete Nucleotide-Sequence of Dengue Type-3 Virus Genome Rna. *Virology* (1990) 176:643–7. doi: 10.1016/0042-6822(90)90037-R
5. Holmes EC. Molecular Epidemiology and Evolution of Emerging Infectious Diseases. *Brit Med Bull* (1998) 54:533–43. doi: 10.1093/oxfordjournals.bmb.a011708
6. Mustafa MS, Rasotgi V, Jain S, Gupta V. Discovery of Fifth Serotype of Dengue Virus (DENV-5): A New Public Health Dilemma in Dengue Control. *Med J Armed Forces India* (2015) 71:67–70. doi: 10.1016/j.mjafi.2014.09.011
7. Back AT, Lundkvist A. Dengue Viruses - an Overview. *Infect Ecol Epidemiol* (2013) 3: 1–21. doi: 10.3402/iee.v3i0.19839

8. Low JGH, Ooi EE, Vasudevan SG. Current Status of Dengue Therapeutics Research and Development. *J Infect Dis* (2017) 215:S96–S102. doi: 10.1093/infdis/jiw423
9. Thomas SJ, Yoon IK. A Review of Dengvaxia?: Development to Deployment. *Hum Vacc Immunother* (2019) 15:2295–314. doi: 10.1080/21645515.2019.1658503
10. Gubler DJ, Halstead SB. HEAD TO HEAD Is Dengvaxia a Useful Vaccine for Dengue Endemic Areas? *Bmj Brit Med J* (2019) 367. doi: 10.1136/bmj.l5710
11. Pinheiro-Michelsen JR, Souza RDO, Santana IVR, da Silva PD, Mendez EC, Luiz WB, et al. Anti-Dengue Vaccines: From Development to Clinical Trials. *Front Immunol* (2020) , 11. doi: 10.3389/fimmu.2020.01252
12. Biswal S, Reynales H, Saez-Llorens X, Lopez P, Borja-Tabora C, Kosalaraksa P, et al. Efficacy of a Tetravalent Dengue Vaccine in Healthy Children and Adolescents. *New Engl J Med* (2019) 381:2009–19. doi: 10.1056/NEJMoa1903869
13. Jackson LA, Rupp R, Papadimitriou A, Wallace D, Raanan M, Moss KJ. A Phase 1 Study of Safety and Immunogenicity Following Intradermal Administration of a Tetravalent Dengue Vaccine Candidate. *Vaccine* (2018) 36:3976–83. doi: 10.1016/j.vaccine.2018.05.028
14. Porter KR, Ewing D, Chen L, Wu SJ, Hayes CG, Ferrari M, et al. Immunogenicity and Protective Efficacy of a Vaxfectin-Adjuvanted Tetravalent Dengue DNA Vaccine. *Vaccine* (2012) 30:336–41. doi: 10.1016/j.vaccine.2011.10.085
15. Guzman MG, Alvarez M, Halstead SB. Secondary Infection as a Risk Factor for Dengue Hemorrhagic Fever/Dengue Shock Syndrome: An Historical Perspective and Role of Antibody-Dependent Enhancement of Infection. *Arch Virol* (2013) 158:1445–59. doi: 10.1007/s00705-013-1645-3
16. White LJ, Young EF, Stoops MJ, Henein SR, Adams EC, Baric RS, et al. Defining Levels of Dengue Virus Serotype-Specific Neutralizing Antibodies Induced by a Live Attenuated Tetravalent Dengue Vaccine (TAK-003). *PloS Negl Trop D* (2021) 15:1–15. doi: 10.1371/journal.pntd.0009258
17. Teoh EP, Kukkaro P, Teo EW, Lim APC, Tan TT, Yip A, et al. The Structural Basis for Serotype-Specific Neutralization of Dengue Virus by a Human Antibody. *Sci Transl Med* (2012) 4:1–9. doi: 10.1126/scitranslmed.3003888
18. Zhu SL, Feng Y, Rao PH, Xue XY, Chen S, Li WS, et al. Hepatitis B Virus Surface Antigen as Delivery Vector can Enhance Chlamydia Trachomatis MOMP Multi-Epitope Immune Response in Mice. *Appl Microbiol Biot* (2014) 98:4107–17. doi: 10.1007/s00253-014-5517-x
19. Jiang P, Cai Y, Chen J, Ye X, Mao S, Zhu S, et al. Evaluation of Tandem Chlamydia Trachomatis MOMP Multi-Epitopes Vaccine in BALB/c Mice Model. *Vaccine* (2017) 35:3096–103. doi: 10.1016/j.vaccine.2017.04.031
20. Lin XY, Chen S, Xue XY, Lu LJ, Zhu SL, Li WS, et al. Chimerically Fused Antigen Rich of Overlapped Epitopes From Latent Membrane Protein 2 (LMP2) of Epstein-Barr Virus as a Potential Vaccine and Diagnostic Agent. *Cell Mol Immunol* (2016) 13:492–501. doi: 10.1038/cmi.2015.29
21. Fadaka AO, Sibuyi NRS, Martin DR, Goboza M, Klein A, Madiehe AM, et al. Immunoinformatics Design of a Novel Epitope-Based Vaccine Candidate Against Dengue Virus. *Sci Rep* (2021) 11:19707. doi: 10.1038/s41598-021-99227-7
22. Dixit NK. Design of Monovalent and Chimeric Tetravalent Dengue Vaccine Using an Immunoinformatics Approach. *Int J Pept Res Ther* (2021) 27:2607–24. doi: 10.1007/s10989-021-10277-x
23. Bhardwaj A, Sharma R, Grover A. Immuno-Informatics Guided Designing of a Multi-Epitope Vaccine Against Dengue and Zika. *J Biomol Struct Dyn* (2021), 1–15. doi: 10.1080/07391102.2021.2002720
24. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring Dengue Genome to Construct a Multi-Epitope Based Subunit Vaccine by Utilizing Immunoinformatics Approach to Battle Against Dengue Infection. *Sci Rep* (2017) 7:9232. doi: 10.1038/s41598-017-09199-w
25. Khan AM, Miotto O, Nascimento EJM, Srinivasan KN, Heiny AT, Zhang GL, et al. Conservation and Variability of Dengue Virus Proteins: Implications for Vaccine Design. *PloS Negl Trop D* (2008) 2:1–15. doi: 10.1371/journal.pntd.0000272
26. Mangada MM, Rothman AL. Altered Cytokine Responses of Dengue-Specific CD4(+) T Cells to Heterologous Serotypes. *J Immunol* (2005) 175:2676–83. doi: 10.4049/jimmunol.175.4.2676
27. Chong LC, Khan AM. Identification of Highly Conserved, Serotype-Specific Dengue Virus Sequences: Implications for Vaccine Design (Vol 20, 921, 2019). *BMC Genomics* (2021) 22:1–18. doi: 10.1186/s12864-019-6311-z
28. Krishnan GS, Joshi A, Kaushik V. T Cell Epitope Designing for Dengue Peptide Vaccine Using Docking and Molecular Simulation Studies. *Mol Simulat* (2020) 46:787–95. doi: 10.1080/08927022.2020.1772970
29. Krishnan GS, Joshi A, Akhtar N, Kaushik V. Immunoinformatics Designed T Cell Multi Epitope Dengue Peptide Vaccine Derived From non Structural Proteome. *Microb Pathogenesis* (2021) 150:104728. doi: 10.1016/j.micpath.2020.104728
30. Duangchinda T, Dejnirattisai W, Vasanawathana S, Limpitikul W, Tangthawornchaikul N, Malasit P, et al. Immunodominant T-Cell Responses to Dengue Virus NS3 Are Associated With DHF. *Proc Natl Acad Sci U S A* (2010) 107:16922–7. doi: 10.1073/pnas.1010867107
31. Weiskopf D, Bangs DJ, Sidney J, Kolla RV, De Silva AD, de Silva AM, et al. Dengue Virus Infection Elicits Highly Polarized CX3CR1+ Cytotoxic CD4+ T Cells Associated With Protective Immunity. *Proc Natl Acad Sci USA* (2015) 112:E4256–4263. doi: 10.1073/pnas.1505956112
32. Doytchinova IA, Flower DR. VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinf* (2007) 8:1–7. doi: 10.1186/1471-2105-8-4
33. Doytchinova IA, Flower DR. Identifying Candidate Subunit Vaccines Using an Alignment-Independent Method Based on Principal Amino Acid Properties. *Vaccine* (2007) 25:856–66. doi: 10.1016/j.vaccine.2006.09.032
34. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M. Improved Prediction of MHC II Antigen Presentation Through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J Proteome Res* (2020) 19:2304–15. doi: 10.1021/acs.jproteome.9b00874
35. Dhanda SK, Vir P, Raghava GP. Designing of Interferon-Gamma Inducing MHC Class-II Binders. *Biol Direct* (2013) 8:30. doi: 10.1186/1745-6150-8-30
36. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an Epitope Conservancy Analysis Tool to Facilitate the Design of Epitope-Based Diagnostics and Vaccines. *BMC Bioinf* (2007) 8:1–6. doi: 10.1186/1471-2105-8-361
37. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery C, et al. Open Source Drug Discovery, C. And In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS One* (2013) 8:e73957. doi: 10.1371/journal.pone.0073957
38. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, et al. Humana Press Inc.:Totowa, NJ(2005).
39. Negahdaripour M, Eslami M, Nezafat N, Hajighahramani N, Ghoshoon MB, Shoolian E, et al. A Novel HPV Prophylactic Peptide Vaccine, Designed by Immunoinformatics and Structural Vaccinology Approaches. *Infect Genet Evol* (2017) 54:402–16. doi: 10.1016/j.meegid.2017.08.002
40. Hasan M, Islam S, Chakraborty S, Mustafa AH, Azim KF, Joy ZF, et al. Contriving a Chimeric Polyvalent Vaccine to Prevent Infections Caused by Herpes Simplex Virus (Type-1 and Type-2): An Exploratory Immunoinformatic Approach. *J Biomol Struct Dyn* (2020) 38:2898–915. doi: 10.1080/07391102.2019.1647286
41. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: A Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Res* (2005) 33:W72–76. doi: 10.1093/nar/gki396
42. David A, Islam S, Tankhilevich E, Sternberg MJE. The AlphaFold Database of Protein Structures: A Biologist's Guide. *J Mol Biol* (2022), 434. doi: 10.1016/j.jmb.2021.167336
43. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly Accurate Protein Structure Prediction With AlphaFold. *Nature* (2021) 596:583–+. doi: 10.1038/s41586-021-03819-2
44. van Zundert GCP, Rodrigues J, 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. doi: 10.1016/j.jmb.2015.09.014
45. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* (2015) 1-2:19–25. doi: 10.1016/j.softx.2015.06.001
46. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins* (2010) 78:1950–8. doi: 10.1002/prot.22711
47. Joung IS, Cheatham TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J Phys Chem B* (2008) 112:9020–41. doi: 10.1021/jp8001614

48. Parrinello M, Rahman A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys* (1981) 52:7182–90. doi: 10.1063/1.328693
49. Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* (1993) 98:10089–92. doi: 10.1063/1.464397
50. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* (1997) 18:1463–72. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
51. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graph* (1996) 14:33–38, 27–38. doi: 10.1016/0263-7855(96)00018-5
52. Tree MO, McKellar DR, Kieft KJ, Watson AM, Ryman KD, Conway MJ. Insect-Specific Flavivirus Infection is Restricted by Innate Immunity in the Vertebrate Host. *Virology* (2016) 497:81–91. doi: 10.1016/j.virol.2016.07.005
53. Cuspoca AF, Diaz LL, Acosta AF, Penaloza MK, Mendez YR, Clavijo DC, et al. An Immunoinformatics Approach for SARS-CoV-2 in Latam Populations and Multi-Epitope Vaccine Candidate Directed Towards the World's Population. *Vaccines-Basel* (2021) 9:1–36. doi: 10.3390/vaccines9060581
54. Rappuoli R. Reverse Vaccinology, a Genome-Based Approach to Vaccine Development. *Vaccine* (2001) 19:2688–91. doi: 10.1016/S0264-410X(00)00554-5
55. 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:W449–54. doi: 10.1093/nar/gkaa379
56. Duthie MS, Windish HP, Fox CB, Reed SG. Use of Defined TLR Ligands as Adjuvants Within Human Vaccines. *Immunol Rev* (2011) 239:178–96. doi: 10.1111/j.1600-065X.2010.00978.x
57. Forstneric V, Ivicak-Kocjan K, Plaper T, Jerala R, Bencina M. The Role of the C-Terminal D0 Domain of Flagellin in Activation of Toll Like Receptor 5. *PloS Pathog* (2017) 13:1–20. doi: 10.1371/journal.ppat.1006574
58. Wiederstein M, Sippl MJ. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res* (2007) 35:W407–410. doi: 10.1093/nar/gkm290
59. Jacchieri SG, Torquato R, Brentani RR. Structural Study of Binding of Flagellin by Toll-Like Receptor 5. *J Bacteriol* (2003) 185:4243–7. doi: 10.1128/JB.185.14.4243-4247.2003
60. Karplus M, Kuriyan J. Molecular Dynamics and Protein Function. *Proc Natl Acad Sci U S A* (2005) 102:6679–85. doi: 10.1073/pnas.0408930102
61. Oliva R, Shaikh AR, Petta A, Vangone A, Cavallo L. D936Y and Other Mutations in the Fusion Core of the SARS-CoV-2 Spike Protein Heptad Repeat 1: Frequency, Geographical Distribution, and Structural Effect. *Molecules* (2021) 26:1–13. doi: 10.3390/molecules26092622
62. Hayashi S, Ueno H, Shaikh AR, Umemura M, Kamiya M, Ito Y, et al. Molecular Mechanism of ATP Hydrolysis in F1-ATPase Revealed by Molecular Simulations and Single-Molecule Observations. *J Am Chem Soc* (2012) 134:8447–54. doi: 10.1021/ja211027m
63. Shah D, Shaikh AR. Interaction of Arginine, Lysine, and Guanidine With Surface Residues of Lysozyme: Implication to Protein Stability. *J Biomol Struct Dyn* (2016) 34:104–14. doi: 10.1080/07391102.2015.1013158
64. Chawla M, Gorle S, Shaikh AR, Oliva R, Cavallo L. Replacing Thymine With a Strongly Pairing Fifth Base: A Combined Quantum Mechanics and Molecular Dynamics Study. *Comput Struct Biotechnol J* (2021) 19:1312–24. doi: 10.1016/j.csbj.2021.02.006
65. Chawla M, Autiero I, Oliva R, Cavallo L. Energetics and Dynamics of the non-Natural Fluorescent 4AP:DAP Base Pair. *Phys Chem Chem Phys* (2018) 20:3699–709. doi: 10.1039/C7CP07400J
66. Kalra K, Gorle S, Cavallo L, Oliva R, Chawla M. Occurrence and Stability of Lone Pair-Pi and OH-Pi Interactions Between Water and Nucleobases in Functional RNAs. *Nucleic Acids Res* (2020) 48:5825–38. doi: 10.1093/nar/gkaa345
67. Chawla M, Credendino R, Poater A, Oliva R, Cavallo L. Structural Stability, Acidity, and Halide Selectivity of the Fluoride Riboswitch Recognition Site. *J Am Chem Soc* (2015) 137:299–306. doi: 10.1021/ja510549b
68. Vangone A, Oliva R, Cavallo L. CONS-COCOMAPS: A Novel Tool to Measure and Visualize the Conservation of Inter-Residue Contacts in Multiple Docking Solutions. *BMC Bioinf* (2012) 13:1–9. doi: 10.1186/1471-2105-13-S4-S19
69. Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. COCOMAPS: A Web Application to Analyze and Visualize Contacts at the Interface of Biomolecular Complexes. *Bioinformat* (2011) 27:2915–6. doi: 10.1093/bioinformatics/btr484

Conflict of Interest: Authors UK, and ARS were employed by the company STEMskills Research and Education Lab Private Limited, Faridabad, Haryana, India.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Immunology

OPINION
published: 07 July 2022
doi: 10.3389/fimmu.2022.908459

Comprehending B-Cell Epitope Prediction to Develop Vaccines and Immunodiagnostics

Salvador Eugenio C. Caoili *

Biomedical Innovations Research for Translational Health Science (BIRTHS) Laboratory, Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila, Philippines

Keywords: B-cell epitope prediction, B-cell epitopes, surface accessibility, conformational disorder, immunodominance, peptide-based vaccines, antipeptide antibodies, protective immunity

1 INTRODUCTION

B-cell epitope prediction (BCEP) is the original subject of immunoinformatics (i.e., bioinformatics applied to immunology). This began with protein sequence analysis to identify hydrophilic peptide fragments bound by antibodies that recognize whole proteins (1), to enable the earlier proposed development of synthetic peptide-based vaccines for inducing protective antibody-mediated immunity (2). BCEP was thus initially "prediction of protein antigenic determinants," with each antigenic determinant being a B-cell epitope (BCE): a structural feature (e.g., sequence segment) recognized by a paratope (i.e., antigen-binding site of an immunoglobulin such as an antibody) (3, 4). Now understood as the computational identification of putative BCEs, BCEP has since grown to comprise much more sophisticated methods for analyzing both sequence (5–9) and higher-order structure (10, 11) on ever larger scales (e.g., applying genomics and proteomics for vaccine design (12, 13)). However, the full potential of BCEP for peptide-based vaccine design remains to be realized, for which reason the utility of BCEP as such has been called into question (14–17). Nevertheless, BCEP can support the development of vaccines and immunodiagnostics provided that its limitations are adequately comprehended and addressed.

2 ACCESSIBLE DISORDER (AD) AND IMMUNODOMINANCE

Accessible disorder (AD) is the state of a BCE that is simultaneously both paratope-accessible and disordered (i.e., conformationally unconstrained), such that BCE-paratope binding can occur *via* induced fit and/or conformational selection (8). Clearly, BCEs must be accessible to paratopes for physical contact to occur between them. In the context of antibody-mediated protective (e.g., antipathogen) immunity, this is most readily feasible for BCEs that are on outwardly protruding solvent-accessible molecular surfaces at extracellular sites (e.g., among secreted biomolecules). Hence, vaccine-design initiatives tend to selectively apply BCEP with a focus on surface-exposed sites among biomolecular targets of the pertinent (e.g., pathogen) secretome (i.e., totality of secreted biomolecules) and surfome (i.e., surface proteome) (18, 19). Additionally, surface complementarity must be attained between BCEs and their corresponding paratopes upon physical contact if stable BCE-paratope binding is to occur, for which the BCEs must adopt suitable conformations. This is favored where the BCEs are disordered prior to binding by the paratopes (20, 21). Vaccines can thus

OPEN ACCESS

Edited by:
Marc H V Van Regenmortel, Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Linlin Guo, The Ohio State University, United States

***Correspondence:**
Salvador Eugenio C. Caoili
badong@post.upm.edu.ph

Specialty section:
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

Received: *19 April 2022*
Accepted: *13 June 2022*
Published: *07 July 2022*

Citation:
Caoili SEC (2022) Comprehending B-Cell Epitope Prediction to Develop Vaccines and Immunodiagnostics. Front. Immunol. 13:908459. doi: 10.3389/fimmu.2022.908459

be produced from peptides comprising disordered BCEs of selected target proteins (e.g., pathogen virulence factors), to elicit production of antipeptide antibodies (i.e., peptide antibodies (22)) that can neutralize the biological activity of the proteins (4, 23–36). Such an approach is viable where the BCEs are disordered in both the peptides and the proteins; but if the BCEs are conformationally constrained (e.g., folded) in the proteins, their binding by the antipeptide antibodies may fail to occur, as is thought to be the case among unsuccessful attempts at peptide-based vaccine development (16).

AD among BCEs thus facilitates BCE-paratope binding; but BCE-specific antibody production is also subject to the phenomenon of immunodominance (i.e., bias of immune responses toward subsets of BCEs encountered in the course of immunization), as depicted in **Figure 1**. Driven by Darwinian competition among B-cell clones, immunodominance tends to be favored by greater numbers of functional BCE-recognizing precursor B cells as well as stronger binding of BCEs by B cells in terms of both affinity (i.e., strength of binding per individual BCE-paratope pairwise interaction) and avidity (i.e., overall strength of cooperative binding among paratopes that simultaneously bind two or more BCEs on a single antigen, as is possible with engagement of one or more bivalent immunoglobulin molecules) (37). Consequently, individual host life history of antigenic exposure (e.g., *via* infection and immunization) influences immunodominance. Immunodominance may thus be precluded by immune tolerance (i.e., selective inability to mount immune responses to particular BCEs, due to functional deletion or inactivation of their corresponding B cells), which is often induced by BCEs of host self antigens (i.e., autoantigens) and of other antigens (e.g., in food) to which the host has been exposed in a natural physiologic setting (rather than in the course of infectious disease or vaccination) (38–40). Alternatively, immunodominance may be heightened by the immunological memory of prior immunization (e.g., *via* infection or vaccination), as occurs in the phenonenon of original antigenic sin (i.e., antigenic imprinting) whereby memory B-cell clones generated by past immunization continue to dominate antibody responses to more recent immunizations, possibly even compromising the ability to mount protective immune responses against newly encountered pathogen variants (41, 42). From an evolutionary standpoint, pathogens may co-evolve with their hosts to evade immune destruction in part by altering their BCE repertoires to limit the expression of immunodominant pathogen BCEs on key virulence factors (e.g., *via* molecular mimicry, with pathogen BCEs tending to resemble host self BCEs) while possibly also expressing immunodominant pathogen BCEs that serve as antigenic decoys to detract from protective host immune responses (43). Furthermore, immune tolerance may be broken under certain circumstances (e.g., infection by a pathogen employing molecular mimicry), which may result in antibody-mediated (e.g., autoimmune) disease (44). These various scenarios highlight the potential complexity of vaccine development with the diversity of BCEs and possible immune responses thereto. Peptide-based vaccine design thus provides opportunities to systematically restrict the repertoire of vaccine BCEs and thereby selectively target key biomolecules (e.g., critical virulence factors) while avoiding harmful or otherwise counterproductive antibody responses (e.g., to BCEs of autoantigens and antigenic decoys).

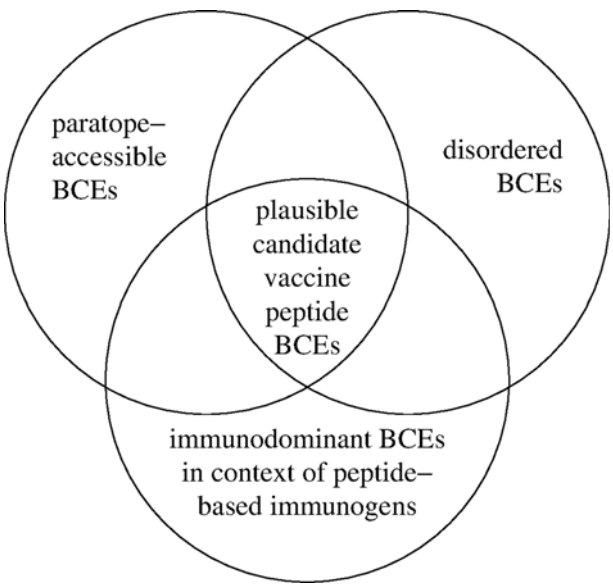


FIGURE 1 | Identification of plausible candidate vaccine peptide BCEs. Accessible disorder (AD) is realized for BCEs that are simultaneously both paratope-accessible and disordered (i.e., conformationally unconstrained) in both peptide-based immunogens and cognate native antigenic targets *in situ* (e.g., extracellular pathogen virulence factors). Immunodominant BCEs are identified empirically as they occur in peptide-based immunogens (e.g., using immunogenic carrier molecules and immunologic adjuvants) versus other contexts (e.g., in native antigenic targets comprising antigenic decoys).

3 TOWARD NEW VACCINES AND IMMUNODIAGNOSTICS

In essence, BCEP consists of two steps: structural partitioning of a prospective target (e.g., protein) into plausible candidate BCEs (e.g., peptidic sequences) and evaluation of these to assign them numerical scores that can inform subsequent decisions (e.g., on selecting components for inclusion in vaccines). Ideally, the scores would directly quantify functional impact (e.g., degree of antibody-mediated host protection against a protein toxin). In practice, functional impact can be estimated from BCE-paratope binding affinity in conjunction with a limited set of other key parameters (e.g., concentrations of antibody and its target), with said affinity itself being estimated as the BCE-paratope standard free-energy change of binding ($\Delta G_b°$) based on target structure (8). This can be simplified by considering only candidate BCEs of paratope-accessible disordered target regions. Viewed more comprehensively, BCEP can be performed largely by excluding target regions that are at least partially inaccessible to paratopes. Such regions may be inaccessible due to biomolecular folding, complex formation (e.g., oligomerization) and anatomic compartmentalization (e.g., due to biomembranes). Accordingly, target regions may be identified for exclusion using appropriate computational tools to predict folding (vis-à-vis disorder, e.g., using AlphaFold (45, 46)) and higher-order structural organization (e.g., among supramolecular assemblies such as biomembranes (47)). For convenience, inaccessibility may be generalized to also include forms of posttranslational modification resulting from covalent linkage of nonprotein moieties (e.g., glycosylation), which can be computationally predicted to mark candidate BCEs for exclusion as well. Likewise, generalization of inaccessibility can also be extended to regions featuring disulfide-bond formation between cysteine residues, as this may impede BCE-paratope binding. After exclusion of implausible or otherwise potentially problematic target regions based on anticipated inaccessibility, $\Delta G_b°$ and in turn functional impact may be estimated for the remaining candidate BCEs.

For the most part, BCEP can thus be regarded as prediction of AD. This varies with the envisioned practical application (e.g., vaccines versus immunodiagnostics). For vaccines, AD is defined by what can be achieved *in vivo* based on endogenous antibody production, with intracellular targets tending to be inaccessible under physiologic conditions, though antibodies are sometimes internalized by host cells in either free or pathogen-bound form to mediate immunity within certain intracellular compartments (48, 49). Such limitation may be overcome *via* immunotherapeutics using exogenously supplied antibodies and derivatives thereof (e.g., antibody fragments), notably with artificially produced cell-penetrating antibodies that can cross plasma membranes to bind intracellular targets (50, 51). For immunodiagnostics, the potential extent of AD is even greater, as constraints *in vivo* (e.g., on temperature and chemical composition) can be transcended *in vitro* (e.g., within a diagnostic test kit). For instance, membrane permeabilization (e.g., using detergents and/or organic solvents (52)) can extend paratope access into virtually all intracellular compartments; and treatment with chaotropic agents (e.g., urea (53) or ammonium thiocyanate (54)) can disrupt both intermolecular and intramolecular interactions, thereby increasing AD *via* order-to-disorder transitions, as in biomolecular disassembly and unfolding. BCEP itself is expanding in scope to support other applications, notably the design of novel antibodies and related constructs. BCEs can thus be identified initially as targets for binding by complementary peptides, which may be computationally designed and subsequently grafted on an antibody scaffold to produce novel antibodies that recognize the BCEs (55). The complementary peptide sequences might themselves be disordered, possibly remaining so on the antibody scaffold and even after binding their targets; but such persistent disorder could pose challenges for biotechnological antibody production, particularly with disordered regions that are recognized to initiate proteasome-mediated degradation (56). Nevertheless, said degradation could be circumvented *via* cell-free antibody production, which can also be used to produce cell-penetrating antibodies (57). Additionally, preexisting (e.g., antihapten) antibodies could be used with a diverse repertoire of peptide-based adaptors (e.g., hapten-labeled complementary peptides) for chemically programmable immunity that entails redirection of the antibodies to particular targets according to the specific choice of adaptors (58). Such repurposing of preexisting antibodies could enable more rapid responses to emerging threats (e.g., novel pathogens) than would be possible *via* development of novel antibodies or similar constructs.

Beyond predicting AD, BCEP also encompasses the more computationally challenging task of predicting conformational BCEs (i.e., BCEs that are to at least some extent conformationally constrained). Among folded proteins, conformational BCEs are surface-exposed regions that each constitute a paratope footprint and are thus typically discontinuous BCEs in the sense of comprising paratope-contacting residues that are noncontiguous along the protein sequence, sometimes even on separate polypeptide chains where proteins form supramolecular complexes such as viral capsids (3). Whereas prediction of AD can be cast as sequence profiling that is reminiscent of the earliest BCEP methods, BCEP for conformational BCEs entails the nontrivial cascading problems of delineating discontinuous candidate BCEs, predicting their hierarchy of immunodominance and estimating $\Delta G_b°$ for cross-reactive binding of their corresponding paratopes to disordered candidate BCEs (e.g., peptidic sequences forming parts of discontinuous candidate BCEs) (59). This is further complicated by possible protein unfolding *in vivo* and consequent uncertainty regarding relevant candidate BCEs, which confounds interpretation of data on binding of antiprotein antibodies to peptide fragments of the cognate protein antigens (60). In light of these considerations, the apparently poor performance of BCEP methods benchmarked against said data (15, 61) is unsurprising and likely reflects the unmitigated complexity of factors underlying BCEP to identify short peptide sequences that are recognized by antiprotein antibodies. Yet, such challenges pose barriers to development

of peptide-based constructs as immunodiagnostic reagents for detecting antiprotein antibodies rather than as vaccine components for eliciting production of antipeptide antibodies (59), though this crucial distinction was unfortunately obscured at the inception of BCEP (1). Confusion has thus resulted mainly from failure to distinguish between capacity of peptides to be bound by antiprotein antibodies (i.e., cross-reactive antigenicity) and capacity of peptides to elicit production of antipeptide antibodies that cross-react with proteins to confer protective immunity (i.e., cross-protective immunogenicity); and this has been the main reason underlying failed attempts at peptide-based vaccine development (17).

4 CONCLUSION

BCEP can be framed mainly as computational identification of putative paratope-accessible disordered peptidic sequences in an appropriate translational context (e.g., *in vivo* versus *in vitro*), thereby transcending limitations of earlier approaches to BCEP. This enables development of vaccines and immunodiagnostics, most notably by selecting BCEs for inclusion among peptide-based constructs that elicit production of antipeptide antibodies. Such antibodies can mediate protective immunity (e.g., by neutralizing pathogen virulence factors *in vivo*) and/or be used for antigen detection (e.g., with the aid of surfactants and other chaotropic agents to increase the extent of AD among pathogen-derived proteins *in vitro*). BCEP thus supports the design of peptide-based constructs that are potentially useful as vaccine components, as companion immunodiagnostics for monitoring antipeptide antibody responses to vaccination, and as means for generating antipeptide antibodies that, apart from mediating protective immunity (e.g., *via* active or passive immunization), may serve as immunodiagnostic reagents for antigen detection.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was funded by the University of the Philippines System, via One UP professorial chair grant 2019-100965.

REFERENCES

1. Hopp TP, Woods KR. Prediction of Protein Antigenic Determinants From Amino Acid Sequences. *Proc Natl Acad Sci USA* (1981) 78:3824–8. doi: 10.1073/pnas.78.6.3824
2. Arnon R. Synthetic Vaccines–a Dream or Reality. *Adv Exp Med Biol* (1972) 31:209–22. doi: 10.1007/978-1-4684-3225-1_17
3. Van Regenmortel MH. What is a B-Cell Epitope? *Methods Mol Biol* (2009) 524:3–20. doi: 10.1007/978-1-59745-450-6_1
4. Caoili SEC. Beyond B-Cell Epitopes: Curating Positive Data on Antipeptide Paratope Binding to Support Peptide-Based Vaccine Design. *Protein Pept Lett* (2021) 28:953–62. doi: 10.2174/0929866528666210218215624
5. El-Manzalawy Y, Dobbs D, Honavar VG. In Silico Prediction of Linear B-Cell Epitopes on Proteins. *Methods Mol Biol* (2017) 1484:255–64. doi: 10.1007/978-1-4939-6406-2_17
6. Ras-Carmona A, Pelaez-Prestel HF, Lafuente EM, Reche PA. BCEPS: A Web Server to Predict Linear B Cell Epitopes With Enhanced Immunogenicity and Cross-Reactivity. *Cells* (2021) 10:2744. doi: 10.3390/cells10102744
7. Zhu J, Gouru A, Wu F, Berzofsky JA, Xie Y, Wang T. BepiTBR: T-B Reciprocity Enhances B Cell Epitope Prediction. *iScience* (2022) 25:103764. doi: 10.1016/j.isci.2022.103764
8. Caoili SEC. Prediction of Variable-Length B-Cell Epitopes for Antipeptide Paratopes Using the Program HAPTIC. *Protein Pept Lett* (2022) 39:328–39. doi: 10.2174/0929866529666220203101808
9. Alghamdi W, Attique M, Alzahrani E, Ullah MZ, Khan YD. LBCEPred: A Machine Learning Model to Predict Linear B-Cell Epitopes. *Brief Bioinform* (2022) 23:bbac035. doi: 10.1093/bib/bbac035
10. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res* (2017) 45:W24–9. doi: 10.1093/nar/gkx346
11. da Silva BM, Myung Y, Ascher DB, Pires DEV. Epitope3D: A Machine Learning Method for Conformational B-Cell Epitope Prediction. *Brief Bioinform* (2022) 23:bbab423. doi: 10.1093/bib/bbab423
12. Rawal K, Sinha R, Abbasi BA, Chaudhary A, Nath SK, Kumari P, et al. Identification of Vaccine Targets in Pathogens and Design of a Vaccine Using Computational Approaches. *Sci Rep* (2021) 11:17626. doi: 10.1038/s41598-021-96863-x
13. Kesarwani V, Gupta R, Vetukuri RR, Kushwaha SK, Gandhi S. Identification of Unique Peptides for SARS-CoV-2 Diagnostics and Vaccine Development by an *In Silico* Proteomics Approach. *Front Immunol* (2021) 12:725240. doi: 10.3389/fimmu.2021.725240
14. Van Regenmortel MH, Pellequer JL. Predicting Antigenic Determinants in Proteins: Looking for Unidimensional Solutions to a Three-Dimensional Problem? *Pept Res* (1994) 7:224–8.
15. Blythe MJ, Flower DR. Benchmarking B Cell Epitope Prediction: Underperformance of Existing Methods. *Protein Sci* (2005) 14:246–8. doi: 10.1110/ps.041059505
16. Chen SW, Van Regenmortel MH, Pellequer JL. Structure-Activity Relationships in Peptide-Antibody Complexes: Implications for Epitope Prediction and Development of Synthetic Peptide Vaccines. *Curr Med Chem* (2009) 16:953–64. doi: 10.2174/092986709787581914
17. Van Regenmortel MHV. *HIV/AIDS: Immunochemistry, Reductionism and Vaccine Design: A Review of 20 Years of Research*. Cham, Switzerland: Springer Nature Switzerland (2019).
18. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of Putative Vaccine Candidates Against *Helicobacter pylori* Exploiting Exoproteome and Secretome: A Reverse Vaccinology Based Approach. *Infect Genet Evol* (2015) 32:280–91. doi: 10.1016/j.meegid.2015.03.027
19. Dwivedi P, Alam SI, Tomar RS. Secretome, Surfome and Immunome: Emerging Approaches for the Discovery of New Vaccine Candidates Against Bacterial Infections. *World J Microbiol Biotechnol* (2016) 32:155. doi: 10.1007/s11274-016-2107-3
20. MacRaild CA, Richards JS, Anders RF, Norton RS. Antibody Recognition of Disordered Antigens. *Structure* (2016) 24:148–57. doi: 10.1016/j.str.2015.10.028
21. Uversky VN, Van Regenmortel MHV. Mobility and Disorder in Antibody and Antigen Binding Sites do Not Prevent Immunochemical Recognition. *Crit Rev Biochem Mol Biol* (2021) 56:149–56. doi: 10.1080/10409238.2020.1869683
22. Trier N, Hansen P, Houen G. Peptides, Antibodies, Peptide Antibodies and More. *Int J Mol Sci* (2019) 20. doi: 10.3390/ijms20246289

23. MacRaild CA, Seow J, Das SC, Norton RS. Disordered Epitopes as Peptide Vaccines. *Pept Sci (Hoboken)* (2018) 110:e24067. doi: 10.1002/pep2.24067
24. Atcheson E, Reyes-Sandoval A. Protective Efficacy of Peptides From *Plasmodium vivax* Circumsporozoite Protein. *Vaccine* (2020) 38:4346–54. doi: 10.1016/j.vaccine.2020.03.063
25. Wang C, Li Y, Wang S, Yan X, Xiao J, Chen Y, et al. Evaluation of a Tandem *Chlamydia Psittaci* Pgp3 Multiepitope Peptide Vaccine Against a Pulmonary Chlamydial Challenge in Mice. *Microb Pathog* (2020) 147:104256. doi: 10.1016/j.micpath.2020.104256
26. Du X, Xue J, Jiang M, Lin S, Huang Y, Deng K, et al. A Multiepitope Peptide, Romp22, Encapsulated in Chitosan-PLGA Nanoparticles as a Candidate Vaccine Against *Acinetobacter Baumannii* Infection. *Int J Nanomedicine* (2021) 16:1819–36. doi: 10.2147/IJN.S296527
27. Chen Y, Wu Y, Qin L, Yu L, Luo H, Li Y, et al. T-B Cell Epitope Peptides Induce Protective Immunity Against *Mycoplasma pneumoniae* Respiratory Tract Infection in BALB/c Mice. *Immunobiology* (2021) 226:152077. doi: 10.1016/j.imbio.2021.152077
28. Verma S, Singh K, Bansal A. Multi-Epitope DnaK Peptide Vaccine Accords Protection Against Lethal *S. typhimurium* Challenge: Elicits Both Cell Mediated Immunity and Long-Lasting Serum-Neutralizing Antibody Titers. *Pharmacol Res* (2021) 169:105652. doi: 10.1016/j.phrs.2021.105652
29. Deng K, Xu Z, Chen M, Liu X. Keyhole Limpet Hemocyanin-Conjugated Peptides From Hepatitis C Virus Glycoproteins Elicit Neutralizing Antibodies in BALB/c Mice. *J Immunol Res* (2021) 2021:3108157. doi: 10.1155/2021/3108157
30. Lei L, Qin H, Luo J, Tan Y, Yang J, Pan Z. Construction and Immunological Evaluation of Hepatitis B Virus Core Virus-Like Particles Containing Multiple Antigenic Peptides of Respiratory Syncytial Virus. *Virus Res* (2021) 298:198410. doi: 10.1016/j.virusres.2021.198410
31. Murdocca M, Citro G, Romeo I, Lupia A, Miersch S, Amadio B, et al. Peptide Platform as a Powerful Tool in the Fight Against COVID-19. *Viruses* (2021) 13:1667. doi: 10.3390/v13081667
32. Patarroyo ME, Patarroyo MA, Alba MP, Pabon L, Rugeles MT, Aguilar-Jimenez W, et al. The First Chemically-Synthesised, Highly Immunogenic Anti-SARS-CoV-2 Peptides in DNA *Genotyped aotus* Monkeys for Human Use. *Front Immunol* (2021) 12:724060. doi: 10.3389/fimmu.2021.724060
33. Long Q, Yang Y, Yang M, Bai H, Sun W, Yang X, et al. Recombinant VLPs Empower RBM Peptides Showing No Immunogenicity in Native SARS-COV-2 Protein to Elicit a Robust Neutralizing Antibody Response. *Nanomedicine* (2022) 41:102527. doi: 10.1016/j.nano.2022.102527
34. Zykova AA, Blokhina EA, Stepanova LA, Shuklina MA, Tsybalova LM, Kuprianov VV, et al. Nanoparticles Based on Artificial Self-Assembling Peptide and Displaying M2e Peptide and Stalk HA Epitopes of Influenza A Virus Induce Potent Humoral and T-Cell Responses and Protect Against the Viral Infection. *Nanomedicine* (2022) 39:102463. doi: 10.1016/j.nano.2021.102463
35. Bonduelle O, Chaudesaigues C, Tolazzi M, Suleiman E, de Bernard S, Alves K, et al. Dichotomy in Neutralizing Antibody Induction to Peptide-Conjugated Vaccine in Squalene Emulsion Contrast With Aluminum Hydroxide Formulation. *Front Immunol* (2022) 13:848571. doi: 10.3389/fimmu.2022.848571
36. Dunér P, Mattisson IY, Fogelstrand P, Glise L, Ruiz S, Farina C, et al. Antibodies Against apoB100 Peptide 210 Inhibit Atherosclerosis in apoE$^{-/-}$ Mice. *Sci Rep* (2021) 11:9022. doi: 10.1038/s41598-021-88430-1
37. Abbott RK, Crotty S. Factors in B Cell Competition and Immunodominance. *Immunol Rev* (2020) 296:120–31. doi: 10.1111/imr.12861
38. Watanabe A, Su KY, Kuraoka M, Yang G, Reynolds AE, Schmidt AG, et al. Self-Tolerance Curtails the B Cell Repertoire to Microbial Epitopes. *JCI Insight* (2019) 4:e122551. doi: 10.1172/jci.insight.122551
39. Finney J, Watanabe A, Kelsoe G, Kuraoka M. Minding the Gap: The Impact of B-Cell Tolerance on the Microbial Antibody Repertoire. *Immunol Rev* (2019) 292:24–36. doi: 10.1111/imr.12805
40. Freidl R, Gstöttner A, Baranyi U, Swoboda I, Stolz F, Focke-Tejkl M, et al. Resistance of Parvalbumin to Gastrointestinal Digestion is Required for Profound and Long-Lasting Prophylactic Oral Tolerance. *Allergy* (2020) 75:326–35. doi: 10.1111/all.13994
41. Vatti A, Monsalve DM, Pacheco Y, Chang C, Anaya JM, Gershwin ME. Original Antigenic Sin: A Comprehensive Review. *J Autoimmun* (2017) 83:12–21. doi: 10.1016/j.jaut.2017.04.008
42. Petráš M, Králová Lesná I. SARS-CoV-2 Vaccination in the Context of Original Antigenic Sin. *Hum Vaccin Immunother* (2022) 18:1949953. doi: 10.1080/21645515.2021.1949953
43. Tobin GJ, Trujillo JD, Bushnell RV, Lin G, Chaudhuri AR, Long J, et al. Deceptive Imprinting and Immune Refocusing in Vaccine Design. *Vaccine* (2008) 26:6189–99. doi: 10.1016/j.vaccine.2008.09.080
44. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular Mimicry and Autoimmunity. *J Autoimmun* (2018) 95:100–23. doi: 10.1016/j.jaut.2018.10.012
45. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly Accurate Protein Structure Prediction With AlphaFold. *Nature* (2021) 596:583–9. doi: 10.1038/s41586-021-03819-2
46. Ruff KM, Pappu RV. AlphaFold and Implications for Intrinsically Disordered Proteins. *J Mol Biol* (2021) 433:167208. doi: 10.1016/j.jmb.2021.167208
47. Morales RAV, MacRaild CA, Seow J, Krishnarjuna B, Drinkwater N, Rouet R, et al. Structural Basis for Epitope Masking and Strain Specificity of a Conserved Epitope in an Intrinsically Disordered Malaria Vaccine Candidate. *Sci Rep* (2015) 5:10103. doi: 10.1038/srep10103
48. Casadevall A. Antibody-Based Vaccine Strategies Against Intracellular Pathogens. *Curr Opin Immunol* (2018) 53:74–80. doi: 10.1016/j.coi.2018.04.011
49. Bottermann M, Caddy SL. Virus Neutralisation by Intracellular Antibodies. *Semin Cell Dev Biol* (2021) 126:108–16. doi: 10.1016/j.semcdb.2021.10.010
50. Herrmann A, Nagao T, Zhang C, Lahtz C, Li YJ, Yue C, et al. An Effective Cell-Penetrating Antibody Delivery Platform. *JCI Insight* (2019) 4:e127474. doi: 10.1172/jci.insight.127474
51. Schneider AFL, Kithil M, Cardoso MC, Lehmann M, Hackenberger CPR. Cellular Uptake of Large Biomolecules Enabled by Cell-Surface-Reactive Cell-Penetrating Peptide Additives. *Nat Chem* (2021) 13:530–9. doi: 10.1038/s41557-021-00661-x
52. Koester SK, Bolton WE. Intracellular Markers. *J Immunol Methods* (2000) 243:99–106. doi: 10.1016/s0022-1759(00)00239-8
53. Rieder FJ, Biebl J, Kastner MT, Schneider M, Jungbauer C, Redlberger-Fritz M, et al. Microbial Cryptotopes are Prominent Targets of B-Cell Immunity. *Sci Rep* (2016) 6:31657. doi: 10.1038/srep31657
54. Ferreira MU, Katzin AM. The Assessment of Antibody Affinity Distribution by Thiocyanate Elution: A Simple Dose-Response Approach. *J Immunol Methods* (1995) 187:297–305. doi: 10.1016/0022-1759(95)00186-4
55. Sormanni P, Aprile FA, Vendruscolo M. Rational Design of Antibodies Targeting Specific Epitopes Within Intrinsically Disordered Proteins. *Proc Natl Acad Sci USA* (2015) 112:9902–7. doi: 10.1073/pnas.1422401112
56. Tomita T. Structural and Biochemical Elements of Efficiently Degradable Proteasome Substrates. *J Biochem* (2022) 171:261–8. doi: 10.1093/jb/mvab157
57. Min SE, Lee KH, Park SW, Yoo TH, Oh CH, Park JH, et al. Cell-Free Production and Streamlined Assay of Cytosol-Penetrating Antibodies. *Biotechnol Bioeng* (2016) 113:2107–12. doi: 10.1002/bit.25985
58. Caoili SE. Antibodies, Synthetic Peptides and Related Constructs for Planetary Health Based on Green Chemistry in the Anthropocene. *Future Sci OA* (2018) 4:FSO275. doi: 10.4155/fsoa-2017-0101
59. Caoili SE. A Structural-Energetic Basis for B-Cell Epitope Prediction. *Protein Pept Lett* (2006) 13:743–51. doi: 10.2174/092986606777790502
60. Van Regenmortel MH. Immunoinformatics may Lead to a Reappraisal of the Nature of B Cell Epitopes and of the Feasibility of Synthetic Peptide Vaccines. *J Mol Recognit* (2006) 19:183–7. doi: 10.1002/jmr.768
61. Rahman K, Chowdhury EU, Sachse K, Kaltenboeck B. Inadequate Reference Datasets Biased toward Short Non-epitopes Confound B-cell Epitope Prediction. *J Biol Chem* (2016) 291:14585–99. doi: 10.1074/jbc.M116.729020

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 08 July 2022
doi: 10.3389/fmed.2022.909660

Presumed Protective Role for Anti-Hepatitis B Virus Antibodies Against COVID-19 Severe Cases: A Clinical Study Confirming *in silico* Hypothesis

Mariem Gdoura[1,2,3]*, *Raoua Touati*[1,2], *Sana Kalthoum*[4], *Rania Ben Slama*[1,2], *Nouel Fatnassi*[2,4], *Mehdi Mrad*[5], *Lamia Ammari*[2,6], *Nozha Brahmi*[7], *Amira Ben Jazia*[7], *Nahed Hogga*[1,2], *Henda Triki*[1,2] and *Sondes Haddad-Boubaker*[1,2]*

[1] Laboratory of Clinical Virology, WHO Regional Reference Laboratory for Poliomyelitis and Measles for the EMR, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis, Tunisia, [2] LR20IPT10 Laboratory of Virus, Host and Vectors, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis, Tunisia, [3] Department of Clinical Biology, Faculty of Pharmacy of Monastir, University of Monastir, Monastir, Tunisia, [4] Centre National de Veille Zoosanitaire, Tunis, Tunisia, [5] Laboratory of Biochemistry, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis, Tunisia, [6] Infectious Diseases Departement, Rabta Hospital, Tunis, Tunisia, [7] Intensive Care Service, Emergence Medical Assistance Center, Tunis, Tunisia

OPEN ACCESS

Edited by:
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

Reviewed by:
Ahmed Yaqinuddin,
Alfaisal University, Saudi Arabia
Mohammad Uzzal Hossain,
National Institute of Biotechnology (NIB), Bangladesh
Waleed Mahallawi,
Taibah University, Saudi Arabia

***Correspondence:**
Mariem Gdoura
mariemgdoura@gmail.com
Sondes Haddad-Boubaker
sondeshaddadboubaker@gmail.com

Specialty section:
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *31 March 2022*
Accepted: *13 June 2022*
Published: *08 July 2022*

Citation:
Gdoura M, Touati R, Kalthoum S, Ben Slama R, Fatnassi N, Mrad M, Ammari L, Brahmi N, Ben Jazia A, Hogga N, Triki H and Haddad-Boubaker S (2022) Presumed Protective Role for Anti-Hepatitis B Virus Antibodies Against COVID-19 Severe Cases: A Clinical Study Confirming in silico Hypothesis. Front. Med. 9:909660. doi: 10.3389/fmed.2022.909660

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for COVID-19 disease which is known to have a broad clinical spectrum, from asymptomatic to critical presentation leading to death. Many researchers have investigated the factors impacting the course of the disease. Our previous *in silico* study suggested a possible protective effect of Hepatitis B, Tetanus and Measles vaccines against COVID-19. In continuity, we conducted a cross-sectional clinical study in order to confirm our *in silico* assumptions regarding the HBs-Ag antibodies.

Methods: A representative sex- and age-matched sample of patients with confirmed COVID-19 was selected ($n = 340$). All clinical presentations were equally represented. Using an ELISA test, each patient benefited of a serology for the detection and measurement of the anti-HBs specific IgG antibodies. The obtained results allowed determining the different correlations between these antibody titers and the disease severity. The R® software and the MedCalc® software served to calculate the Spearman's coefficient of rank correlation (rho) for the obtained titers per severity group as well as the different other calculations and figure representations.

Results: A significant positive correlation was found with the anti-HBs titers (rho $= 0.107$; $p = 0.04$). High anti-HBs titers were significantly associated with the mild presentation of COVID-19. A significant difference was found between the obtained titers per severity class (chi-2 test, $p = 0.03$).

Discussion/Conclusion: Our findings demonstrated that anti-HBs titers were significantly higher for patients having mild COVID-19 presentations. We presume that being immunized against the HB may play a protective role in the course of the disease. Our study provided more key elements in understanding the disparity of the clinical spectrum among regions.

Keywords: COVID-19, correlation, antibody titer, Hepatitis B, SARS-CoV-2

INTRODUCTION

The emergence and the spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for COVID-19, have caused global public health, economic, and social crises (1, 2). While most cases of COVID-19 are mild to moderate, a significant fraction may develop severe to critical clinical presentations leading to death (3). As of 29 March 2022, 6,127,981 COVID-19-related deaths, worldwide, have been reported to the WHO (4). An uneven mortality distribution across the WHO regions is noteworthy. This unequal distribution is even deeper between countries throughout the regions as they are reporting very different case fatality ratios (5). As an example, within the region of the Americas, Venezuela reported only 19 deaths per 100,000 inhabitants, whereas the rates reported by Peru exceeded 600 per 100,000 inhabitants (4). Thomas et al. (6) have gone even further by examining 19 different cities in the United States and concluded to non-uniform COVID-19 severity within the same city level. Presently, almost 2 years after the virus emergence, we are still learning about the risk factors of severe COVID-19 outcomes (7). Demographic risk factors such as advanced age as well as underlying medical conditions have evident impact on the COVID-19 severity (8). In addition, many other assumptions have been made in order to explain these substantial striking disparities such as human leukocyte antigen (HLA) polymorphism, race and ethnicity, environmental conditions, nutrition, and microbiome (9–12). The most plausible hypotheses were those involving the immune response mechanisms (13, 14), and it was suggested that some vaccines could enhance the innate immune response (15, 16). Indeed, even before the emergence of the COVID-19 pandemic, it was described that the large-scale use of the BCG vaccine (Bacillus Calmette-Guérin) has beneficial nonspecific effects on the immune system by protecting against other infectious diseases and reducing the non-tuberculosis child mortality (17). Recently, Curtis et al. (18) suggested that it might be involved in the protection against COVID-19 as it could reduce viraemia and thus enhance rapid recovery and less severity. Anbarasu et al. (19) also suggested that the extensive pediatric vaccination with MMR vaccines (Measles, Mumps, and Rubella vaccines) has induced interferon (IFN) secretion and activated natural killer cells, offering, thereby, natural immunity against SARS-CoV-2 in the young population.

In a previous paper, we investigated the putative protective role against COVID-19 of 12 widely used vaccines, including live attenuated (BCG, Oral Polio Vaccine, MMR vaccines) and inactivated ones [tetanus, *Corynebacterium diphtheriae*, *Bordetella pertussis*, Hepatitis B (HB), Hepatitis A, *Haemophilus influenzae* type B (Hib), and *Streptococcus pneumoniae* vaccines (PCV10)] (20). A total of 30 antigenic proteins were investigated. Using a package of the *in silico* analysis tools, we performed amino acid sequence alignments and hot spot analysis. Among the investigated antigenic proteins, 14 proteins presented similar amino acid patterns in eight different vaccines. Structural and antigenicity tests (B-cell and T-cell epitope predictions) identified three segments in Hepatitis B, Measles and Tetanus proteins presented antigenic properties that can induce putative protective effect against SARS-CoV2 (**Figure 1**).

Hepatitis B is a vaccine-preventable disease; however, it is the world's most common liver infection. The WHO estimates that 296 million people were living with chronic HB infection in 2019, with 1.5 million new infections each year (21). National and international efforts are being made in order to improve the vaccine coverage through systematic childhood and health workers vaccination. We were wondering whether this unequal immunization between countries may justify the COVID-19 severity variability. Within this vision, and in continuity to the previous *in silico* study (20), we investigate in this work the putative protective role of the anti-HBs-Ag specific IgG of the HBV against SARS-CoV-2 using real samples from patients who recovered from COVID-19. Our aim was to identify the statistical correlation between the corresponding antibody titers and the disease severity among a large sample size.

METHODS

Ethics Statement

This study was performed under ethical standards according to the 1964 Declaration of Helsinki and its later amendments. The samples were collected in the context of COVID-19 diagnostic activities. They were used in this study after de-identification with respect to patient anonymity and after the approval of the Bio-Medical Ethics Committee of Pasteur Institute of Tunis, Tunisia (2020/14/I/LR16IPT/V1).

Study Population

This cross-sectional study was conducted between May and June, 2021, in the Laboratory of Clinical Virology (LCV) of Pasteur Institute of Tunis. The included sera were randomly selected from the LCV bio-bank; they were collected during the pandemic from patients with different clinical presentations and then carefully stored in a −40°C freezer in accordance with the BAOBAB® LIMS storing application (22). A total of 340 patients with COVID-19 infection were enrolled, matched for age and sex, and classified into "asymptomatic," "mild," "moderate," and "severe" according to the United Stated National Institutes of Health (NIH) definitions update of 19 October 2021 (3). Each group contained 85 patients. All selected patients were not vaccinated against SARS-CoV-2 and the infection with COVID-19 was confirmed either by real-time reverse transcription PCR (RT-qPCR) assessed on nasopharyngeal swab by a WHO-approved in-house protocol (the Hong Kong University, China, RT-qPCR protocol) (23) or by SARS-CoV-2-specific antibody seroconversion detected by the commercial test Elecsys® Roche® total anti-nucleoprotein antibodies.

Clinical Data Collection

The socio-demographic data of patients and information on clinical features, co-morbidities, and exposure or contact history with COVID-19 patients were collected.

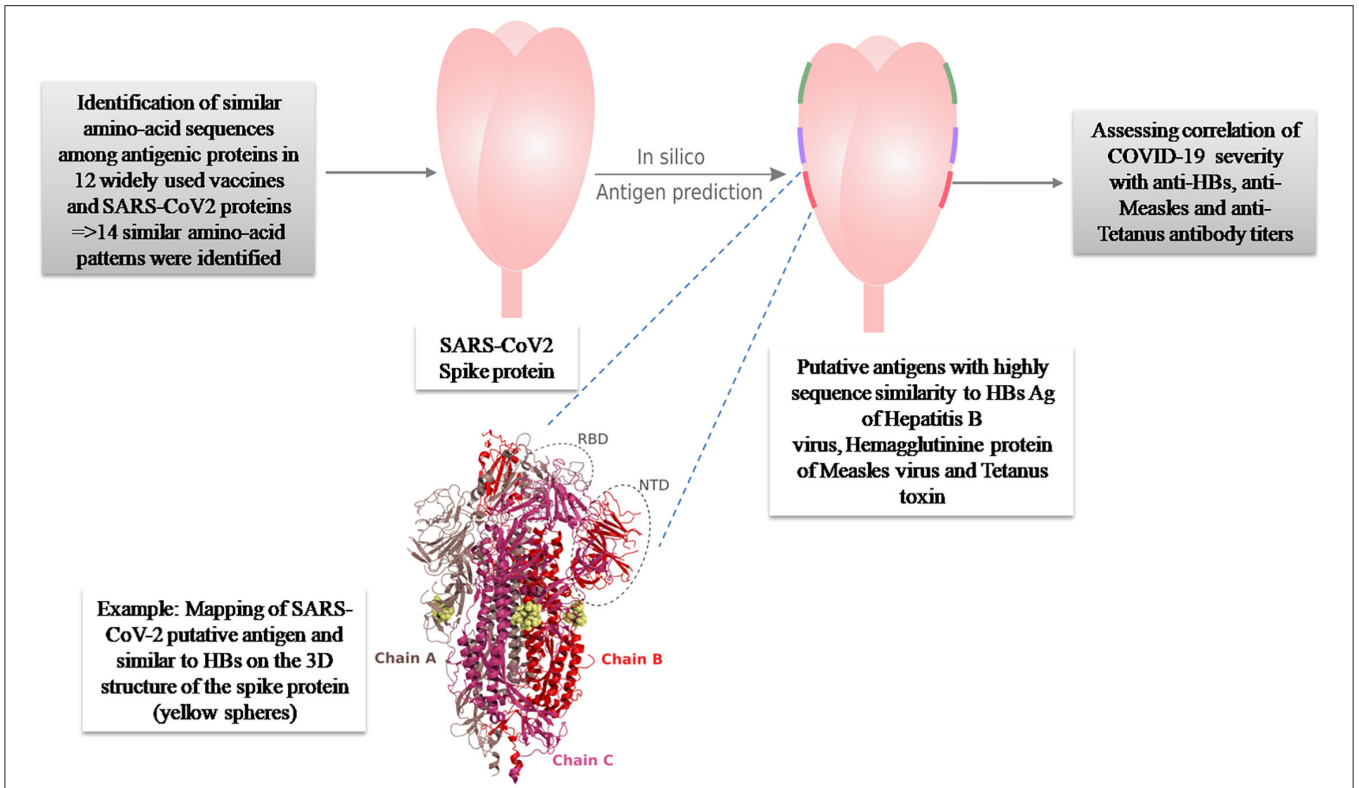


FIGURE 1 | Summary of the original *in silico* hypothesis suggesting an antigenic potential of three amino-acid patterns present in HBs Ag of Hepatitis B, Hemagglutinine protein of Measles virus and Tetanus toxin, against SARS-CoV-2 (20).

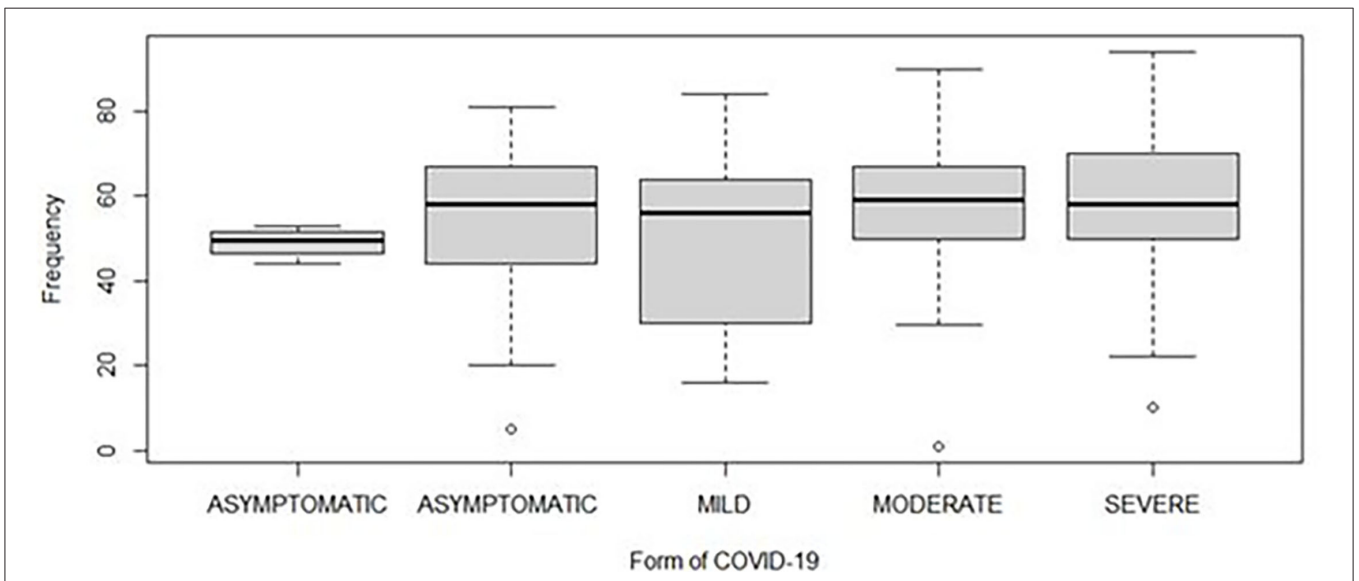


FIGURE 2 | Distribution of the age of patients according to different severities of COVID-19. The horizontal line inside each box represents the median value of the age with interquartile ranges. The dots represent patients.

Detection and Measurement of the Anti-HBs

Detection and measurement of the anti-HBs IgG-specific antibodies were carried out by the commercial *in vitro* diagnostics (IVD)-validated immunoassay: an anti-HBs enzyme immunoassay kit, ETI-AB-AUK-3, manufactured by the DiaSorin® S.p.A., Italy (REF: P001603). This assay is based on a direct and non-competitive sandwich ELISA. It enables

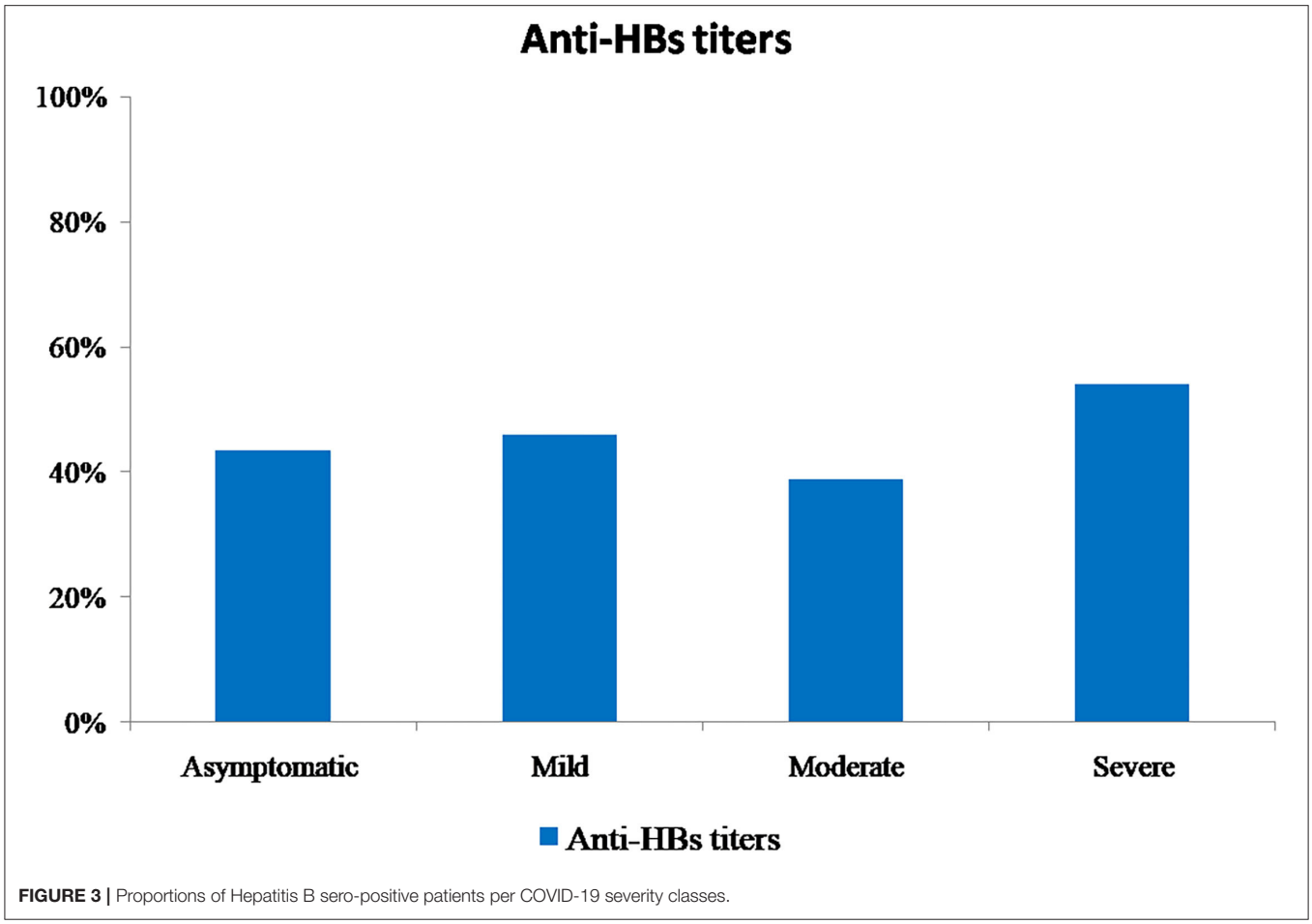


FIGURE 3 | Proportions of Hepatitis B sero-positive patients per COVID-19 severity classes.

detection of anti-HBs IgG-specific antibodies using wells coated with heat-treated human HBs-Ag. The measurement of the anti-HBs specific antibodies depends on the use of standardized calibrators that were referenced against WHO anti-Hepatitis B Immunoglobulin 1st International Reference Preparation, 1977. This kit is recommended for measuring the anti-HBs titers whether acquired as a result of natural HBV infection or after vaccination. The sensitivity and specificity of this kit are 99.11% CI 95% (98.18–99.64%) and 98.21% CI 95% (97.07–99.00%), respectively, following the manufacturer's instructions. The testing procedures and result interpretation were conducted according to the kit instructions: a titer lower than 10 mIU/mL indicates that the patient is not immunized against the HBV; a titer higher than 10 mIU/mL is correlated to an efficient immunity against the virus; however, a titer more than 100 mIU/mL is recommended for the vulnerable populations such as the health care workers. This classification also aligns with the WHO recommendations (24).

Statistical Analysis

To explore the relationship between anti-HBs specific antibody titers and the severity of COVID-19 cases, the sample size was first calculated with 80% power of the test, an alpha value set to 0.05, and a correlation value of 0.3 using Epitools® (25). The sample size was calculated using the Sample Size Calculators website (26). The Spearman's rank correlation coefficient was used to determine the correlation between anti-HBs specific antibody levels and the SARS-CoV-2 severity of cases. Continuous data were presented in median and ranges and categorical data were presented in numbers and percentages. All statistical analyses were performed using the MedCalc® Software (version 18.2.1) and the R® software (version 3.4). A p-value of <0.05 was considered statistically significant.

RESULTS

Demographic Features of the Tested Population

This study included 340 patients that were stratified according to COVID-19 disease severity into four groups: "asymptomatic," "mild," "moderate," and "severe." The required sample size was 85 patients in each group. Among the 340 patients, 160 were female (47%) and 180 (53%) were male. The sex ratio male/female (M/F) was estimated at 1.1. The mean age for all groups was 54 years (1–94 range; **Figure 2**).

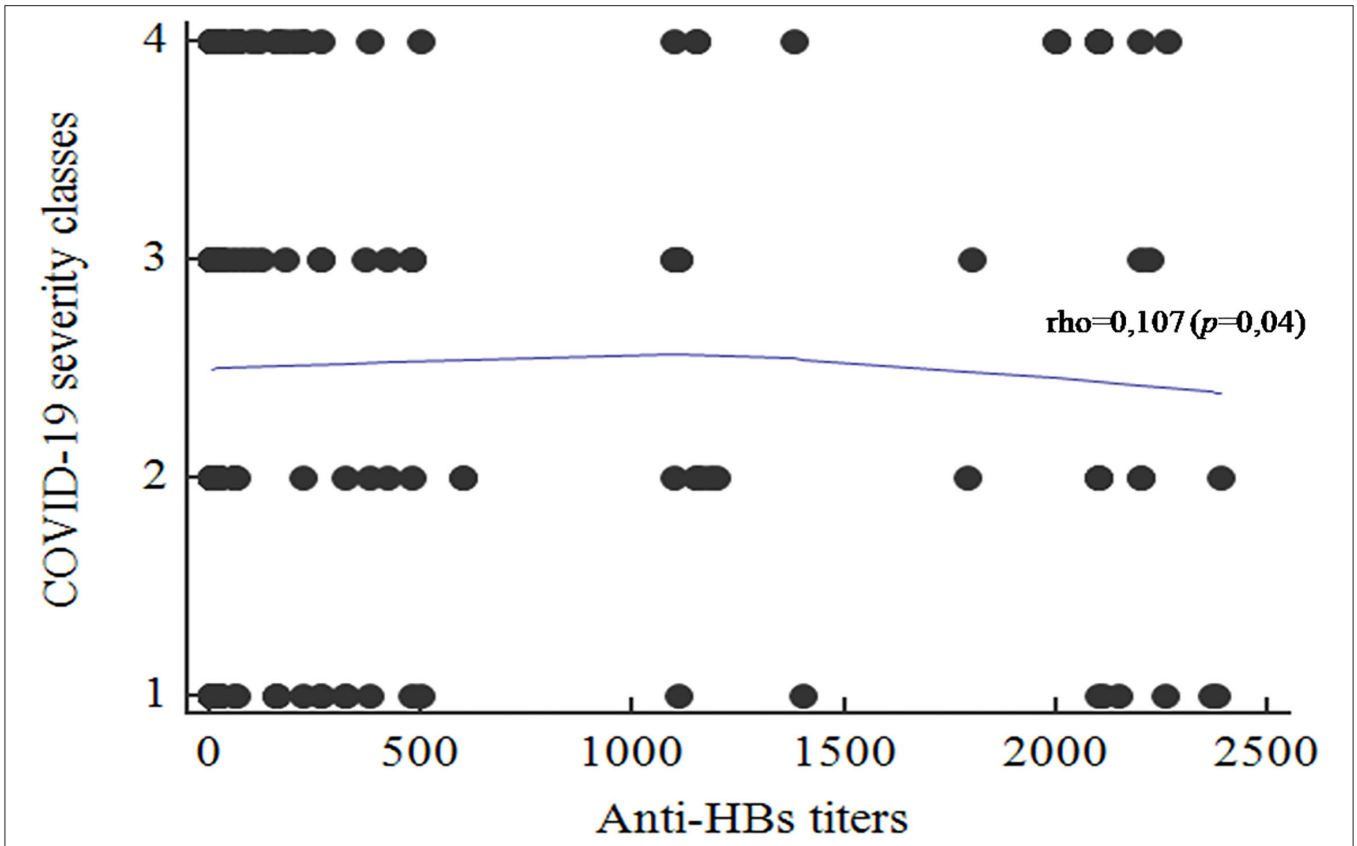


FIGURE 4 | Correlation between the anti-HBs titers and the COVID-19 disease severity classes (1, Asymptomatic; 2, Mild; 3, Moderate; 4, Severe.

Serological Results

All sera were tested for the presence of anti-HBs antibodies. In total, 54.4% ($n = 185$) of patients did not have detectable anti-HBs antibodies. For the remaining proportion of patients who had positive anti-HBs antibodies ($n = 155$, 45.6%), it was distributed according to the COVID-19 disease severity group as shown in **Figure 3**. Regarding the antibody measurement, the titers ranged from 14 to 2,390 mIU/mL, with a median of 100 mIU/mL. For the "asymptomatic" class, titers ranged from 15 to 2,380 mIU/mL; for the "mild" class, titers ranged from 15 to 2,390 mIU/mL; for the "moderate" class, titers ranged from 15 to 2,220 mIU/mL; and for the "severe" class, titers ranged from 14 to 2,265 mIU/mL. A significant difference was found between the obtained titers per severity class (chi-2 test, $p = 0.03$). In addition, a significant reverse correlation was found between the patients' ages and the obtained anti-HBs antibody titers (rho $= -0.176$; $p < 0.05$); this association demonstrated that higher anti-HBs titers were detected in children and the adult patients than that in older age.

Correlation Analysis

Taking into consideration the different obtained titers, the correlation with the COVID-19 severity classes was performed. A significant positive correlation was found between the anti-HBs titers and the COVID-19 severity classes (rho $= 0.107$; $p = 0.04$; **Figure 4**). High anti-HBs titers were significantly associated with mild presentation of COVID-19. In contrast, patients with severe clinical presentations had lower antibody titers.

DISCUSSION

In this article, we investigated the potential role of a routinely used vaccine against HBV in preventing severe clinical presentations of the COVID-19 disease. Our aim was to confirm our previous *in silico* findings. Indeed, we identified similar amino acid pattern in HBs-Ag protein of HBV and SARS-CoV-2 with antigenic properties (20) (**Figure 1**). We have attempted to identify whether there is a correlation between HBV antibody titers and the severity of SARS-CoV-2 infection, using sera of patients who recovered from COVID-19.

Several studies have previously suggested a potential cross protection of post-vaccine antibodies toward severe to critical outcomes of the SARS-CoV-2 infection (17–19, 27, 28) and, as the experimental identification of the incriminated epitopes is challenging, many studies had recourse to *in silico* approach. Indeed, bioinformatic and computational biology provides a valuable contribution in the current COVID-19 emergency context, thanks to its rapidity and low cost in comparison to wet-laboratory and clinical investigations (29). However, *in vitro* investigations remain a more solid way to support conclusions.

Thus, a total of 340 patients with confirmed SARS-CoV-2 infection were enrolled in this study, matched for age and sex, and classified into different clinical presentations ($n = 85$/group) according to the National Institute of Health definitions (3). To our knowledge, our study investigated the largest population for the determination of vaccine correlation as compared to previous ones and reported, for the first time, that the presence of high anti-HBs titers in patients' sera may be associated with a significant protective role against the COVID-19 disease. Gold et al. investigated only a total of 80 patients with COVID-19 for the relationship with MMR vaccine (30). Al Balakosy et al. also determined the anti-Measles IgG antibodies on a total of only 78 patients with COVID-19 (31).

The active substance in the HB vaccine is the viral surface protein HBs-Ag, obtained by the yeast-derived recombinant vaccine biotechnology, in most commercially available vaccines (32). The immunogenic fraction is the amino acid hydrophilic region, referred as the common *a* determinant present in the HBs-Ag (subtype HBs-Ag-adw2 or HBs-Ag-adr) (32, 33). These two antigenic epitopes were investigated in our previous *in silico* study and we showed that the amino-acid sequence PGTNTSN in the Spike protein of SARS-CoV-2 (position 600–606) matches with the predicted epitope TNTSN in the HBs-Ag-adr (20). The pattern PGTNTSN corresponded to an exposed site in the S protein with a high accessible surface area value and presented probing spheres mimicking the CDR antibodies, that was in line with a potential implication in the B-cell mediated response. Furthermore, it was also described by Tajiri et al. (34) that two regions of HBsAg (residues 104–123 and 108–123), containing the epitope matching the PGTNTSN segment of SARS-CoV-2, were able to bind with two human monoclonal antibodies. This highlighted the immunogenic ability of these segments. Indeed, high-antibody titers were found among the studied population within the mildly infected group, with a significant positive correlation (rho $= 0.107$; $p = 0.04$) which is in concordance with our previous *in silico* findings (20). Therefore, we conclude that epitopes in regions of HBsAg (residues 104–123 and 108–123) matching the PGTNTSN segment of RBD may induce a cross protection against SARS-CoV-2. *In vivo*, this might lead either to a direct inhibition of the early cell-entry phases or to an indirect non-specific steric clutter. Nevertheless, only experiments such as sero-neutralization assay may support such findings.

Relying on clinical observations, Chen et al. (35) and Wu et al. (36) reported that SARS-CoV-2 infection in patients infected with HBV could facilitate the development of liver injury which is associated with adverse outcomes. Accordingly, it can also be suggested that HBV vaccination may also indirectly protect patients from these adverse outcomes (37). Furthermore, the possible protective role of HBV vaccine against other diseases, such as lymphoma, was reported by two previous studies (38, 39). Presently, the safe and efficient HB vaccination is highly recommended for all children worldwide. However, the WHO reports widely variable coverage rates from country to country, depending on the respective national strategies (24). From another point of view, it is well known, today, that after an initial vaccination, anti-HBs antibody titers may decline over years (40), making elderly less immunized than the other age groups. So, taking into consideration that the severe COVID-19 presentations are more frequent among elderly (8), we attempt to speculate that the loss of anti-HBs antibodies may be in line with our hypothesis. In our study, we performed a serum measurement of the anti-HBs antibodies regardless of the vaccination history because only the presence of anti-HBs at a protective level may indicate an effective protection whatever it was obtained through vaccination or natural infection. Furthermore, the non-response rates to the HBV vaccine range from 5 to 10% in total population and can reach 40% among patients with diabetes and kidney failure (40). For this study, the patients came from Tunisia which is a low-endemic country as result to the introduction of vaccination since 1995. Indeed, on the basis of the most recent national household-based cross sectional and serological survey in 2015, the national point prevalence of HBs-Ag was 1.7% [95% CI (1.6–1.9%)] and the vaccine effectiveness was 88.6% [95% CI (81.5–93.0%)] (41).

Epitopic similarity between different virus' antigens is a well-described phenomenon that has various implications on pathogenicity comprehension, diagnosis methods, and even treatment opportunities. It has been demonstrated during the emergence of Zika virus, for example, that there is a strong structural similarity with the matrix and the envelope antigens of dengue and West Nile viruses, i.e., within the same family of Flaviviridae (41). It has been reported, recently, that these antigenic cross-reactivities have impeded the IgM-specific antibody serology assays (42). Furthermore, beyond the viruses, it was demonstrated that there is a molecular mimicry between self-human and viral antigens, which might trigger autoimmune diseases in genetically predisposed individuals (43). For instances, viruses such as Coxsackievirus, Mumps virus, Rubella virus, and Hepatitis C virus were incriminated in inducing the type 2 diabetes, again by exhibiting molecular mimicry with the host proteins (44). For the SARS-CoV-2, the previously reported epitopic similarity (20) and the current findings may suggest a possible cross protection. Nevertheless, it should be supported by larger statistic investigation and further lab experiments such as sero-neutralization assays. Our findings may be unique and encourage other studies targeting previously used antiviral to be tested against SARS-CoV-2 infection, especially those used against HBV chronic infection. Indeed, García-Trejo et al. published an argument repurposing the Lamivudine (nucleotide/nucleoside analogs); they provided *in silico* docking evidence suggesting that Lamivudine may bind and possibly work as an inhibitor of the SARS-CoV-2 RdRp RNA polymerase (45). The telbivudine also was suggested as a fighting drug for COVID-19 (46). These data underline the possible similarity with gene sequences between HBV and SARS-CoV-2.

CONCLUSION

In the COVID-19 crisis context, clinical research is escalating and providing mounting evidence that immunity background plays a crucial role in deciding the course of the disease. Our findings have placed emphasis on HBV, linking anti-HBs

high sero-positivity to the COVID-19 minor severity, as its antigenic properties were consistent with its putative neutralizing capacity. Although findings were significant, larger population investigation may further support the obtained correlation. Also, it will be interesting to investigate the sero-neutralization effects of anti-HBs antibodies using different SARS-CoV-2 variants of interest.

The observed associations between anti-HBs antibody titers and the COVID-19 disease course may explain the geographical disparity worldwide of the COVID-19 severity, along with all the suggested risk factors. We believe that it is still important to dig into the protective and risk factors that have led to the large number of deaths inherent to COVID-19 especially in the context of SARS-CoV-2 variant emergence.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Bio-Medical Ethics Committee of Pasteur Institute of Tunis, Tunisia. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MG, SH-B, and SK: conceptualization, methodology, and original draft preparation. RT, RS, and NH: investigation. RT, SK, and NF: data curation. MM, LA, NB, and AJ: sample and data collection. MG, SH-B, SK, and HT: writing—review and editing, supervision, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Tunisian Ministry of Higher Education and Scientific Research (LR20IPT10).

ACKNOWLEDGMENTS

The authors gratefully acknowledge Houcemeddine Othman from Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa, for his valuable scientific contribution in this work, Ali Bouattour for logistical support, Ezzeddine Nouiri and Nejla Mekki for technical advises, Ali Neili and Wafa Chamsa for the design of the bio-bank by the BAOBAB® LIMS, and Habib Halouani and Donia Sahli for the database entry.

REFERENCES

1. Pollard CA, Morran MP, Nestor-Kalinoski AL. The COVID-19 pandemic: a global health crisis. *Physiol Genomics*. (2020) 52:549–57. doi: 10.1152/physiolgenomics.00089.2020
2. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. *Int J Surg*. (2020) 78:185–93. doi: 10.1016/j.ijsu.2020.04.018
3. Clinical Spectrum. *COVID-19 Treatment Guidelines*. Available online at: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/ (accessed Mars 3, 2022).
4. *WHO Coronavirus (COVID-19)* Dashboard. Available online at: https://covid19.who.int (accessed Mars 3, 2022).
5. Mortality Analyses. *Johns Hopkins Coronavirus Resource Center*. Available online at: https://coronavirus.jhu.edu/data/mortality (accessed Févr 26, 2022).
6. Thomas LJ, Huang P, Yin F, Luo XI, Almquist ZW, Hipp JR, et al. Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proc Natl Acad Sci USA*. (2020) 117:24180–7. doi: 10.1073/pnas.2011656117
7. CDC. *Healthcare* Workers. Centers for Disease Control and Prevention. (2020). Available online at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html (accessed mars 3, 2022).
8. Rossen LM. Excess deaths associated with COVID-19, by age and race and ethnicity — United States, January 26–October 3, 2020. *MMWR Morb Mortal Wkly Rep*. (2020) 69:1522–7. doi: 10.15585/mmwr.mm6942e2
9. Mackey K, Ayers CK, Kondo KK, Saha S, Advani SM, Young S, et al. Racial and ethnic disparities in COVID-19-related infections, hospitalizations, and deaths : a systematic review. *Ann Intern Med*. (2021) 174:362–73. doi: 10.7326/M20-6306
10. Sakuraba A, Haider H, Sato T. Population difference in allele frequency of HLA-C*05 and its correlation with COVID-19 mortality. *Viruses*. (2020) 12:E1333. doi: 10.3390/v12111333
11. Ganslmeier M, Furceri D, Ostry JD. The impact of weather on COVID-19 pandemic. *Sci Rep*. (2021) 11:22027. doi: 10.1038/s41598-021-01189-3
12. Kumar P, Chander B. COVID 19 mortality: probable role of microbiome to explain disparity. *Med Hypotheses*. (2020) 144:110209. doi: 10.1016/j.mehy.2020.110209
13. Melenotte C, Silvin A, Goubet AG, Lahmar I, Dubuisson A, Zumla A, et al. Immune responses during COVID-19 infection. *Oncoimmunology*. (2020) 9:1807836. doi: 10.1080/2162402X.2020.1807836
14. Esenboga S, Ocak M, Akarsu A, Bildik HN, Cagdas D, Iskit AT, et al. COVID-19 in patients with primary immunodeficiency. *J Clin Immunol*. (2021) 41:1515–22. doi: 10.1007/s10875-021-01065-9
15. Mina MJ. Measles, immune suppression and vaccination: direct and indirect nonspecific vaccine benefits. *J Infect*. (2017) 74(Suppl 1):S10–7. doi: 10.1016/S0163-4453(17)30185-8
16. Fidel PL, Noverr MC. Could an unrelated live attenuated vaccine serve as a preventive measure to dampen septic inflammation associated with COVID-19 infection? *MBio*. (2020) 11:e00907–20. doi: 10.1128/mBio.00907-20
17. Higgins JPT, Soares-Weiser K, López-López JA, Kakourou A, Chaplin K, Christensen H, et al. Association of BCG, DTP, and measles containing vaccines with childhood mortality: systematic review. *BMJ*. (2016) 355:i5170. doi: 10.1136/bmj.i5170
18. Curtis N, Sparrow A, Ghebreyesus TA, Netea MG. Considering BCG vaccination to reduce the impact of COVID-19. *Lancet*. (2020) 395:1545–6. doi: 10.1016/S0140-6736(20)31025-4
19. Anbarasu A, Ramaiah S, Livingstone P. Vaccine repurposing approach for preventing COVID 19: can MMR vaccines reduce morbidity and mortality? *Hum Vaccin Immunother*. 16:2217–8. doi: 10.1080/21645515.2020.1773141
20. Haddad-Boubaker S, Othman H, Touati R, Ayouni K, Lakhal M, Ben Mustapha I, et al. In silico comparative study of SARS-CoV-2 proteins

and antigenic proteins in BCG, OPV, MMR and other vaccines: evidence of a possible putative protective effect. *BMC Bioinformatics*. (2021) 22:163. doi: 10.1186/s12859-021-04045-3
21. Hepatitis B. Available online at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b (accessed Mai 13, 2022).
22. Bendou H, Sizani L, Reid T, Swanepoel C, Ademuyiwa T, Merino-Martinez R, et al. Baobab laboratory information management system: development of an open-source laboratory information management system for biobanking. *Biopreserv Biobank*. (2017) 15:116–20. doi: 10.1089/bio.2017.0014
23. WHO IN HOUSE ASSAYS. Available online at: https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf (accessed Févr 23, 2022).
24. Guidelines_development_recommendations.pdf. Available online at: https://www.who.int/immunization/sage/Guidelines_development_recommendations.pdf?ua=1 (accessed Mars 3, 2022).
25. Epitools - Home. Available online at: https://epitools.ausvet.com.au/ (accessed Mars 7, 2022).
26. Kohn JS Michael. *Correlation sample size | Sample Size* Calculators. Available online at: https://sample-size.net/correlation-sample-size/ (accessed Mars 7, 2022).
27. Young A, Neumann B, Mendez RF, Reyahi A, Joannides A, Modis Y, et al. Homologous protein domains in SARS-CoV-2 and measles, mumps and rubella viruses: preliminary evidence that MMR vaccine might provide protection against COVID-19. *medRxiv. [Preprint]. Infect Dis*. (2020). doi: 10.1101/2020.04.10.20053207
28. Marakasova E, Baranova A, MMR. Vaccine and COVID-19: Measles protein homology may contribute to cross-reactivity or to complement activation protection. *MBio*. (2021) 12:e03447–20. doi: 10.1128/mBio.03447-20
29. Basu S, Ramaiah S, Anbarasu A. *In-silico* strategies to combat COVID-19: a comprehensive review. *Biotechnol Genet Eng Rev*. (2021) 37:64–81. doi: 10.1080/02648725.2021.1966920
30. Gold JE, Baumgartl WH, Okyay RA, Licht WE, FidelJr PL, Noverr MC, et al. Analysis of Measles-Mumps-Rubella (MMR) titers of recovered COVID-19 patients. *mBio*. (2020) 11:e02628–20. doi: 10.1128/mBio.02628-20
31. Al Balakosy A, Alfishawy M, Elnabawy O, Hassan A, Shamkh M, Mahmoud M, et al. Measles IgG antibodies: is there a protective role in COVID 19 pandemic?? ?*Afro-Egypt J Infect Endem Dis*. (2021) 11:306–13. doi: 10.21608/aeji.2021.72615.1144
32. Heijtink RA, Bergen P, Melber K, Janowicz ZA, Osterhaus AD. Hepatitis B surface antigen (HBsAg) derived from yeast cells (*Hansenula polymorpha*) used to establish an influence of antigenic subtype (adw2, adr, ayw3) in measuring the immune response after vaccination. *Vaccine*. (2002) 20:2191–6. doi: 10.1016/S0264-410X(02)00145-7
33. Romanò L, Paladini S, Galli C, Raimondo G, Pollicino T, Zanetti AR. Hepatitis B vaccination. *Hum Vaccin Immunother*. (2015) 11:53–7. doi: 10.4161/hv.34306
34. Tajiri K, Ozawa T, Jin A, Tokimitsu Y, Minemura M, Kishi H, et al. Analysis of the epitope and neutralizing capacity of human monoclonal antibodies induced by hepatitis B vaccine. *Antiviral Res*. (2010) 87:40–9. doi: 10.1016/j.antiviral.2010.04.006
35. Chen X, Jiang Q, Ma Z, Ling J, Hu W, Cao Q, et al. Clinical characteristics of hospitalized patients with SARS-CoV-2 and hepatitis B virus co-infection. *Virol Sin*. (2020) 35:842–5. doi: 10.1007/s12250-020-00276-5
36. Wu J, Yu J, Shi X, Li W, Song S, Zhao L, et al. epidemiological and clinical characteristics of 70 cases of coronavirus disease and concomitant hepatitis B virus infection: a multicentre descriptive study. *J Viral Hepatitis*. (2021) 28:80–8. doi: 10.1111/jvh.13404
37. Xiang TD, Zheng X. Interaction between hepatitis B virus and SARS-CoV-2 infections. *World J Gastroenterol*. (2021) 27:782–93. doi: 10.3748/wjg.v27.i9.782
38. Lia L, Grima D, Amici F, Manzi L, Monaci A, Torre GL. The possible protective effect of hepatitis B vaccine against lymphomas: a systematic review. *Curr Pharm Biotechnol*. (2022). doi: 10.2174/1389201023666220113111946. [Epub ahead of print].
39. Marcucci F, Spada E, Mele A, Caserta CA, Pulsoni A. The association of hepatitis B virus infection with B-cell non-Hodgkin lymphoma - a review. *Am J Blood Res*. (2012) 2:18–28.
40. Phattraprayoon N, Kakheaw J, Soonklang K, Cheirsilpa K, Ungtrakul T, Auewarakul C, et al. Duration of hepatitis B vaccine-induced protection among medical students and healthcare workers following primary vaccination in infancy and rate of immunity decline. *Vaccines*. (2022) 10:267. doi: 10.3390/vaccines10020267
41. Ben Hadj M, Bouguerra H, Saffar F, Chelly S, Hechaichi A, Talmoudi K, et al. Observational study of vaccine effectiveness 20 years after the introduction of universal hepatitis B vaccination in Tunisia. *Vaccine*. (2018) 36:5858–64. doi: 10.1016/j.vaccine.2018.08.038
42. Stiasny K, Malafa S, Aberle SW, Medits I, Tsouchnikas G, Aberle JH, et al. Different cross-reactivities of IgM responses in dengue, zika and tick-borne encephalitis virus infections. *Viruses*. (2021) 13:596. doi: 10.3390/v13040596
43. Sousa M de AC, Parana R, Andrade LJ de O. SEQUENCE SIMILARITY BETWEEN THYROID SELF-PROTEIN AND HEPATITIS C VIRUS POLYPROTEIN: possible triggering mechanism of autoimmune thyroiditis. *Arq Gastroenterol*. (2016) 53:185–91. doi: 10.1590/S0004-28032016000300012
44. Antonelli A, Ferrari SM, Di Domenicantonio A, Ferrannini E, Fallahi P. Viral Infections and Type 1 Diabetes. In: *Infection and Autoimmunity [Internet]*. Elsevier (2015). p. 877–89. Available online at: https://linkinghub.elsevier.com/retrieve/pii/B9780444632692000477
45. Garcia-Trejo JJ, Ortega R, Zarco-Zavala M. Putative repurposing of lamivudine, a nucleoside/nucleotide analogue and antiretroviral to improve the outcome of cancer and COVID-19 patients. *Front Oncol*. (2021) 11:664794. doi: 10.3389/fonc.2021.664794
46. Kandeel M, Al-Nazawi M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci*. (2020) 251:117627. doi: 10.1016/j.lfs.2020.117627

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fmed.2022.902611

Immunoinformatic Approach to Contrive a Next Generation Multi-Epitope Vaccine Against *Achromobacter xylosoxidans* Infections

*Kashaf Khalid[1], Umar Saeed[2], Mohammad Aljuaid[3], Mohammad Ishtiaq Ali[4], Awais Anjum[5] and Yasir Waheed[6,7]**

[1] Multidisciplinary Laboratory, Foundation University Islamabad, Islamabad, Pakistan, [2] Biological Production Division, National Institute of Health, Islamabad, Pakistan, [3] Department of Health Administration, College of Business Administration, King Saud University, Riyadh, Saudi Arabia, [4] Department of Microbiology, Quaid-i-Azam University, Islamabad, Pakistan, [5] PerkinElmer Inc., Newport, United Kingdom, [6] Clinical and Biomedical Research Center, Foundation University Islamabad, Islamabad, Pakistan, [7] Office of Research, Innovation and Commercialization (ORIC), Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, Pakistan

OPEN ACCESS

Edited by:
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

Reviewed by:
Bilal Shaker,
Chung-Ang University, South Korea
Mohammad M. Pourseif,
Tabriz University of Medical Sciences, Iran
Ricardo Martins Ramos,
Science and Technology of Piauí, Brazil

***Correspondence:**
Yasir Waheed
yasir_waheed_199@hotmail.com

Specialty section:
This article was submitted to Infectious Diseases—Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: 23 March 2022
Accepted: 21 June 2022
Published: 11 July 2022

Citation:
Khalid K, Saeed U, Aljuaid M, Ali MI, Anjum A and Waheed Y (2022) Immunoinformatic Approach to Contrive a Next Generation Multi-Epitope Vaccine Against Achromobacter xylosoxidans Infections. Front. Med. 9:902611. doi: 10.3389/fmed.2022.902611

Achromobacter xylosoxidans, previously identified as *Alcaligenes xylosoxidans*, is a rod-shaped, flagellated, non-fermenting Gram-negative bacterium that has the ability to cause diverse infections in humans. As a part of its intrinsic resistance to different antibiotics, *Achromobacter* spp. is also increasingly becoming resistant to Carbapenems. Lack of knowledge regarding the pathogen's clinical features has led to limited efforts to develop countermeasures against infection. The current study utilized an immunoinformatic method to map antigenic epitopes (Helper T cells, B-cell and Cytotoxic-T cells) to design a vaccine construct. We found that 20 different epitopes contribute significantly to immune response instigation that was further supported by physicochemical analysis and experimental viability. The safety profile of our vaccine was tested for antigenicity, allergenicity, and toxicity against all the identified epitopes before they were used as vaccine candidates. The disulfide engineering was carried out in an area of high mobility to increase the stability of vaccine proteins. In order to determine if the constructed vaccine is compatible with toll-like receptor, the binding affinity of vaccine was investigated *via* molecular docking approach. With the *in silico* expression in host cells and subsequent immune simulations, we were able to detect the induction of both arms of the immune response, i.e., humoral response and cytokine induced response. To demonstrate its safety and efficacy, further experimental research is necessary.

Keywords: *Achromobacter xylosoxidans*, gram-negative bacteria, Immunoinformatic approaches, multi-epitope vaccine, docking

INTRODUCTION

Achromobacter xylosoxidans is a non-fermenter, Gram-negative bacterium that belongs to the *Alcaligenaceae* family. Since the past few years, it has gained notoriety as a pathogen responsible for nosocomially acquired infections among immunocompromised as well as immunocompetent people (1, 2). Researchers Ybuuchi and Ohyama first isolated it in 1971 after observing chronic otitis media patients' ear discharge (3). The opportunistic pathogen is associated with a plethora

of lethal infections such as septicemia (4), pneumonia (2), urinary tract infection (5), meningitis, peritonitis and other infections (6–8). Recently, it has been discovered that the bacterium plays a prominent role in causing cystic fibrosis worldwide (9).

A. xylosoxidans is an opportunistic motile bacterium with peritrichous flagella and often confused with *Pseudomonas* (10). The pathogen is extensively found in water, soil and hospital environments (11) and has proven to be extensively resistant to antibiotics as they harbor excellently characterized resistance mechanisms and at least 50 intrinsic resistance genes (12). In addition to intrinsic resistance, acquired resistance is widely being reported across several regions of the world such as the United States (13), Europe (14), China (15) along with many other regions. In spite of that, the optimal treatment for *Achromobacter spp.* is still unknown. In the face of increasingly prevalent pathogenic bacterium, limited attention is being put toward its treatment, posing an insurmountable clinical challenge and a state of worldwide emergency.

In the current post-genomic era, emphasis has majorly been shifted from antibiotic-based approaches to non-antibiotic-based approaches for the treatment of pathogen induced infections. Particularly, multi epitope vaccine constructs (MEVC) have been shown to be a credible alternative to the obsolete antibiotic-based treatment regimens (16). A multitude of attempts have been made to design potent, safe, and immunostimulant vaccines that are superior to traditional vaccines using AI and immunoinformatic tools (17). In spite of the plethora of available information and the computational resources, no attempt has been made yet to develop a vaccine against this looming threat. Therefore, we made an effort to build a rational multi-epitope next generation therapeutic vaccine against the MDR pathogen by using several *in silico* approaches.

To accomplish this purpose, we used the reverse vaccinology (RV) technique to perform genome mining to filter the proteins that are most immunogenic, as well as non-allergenic and safe to use. A successful working example of the RV approach is the vaccine design against Meningococcus B pathogen where genomic based approaches were used to predict ninety novel antigens (18). This invention led to initiation of plethora of studies on development of vaccines against several dominant and emerging pathogens such as *Klebsiella pneumoniae* (19), *Acinetobacter baumannii* (20), *Helicobacter pylori* (21), *Yersinia pestis* (22) and many others.

MATERIALS AND METHODS

Retrieval of Complete Proteome

As a first step, GenBank Database was employed to retrieve the complete proteome of *A. xylosoxidans* comprising 5,729 proteins (strain GAD3: Accession no. GCF_003031105.1). **Figure 1** presents the study's step-by-step process.

Determination of CTL Epitopes

Using NetCTL 1.2, we identified CTL epitopes in the polyprotein sequence (23). Three elements are combined in this prediction: first, the binding peptide for MHC-I is predicted, then the C-terminal proteasome cleavage is predicted, and it is finally predicted that the Transporter Associated with Antigen Processing (TAP) program is performed. Artificial neural networks were used to estimate the first two parameters, whereas weight matrices were used to determine the TAP transporter's efficiency. In the prediction of CTL epitopes, a cut-off value of 0.75 was permitted. We used the PSORTb to determine the localization of the proteins from which all the epitopes were derived (24).

Determination of HTL Epitopes

With the help of IEDB, 15-mer amino acid length HTL epitopes with excellent affinity were predicted for a reference panel of seven alleles. Our predicted peptides were sorted based on their IC_{50} scores; good binders had an IC_{50} of < 50 nM, intermediate binders had < 500 nM, and low affinity binders had $< 5{,}000$ nM. A lower percentile rank represents a higher binding affinity, as revealed by the inverse correlation between percentile ranking and binding affinity (25).

Study of Population Coverage

Analyses of the world population coverage was conducted with the IEDB server (25). We used the selected epitopes and compared them to the respective allele sets in a wide range of global populations. In the coverage analyses, it was primarily measured whether the selected epitopes were able to cover large populations. Analysis was done on the United States (13), European countries (14), and China (15), which are being hit hardest by the outbreak of *Achromobacter.*

Determination of Continuous B-Cell Epitope Regions

B-cells produce antibodies that serve as a long-term source of immunity. Therefore, to stimulate the protective immune response of the host, the continuous B-cell epitopes were anticipated in the bacterial proteome using an online webserver, BCPred (26). To predict epitopes, BCPred employs a support vector machine (SVM) algorithm and a successive kernel approach. A threshold of 0.8 was established initially as a cut-off point to eliminate the top predicted B-cell epitopes.

Contriving the Immunogenic Construct

Having prioritized carefully determined epitopes from previous steps, we were able to design a multi epitope vaccine construct. Linkers AAY, GPGPG and KK were, respectively, used between CTL, HTL, and B cell epitopes to accomplish this (27). Linking sequences are essential for extending the conformation of proteins (flexibility), allowing foldability, and separating functional domains, which ultimately makes the overall structure of protein more stable and therefore facilitates the protein expression (28). The complex was adjuvanated by connecting an immune potentiating protein (50S ribosomal protein L7/L12, Accession: P9WHE3) at the amine terminus so as to trigger an adequate amount of immune response against the pathogen (29).

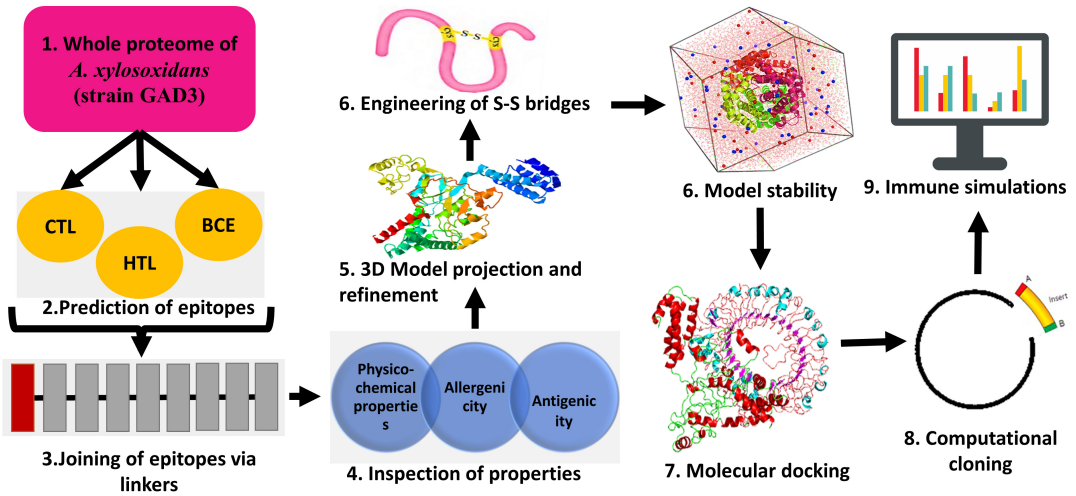


FIGURE 1 | A general workflow employed during the vaccinomics study throughout.

Determination of Allergenicity and Antigenic Potential of the Proposed Construct

The vaccine sequence's allergenicity was assessed using AlgPred, an established allergy prediction server (30). According to the server, allergenic sequences are predicted to an accuracy of 85%. Identifying allergens can be done based on their scores (> 0.4). The vaccine construct must have antigenicity to trigger the proper immunogenic response. To further confirm the non-allergenicity of the proposed vaccine, AllerTop v2 was employed (31). We employed the VaxiJen server to determine the construct's antigenicity, retaining the default threshold value of 0.45 (32). By converting protein sequences into a vector of key amino acid properties using automatic cross correlation, VaxiJen predicts antigens without alignment. Multiple biological functions depend on protein solubility and stability. Therefore, vaccine's solubility was determined using two web servers SOLpro (33) and Protein-Sol (34). In SOLpro, solubility is predicted using the probability scores. Thus, soluble proteins receive scores of ≥ 0.5 and insoluble proteins receive scores of < 0.5. In contrast, Protein-Sol predicts the vaccine construct's solubility based on solubility data.

Physicochemical Profiling of the Proposed Construct

On the ProtParam website, the vaccine's physicochemical properties were revealed which involve theoretical pI, molecular weight, construct's half-life, instability index, and GRAVY (35). 2-D assembly of the designed construct was calculated using PSIPRED (36).

3D Modeling, Refinement and Quality-Check

For 3D model prediction, I-TASSER was harnessed which is a web tool for computer-assisted function and structure determination of proteins using sequence-to-structure-to-function analysis, with the PDB used to identify similar structure patterns (37). The initial 3D atomic models produced by I-TASSER originate from numerous threading alignments and reiterative simulations of structure beginning with a protein sequence. Models with TM-values greater than 0.5 generally express precise topology, while those with TM-values lesser than 0.17 show random similarity. A protein's length has no bearing on the cutoff value (38). The community-wide CASP analysis of vaccine's 3D structure evaluation using I-TASSER projection, refinement, and validation has consistently listed it as the best server for protein 3D structure determination (39).

The projected 3D structure was subjected to further refinement to improve local and global structure quality using ModRefiner (40) and Galaxy Refine (41) web servers. CASP10 refinement was (42) used to reconstruct and repack the side chain of the protein, which was then relaxed through MD simulation. Thereafter, a model validation step was performed to pinpoint probable limitations in the refined structure (39). Initially, ProSA-web server was employed to gauge the overall quality in terms of Z-value (43). The Z-scores are erroneous if they do not fall within the range of properties for pre-determined protein structures. In order to investigate non-bonding atoms–atom interactions, the ERRAT web server was used (44). By displaying the percentages of residues around favorable and unfavorable regions on the Ramachandran plot, the overall quality of the modeled structure was described using RAMPAGE's web server (45). The loops were optimized using ModLoop, a server that analyzes the proteins' angles so that the loops have proper φ and ψ angles (46).

Stabilization of Vaccine Through Loop Remodeling and Disulfide Engineering

Providing more stability to the refined protein model was required before proceeding with the next step. Loop re-modeling in the 3D structure of vaccine was carried out by using

Galaxy Loop server (47). Furthermore, disulfide bonds were introduced into the protein model *via* an online server-Disulfide by Design 2.0 (48). Disulfide bond is a covalent interaction that mimics stable molecular interactions and demonstrates the precise geometric conformation of proteins and therefore contributes to their stability. Using disulfide engineering, one can introduce disulfide bonds into a protein structure. Therefore, the initial model of the refined protein was uploaded and used for the residue pair search necessary to engineer disulfides. Total 4 residue pairs were selected to mutate them with cysteine residue using create mutate function of the Disulfide by Design 2.0 server.

TABLE 1 | Predicted 9-mer peptides as the potential cytotoxic-T cell epitopes.

Sr. no	Residue ID.	Peptide Sequence (9 mer)	MHC Binding Affinity	Rescale Binding Affinity	C-Terminal Cleavage Affinity	TAP score	combined Score	MHC-I Binding	Antigenicity Score	ToxinPred	AllerTop
1	79	PTDKTVDAL	0.1763	0.7484	0.8108	0.1380	0.8770	✓	Non-antigen	-	-
2	13	STAGTGHFY	0.7048	2.9925	0.8299	2.8060	3.2573	✓	0.99	Non-toxin	Allergen
3	41	PVARKHVDY	0.1391	0.5907	0.9643	2.4610	0.8584	✓	Non-antigen	-	-
4	29	LSEQGFNVF	0.2172	0.9221	0.8489	2.3830	1.1686	✓	Non-antigen	-	-
5	99	GLEMGADDY	0.2783	1.1815	0.7952	0.7952	2.6680	✓	1.06	Non-toxin	Allergen
6	228	IQTVWGLGY	0.1721	0.7309	0.9704	2.8802	1.0206	✓	0.46	Non-toxin	Allergen
7	**05**	**MCLICGWVY (O)**	**0.1306**	**0.5543**	**0.5592**	**2.9540**	**0.7859**	✓	**1.04**	**Non-toxin**	**Non-allergen**
8	64	TTVQQAIDY	0.4884	2.0736	0.8348	2.9690	2.3473	✓	Non-antigen	-	-
9	**87**	**VKDLPGVRY (O)**	**0.1157**	**0.4912**	**0.9662**	**2.9920**	**0.7857**	✓	**1.18**	**Non-toxin**	**Non-allergen**
10	**133**	**PTTTVHGVF (O)**	**0.1261**	**0.5355**	**0.9066**	**2.1510**	**0.7790**	✓	**0.59**	**Non-toxin**	**Non-allergen**
11	**141**	**FLDVLGLGV (O)**	**0.2209**	**0.9381**	**0.8986**	**0.1100**	**1.0784**	✓	**0.88**	**Non-toxin**	**Non-allergen**
12	09	FSSDMAIDL	0.1743	0.77402	0.5425	0.8440	0.8638	✓	0.68	Non-toxin	Allergen
13	156	VSDASGSMV	0.4340	1.8425	0.1304	0.3950	1.8818	✓	0.86	Non-toxin	Allergen
14	**171**	**TTEVAVISL (P)**	**0.1609**	**0.6830**	**0.6335**	**0.7380**	**038149**	✓	**0.88**	**Non-toxin**	**Non-allergen**
15	176	VISLGGMVY	0.1991	0.8455	0.9705	3.2650	1.1543	✓	Non-antigen	-	-
16	261	LTDPLNQIV	0.5305	2.2523	0.7255	0.1750	2.3699	✓	Non-antigen	-	-
17	44	ILLSSMTGY	0.1260	0.5350	0.9524	3.1170	0.8337	✓	Non-antigen	-	-
18	58	QMTGVVHEY	0.3152	1.3382	0.9744	3.0700	1.6379	✓	Non-antigen	-	-
19	131	LTDAGKLEM	0.3861	1.6394	0.8950	0.2520	1.7862	✓	Non-antigen	-	-
20	192	RTDLDKLVL	0.2002	0.8500	0.9140	0.9520	1.0347	✓	Non-antigen	-	-
21	**268**	**CLKAENIYY (P)**	**0.2354**	**0.9993**	**0.8162**	**3.0420**	**1.2738**	✓	**1.23**	**Non-toxin**	**Non-allergen**
22	20	VSDGIAHVH	0.2379	1.0101	0.3047	0.4680	1.0324	✓	Non-antigen	-	-
23	**38**	**ITDRQGNAL (O)**	**0.2231**	**0.9472**	**0.9165**	**0.7900**	**1.1242**	✓	**0.83**	**Non-toxin**	**Non-allergen**
24	**73**	**TAGRVALEY (P)**	**0.2929**	**1.2435**	**0.9378**	**2.6260**	**1.5155**	✓	**0.81**	**Non-toxin**	**Non-allergen**
25	03	MSDPIADML	0.3963	1.6826	0.6579	0.9200	1.8273	✓	Non-antigen	-	-
26	**121**	**GVGGEVLCY (P)**	**0.1417**	**0.6015**	**0.9548**	**2.8620**	**0.8878**	✓	**1.34**	**Non-toxin**	**Non-allergen**
27	01	MSETQNTQV	0.3647	1.5484	0.9547	0.2560	1.7044	✓	1.42	Non-toxin	Allergen
28	47	HVDTGDYIV	0.2039	0.8658	0.6309	0.0410	0.9624	✓	Non-antigen	-	-
29	108	MLPKGPLGY	0.1866	0.7923	0.9546	2.9460	1.0828	✓	Non-antigen	-	-
30	181	GTDGHAHIV	0.3340	1.4182	0.9094	0.2290	1.5432	✓	1.74	Non-toxin	Allergen
31	**252**	**DAEEIKLRY (P)**	**0.3619**	**1.5366**	**0.9363**	**2.3720**	**1.7957**	✓	**2.44**	**Non-toxin**	**Non-allergen**
32	**100**	**GSCANGGGY (P)**	**0.3658**	**1.5533**	**0.3409**	**2.8310**	**1.7459**	✓	**2.48**	**Non-toxin**	**Non-allergen**
33	13	FTQQNPLFK	0.1771	0.7517	0.22313	0.4170	0.8058	✓	0.41	-	-
34	306	NAESVSSLF	0.1621	0.6883	0.5846	2.5340	0.9026	✓	Non-antigen	-	-
35	412	MTGSQPMLF	0.2936	1.2466	0.3933	2.2300	1.4171	✓	0.78	Non-toxin	Allergen

Bracketed letters describe the location of the protein from which the epitope was obtained. The letter P indicates periplasmic origin, and O indicates outer membrane origin. The selection criteria included (i) TAP values, (ii) C-terminal cleavage, and (iii) antigenicity values. Bold lettering refers to CTLs that have been finalized for further review.

TABLE 2 | Predicted 15-mer peptides as the potential helper T lymphocyte epitopes.

Sr. no	Position (Start-end)	Sequence (15-mer)	HLA allele	Percentile rank	Antigenic values	ToxinPred	Allertop
1	**24–38**	**AQFGLRMQKATQQLA (P)**	**HLA-DRB4*01:01**	**0.16**	**0.97**	**Non-toxin**	**Non-allergen**
2	**145–159**	**AFGPYVLNLSTRTLT (P)**	**HLA-DRB3*02:02**	**0.01**	**0.43**	**Non-toxin**	**Non-allergen**
3	11–25	KILVVDDDPRLRDLL	HLA-DRB1*03:01	0.03	Non-antigen	-	-
4	175–189	ADDAVEFSRTAPNMI	HLA-DRB1*07:01	0.03	Non-antigen	-	-
5	117–131	AQLIDLLRIYLGKKL	HLA-DRB1*15:01	0.08	Non-antigen	-	-
6	199–213	IVNYIRRNYGMLIGE	HLA-DRB3*02:02	0.22	Non-antigen	-	-
7	35–49	SSRVLVDTPEVRGMI	HLA-DRB1*03:01	0.16	Non-antigen	-	-
8	4–18	REGYRPNVGIILVNS	HLA-DRB3*02:02	0.05	0.45	Non-toxin	Allergen
9	393–407	GNFKSLLARMKEWFM	HLA-DRB5*01:01	0.10	Non-antigen	-	-
10	**54–68**	**KRHYKRIRSQQLPP (O)**	**HLA-DRB4*01:01**	**0.07**	**0.92**	**Non-toxin**	**Non-allergen**
11	71–85	EQLIARNAPLPIVFI	HLA-DRB3*02:02	0.11	0.72	Non-toxin	Allergen
12	54–68	RGLLRMVSRRRKLLD	HLA-DRB5*01:01	0.02	Non-antigen	-	-
12	98–112	RARIVAQAAQDAAQE	HLA-DRB4*01:01	0.38	Non-antigen	-	-
14	76–90	AVYGIRGILARGKFD	HLA-DRB5*01:01	0.06	Non-antigen	-	-
15	251–265	LQYQKHLRLQEARRL	HLA-DRB5*01:01	0.41	0.57	Non-toxin	Allergen
16	**114–128**	**KKSLARLQLDHIDLY (P)**	**HLA-DRB4*01:01**	**0.39**	**0.61**	**Non-toxin**	**Non-allergen**

Bracketed letters describe the location of the protein from which the epitope was obtained. The letter P indicates periplasmic origin, and O indicates outer membrane origin. The selection criteria included (i) low percentile rank, (ii) antigenicity, (iii) allergenicity, (iv) toxicity. Bold lettering refers to HTLs that have been finalized for further review.

Non-linear B-Cell Epitope Determination

Nearly 90% of the B-cell epitopes, owing to their spatial arrangements, are discontinuous and found close together (49). Due to conformational B-cell epitopes playing a significant part in eliciting an antibody-mediated immune response, it is imperative to determine their presence in the developed vaccine construct (42). This was accomplished by using the ElliPro server (50), which determines epitopes on the basis of PI values, where PI equal to 0.9 includes 90% of residues in an ellipsoid, and excludes 10% from it. The epitopes with the top PI values were selected.

TABLE 3 | Predicted 20-mer peptides with values greater than the threshold value (0.9) as the potential B cell epitopes.

Sr. no	Sequence (20-mer)	Start position	Threshold	ToxinPred	AllerTop
1	NLQEHSVVLVR GGRVKDLPG	114	0.93	Non-toxin	Allergen
2	**ADITDKGIALT GGGALLRDL (O)**	**285**	**0.95**	**Non-toxin**	**Non-allergen**
3	VIADFTVTEQM LKQFIRMVH	81	0.91	Non-toxin	Allergen
4	ERGRASWAPYPD ITPLPPEH	121	0.91	Non-toxin	Allergen
5	**PLPPEHRTLEQ AWSIFRFGP (P)**	**135**	**0.96**	**Non-toxin**	**Non-allergen**
6	YGDLLRHFGRS IVIAHSEGA	195	0.92	Non-toxin	Allergen
7	**NADVQLDL LLDYASNVQKYP (O)**	**205**	**0.98**	**Non-toxin**	**Non-allergen**
8	**TGGSLPGTGSG GGGGPGGAG (O)**	**36**	**0.94**	**Non-toxin**	**Non-allergen**
9	**IDNGFDADPAT DHHKLSVAG (P)**	**156**	**0.97**	**Non-toxin**	**Non-allergen**
10	PADPVTPPDPAK PADPATPA	87	0.92	Non-toxin	Allergen

Bold lettering refers to HTLs that have been finalized for further review.

Folding Stability Analysis *via* MD Simulation Technique

For proper regulation of cellular protein networks, which perform various biological functions, such as cell-cell communication and activation of immune response and so on, appropriate folding of proteins is necessary (27). This study utilized all-atom molecular dynamics simulations approach to determine how the vaccine constructs fold and function. Using a freeware-GROMACS (version 5.0), we set up a simulation system that included water and ions, as well as the designed vaccine construct (51). Initially, pdb2gmx module of GROMACS was harnessed to generate a force-filed compliant topology of the designed construct. The system was then applied the force field with OPLS-AA, after which the editconf module was used to create a rhombic dodecahedron box that later on was solvated with water molecules. The solvated system after going through electro-neutralization stage, was checked for steric clashes and inappropriate geometry *via* the energy minimization (EM) technique. Following relaxation of the structure with EM, ions and solvents were equilibrated under NVT and NPT ensembles for 100 and 1,000 ps, respectively. The equilibrated system was then subjected to dynamic simulations for 50 ns. By examining

the MD trajectories, we were able to compute data such as RMSD, RMSF and RG (52).

Molecular Docking With Immune Receptors

It is imperative that the vaccine interact with the target immune cell receptors in order to cause a stable immunogenic response. Therefore, we conducted a molecular docking analysis to decipher the interaction among the designed construct as well as the human toll like receptor. Specifically, TLR4 was studied as it is actively involved in invoking the immune response against the Gram negative bacteria (53). To fetch the pdb format of TLR4 from the Protein Data Bank, ID: 4G8A was used. The docking process was driven using the ZDOCK server (54). Using rigid-body docking programs, the ZDOCK web server generates quick and accurate complexes. We used the option of selecting contacting/blocking residues and selected the active residues (I48, D50, F54, Y72, S73, F75, S76, S100) from the B chain of the TLR 4 (55). Furthermore, the complexes generated by ZDOCK were sent for refinement and post processing *via* the FireDock server (56). The FireDock server allows high throughput refinement of docked complexes *via* side chain optimization. As a final step, using PDBsum, interactions between the vaccine and the host TLRs were mapped (57). In order to measure the epitope binding efficiency, the crystallographic ligand KDO was used as a positive control.

In silico Cloning Experiment

Optimizing codons in the host cell ensures maximum expression. We therefore optimized the sequence using JCat tool in accordance with our preferred expression organism, *Escherichia coli K12* (58). In addition to GC content, the result provides a codon adaptation index (CAI) score. The score reflects the level of favorability of protein expression. SnapGene software was used for *in silico* cloning using pET28a (+) vector backbone.

Immune Simulations

Foreign particles and the immune system can be studied by simulating their interactions. For this purpose, C-ImmSim was harnessed (59). The reactivity of the host immune system to foreign particles is studied by an agent-based modeling approach. Researchers are exploring immune responses in realistic virtual environments due to the recent interest in computational learning and development. With the PSSM model, the server measures the immune system's reaction to an antigen. Estimates are provided for antibodies, interferon, cytokines, and other immune substances produced by vaccination. During the entire simulation, 1400-time steps were run, which is approximately 15 months (each time step lasts 8 h).

RESULTS

Data Retrieval

From the NCBI database, we downloaded the completely sequenced proteome of *A. xylosoxidans* and examined T cell epitopes and B cell epitopes so as to stimulate the respective arms of the immune system in order to design a potent multi-epitope vaccine against the pathogen.

Proteome Mining for CTLs

Modern vaccines consist of a combination of epitopes, out of which CTLs are important in vaccine's efficient stimulation of the immune response. A thorough analysis of complete proteome of *A. xylosoxidans* revealed a total of 35 potential CTL epitopes as 9-mer peptides (**Table 1**). These peptides were filtered on the bases of good binding affinity. Ultimately, only eleven epitopes showed the greatest antigenic potential and combined score and were found to be localized in the outer membrane and periplasmic regions which made them perfect candidates to be incorporated into the final vaccine construct.

Proteome Mining for HTLs

For predicting HTL lymphocytes with high binding efficiencies, a seven-allele HLA reference set was chosen. When selecting epitopes for the final construct of the multi-epitope vaccine, only those with the lowest percentile ranks were selected. The final selected HTL epitopes as well as the location of the proteins used to extract these epitopes are listed in **Table 2**.

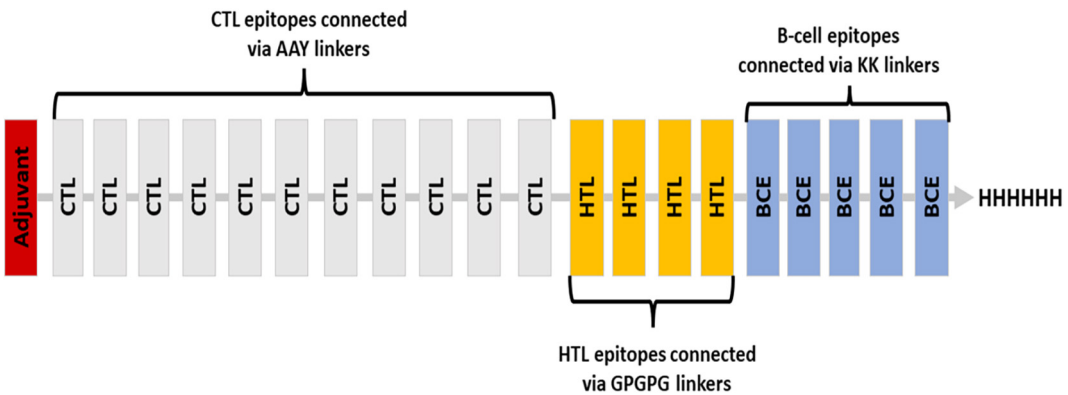


FIGURE 2 | Blueprint of the proposed vaccine ensemble with different colors representing different components.

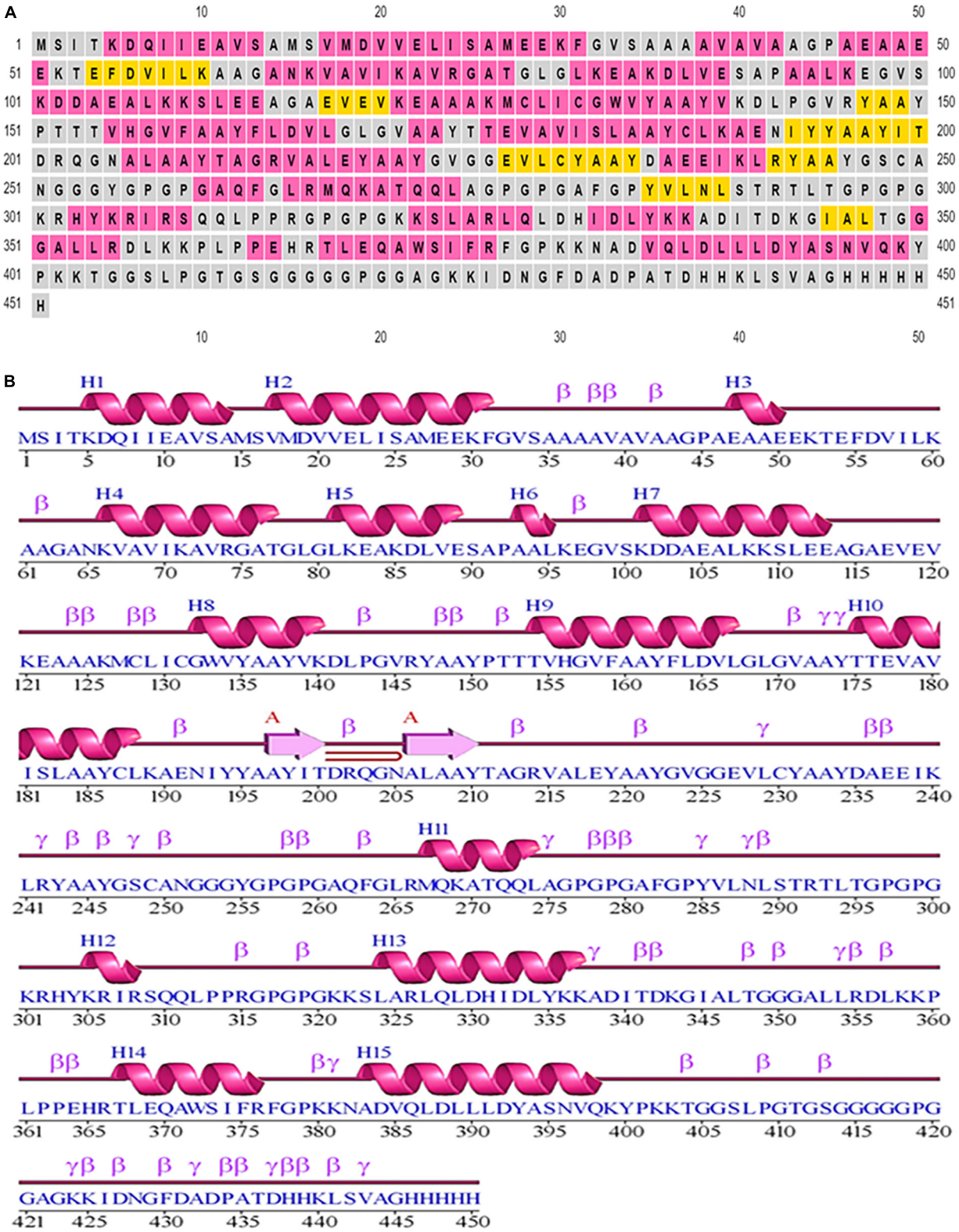


FIGURE 3 | Secondary structural elements of the designed construct. **(A)** Annotation grid. **(B)** Illustration of the structure with each region denoted by its own symbol.

Study of Population Coverage

The epitopes and alleles associated with selected MHC class I and MHC class II epitopes were analyzed in various regions of the world (**Supplementary Table 1**). We found that the selected MHC class I and class II epitopes are shared by 57.55% and 98.9% of the world's population, respectively. Among the countries with the highest coverage of class I MHC epitopes, Italy (88.28%) and China (81.39%) ranked first and second, respectively (**Supplementary Table 2**). As for the MHC II, the United States had the highest coverage of MHC class II epitopes (99.1%) (**Supplementary Table 3**).

Proteome Mining for Linear BCEs

In the humoral immune response, B lymphocytes play a major role. To further incorporate B-cell epitopes into the final construct, we chose those with the highest scores. As a whole, scores greater than 0.9 suggested better potential for incorporation into vaccine formulations. A shortlist of ten optimal epitopes and location of their proteins can be found in **Table 3**.

Linking of Epitopes to Contrive Multi-Epitope Vaccine

Various linkers were used to merge the 11 CTL epitopes, 4 HTL epitopes, and 5 B-cell epitopes. CTL epitopes were linked using AAY linkers (AAY aids the epitope in achieving good binding to TAP transporters as well as facilitates epitope presentation); HTL epitopes were linked using GPGPG linkers (GPGPG catalyzes the HTL responses and preserves the conformational immunogenic potential of helper and antibody epitopes). A further connection of epitopes to B-cell epitopes was made using the KK linkers. An adjuvant was placed at the N-terminus *via* EAAK linker, and a 6-his tag was placed at the C-terminus (**Figure 2**). In total, 451 amino acids were incorporated into the final construct.

Inspection of Physicochemical Properties

Molecular weight of the vaccine was found to be 47.1 kDa. The theoretical PI of the vaccine was 8.33 which demonstrated the basic nature of vaccine. A half-life of approximately 10 h and 20 h was observed in *E. coli* and yeast, respectively. The vaccine was classified as stable by its reported instability score of 33.36 (< 40 represents stability). Moreover, the value 86.05 corresponds to thermo stability, which describes the vaccine's ability to withstand a wide range of temperatures. GRAVY assessed the vaccine as a hydrophilic molecule with a GRAVY value of −0.114.

Allergenicity and Antigenicity and Solubility Analysis

The final sequence of proposed vaccine was sent for allergenicity testing using the AlgPred and AllerTop servers. IgE epitope mapping and amino acid composition indicated the non-allergenic nature of vaccine. Furthermore, the antigenic potential exhibited by the designed construct was observed to be greater than the threshold value as measured by the VaxiJen 2.0 webserver with a value of 0.79 without adjuvant and 0.88 with adjuvant. Therefore, regardless of whether an adjuvant was present, the construct was antigenic in nature. As a result

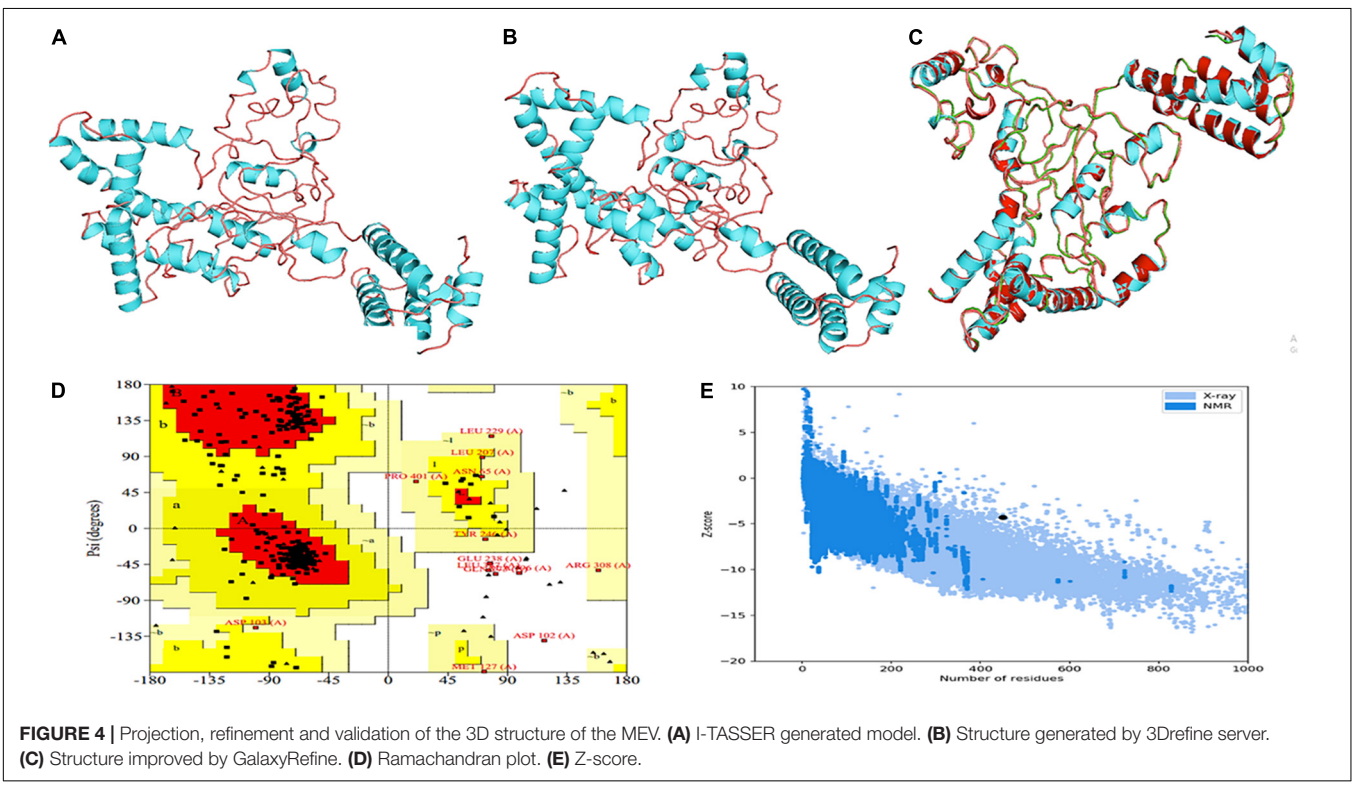


FIGURE 4 | Projection, refinement and validation of the 3D structure of the MEV. **(A)** I-TASSER generated model. **(B)** Structure generated by 3Drefine server. **(C)** Structure improved by GalaxyRefine. **(D)** Ramachandran plot. **(E)** Z-score.

of overexpression in *E. coli*, SOLpro server had a solubility level of 0.949. Using the Protein-Sol webserver, the solubility of the vaccine construct (QuerySol = 0.49) was greater than the population average (PopAvrSol = 0.45), thereby confirming the soluble nature of the proposed ensemble.

Investigation of Secondary Elements in the Designed Construct

Based on PSIPRED, it was determined there were 9.3% strands, 48.11% helical structures, and 42.57% coil structures. **Figure 3** is a diagrammatic depiction of the secondary elements of the designed vaccine.

Tertiary Structure Prediction

With the use of several threading templates (1rquA, 7s0yA, 1rqv, 7louA, 1rqv, 7mexA, 1rqv, 7mexA, 1rquA, and 7eeiA), the I-TASSER forecast the probable 3D model of the designed construct. These 10 templates showed good alignment based on their Z-score values ranging from 1.22 to 3.86. C-score values can range from −5 to 2, with a higher score denoting higher confidence. Among the five predicted models with a C-score between 0.58 and −2.70, model 1 was chosen based on its highest value. TM-scores are used to check for topological similarity between two protein structures. Our vaccine construct had a TM score of 0.58 ± 0.14, whereas RMSD score was 8.6 ± 4.6 Å. TM scores over 0.5 are indicative of accurate topology, while scores below 0.17 indicate non-specific similarity (**Figure 4A**).

Model Refinement and Validation

3Drefine was used to refine the potential chimeric vaccine model. Several parameters were used to shortlist one of five models generated by 3Drefine. Subsequently, the refined model was sent for further refinement by GalaxyWeb, where Model 1 was shortlisted based on poor rotamers score (0.6), clash score (9.8), and Ramachandran plot (90.2%) for further corroboration (**Figures 4B,C**). Based on PROCHECK's Rama plot investigation of the protein model, 86.2 percent of amino acids were found in preferred regions. In addition, 10.5% of the residues were in the allowed regions, while only 1.4% of proteins were in the disallowed boundaries, so the model was of good quality (**Figure 4D**). To refine loops in the disallowed regions, 10

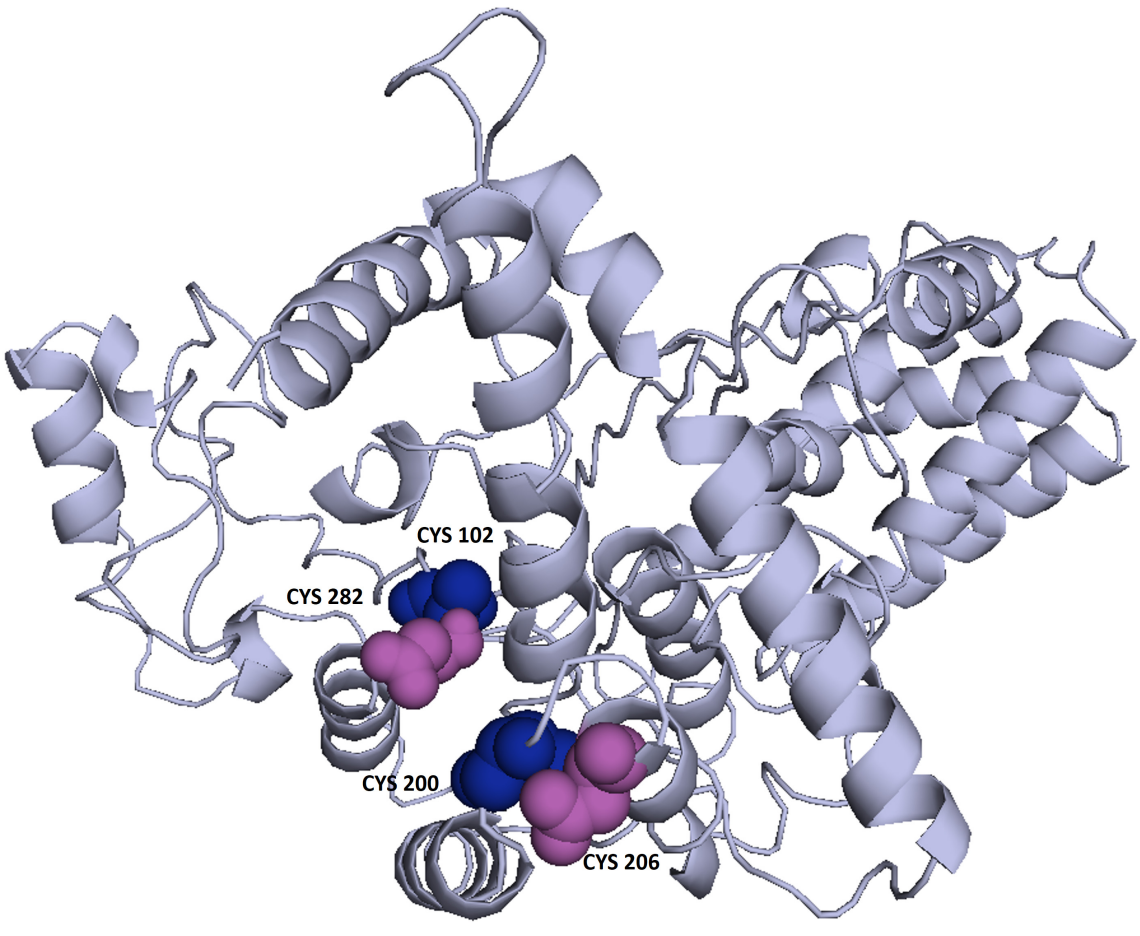


FIGURE 5 | Enhancing protein stability through disulfide engineering. There are two sets of mutated residues displayed in magenta and gray. Residues were chosen according to their energy, chi3 value, and B-factors.

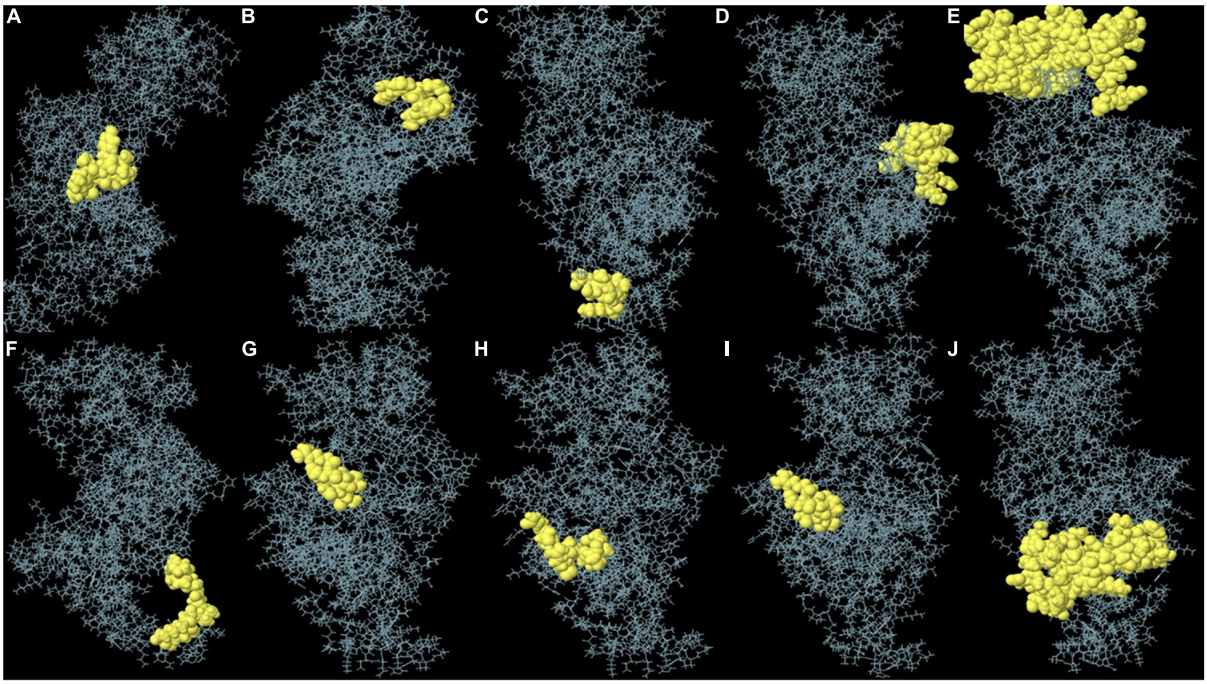


FIGURE 6 | (A–J) Each monomer of the designed vaccine is predicted to contain discontinuous B cell epitopes.

rounds of loop refinement were performed each for Val20-Ala26, Ala84-Glu89, Glu105-Ala114, Glu218-Val224, Cys230-Ala232, Tyr335-Lys337, Asp342-G344, Tyr393-Gln398, Gly423-Ily426 and Pro434-Gly445. Furthermore, we used Disulfide by Design v 2.0 (48) for the S-S engineering of the final construct to increase the steadiness of its modeled structure. It was determined that 32 residue pairs could be utilized for disulfide engineering. Nevertheless, after factors such as Chi3 and energy were evaluated, only four pairs of residues were considered final since their energy values and Chi3 values fell within the permitted range, which is less than 2.5 kCal/mol and between −87° and + 97°, respectively. Hence, a total of four mutations were introduced at residue pairs Asp102-Phe282 and Thr200-Ala206 (**Figure 5**). The ProSA-Web validated the overall quality of the refined model and any potential errors. With a Z-score of −1.26, the refined model appeared to be suitable (**Figure 4E**). Furthermore, the ERRAT web server gave a quality score of 84.32.

Non-linear B-Cell Epitopes Determination

Non-linear B-cell epitopes may be predicted based on the structure and folding of the new protein. By applying the ElliPro web tool, we were able to analyze the refined 3D models of non-linear B-cell epitopes (**Figure 6**). With values ranging from 0.51 to 0.79, ElliPro predicted 10 B-cell epitopes with non-linear characteristics involving 237 residues (**Table 4**).

Stability Check *via* MD Simulation

Prior to determining how the MEVC interacted with the immunoreceptor, the folding stability of MEVC was evaluated. The use of MD simulations to understand protein folding and stability could provide insight into different properties of proteins—their loops, their interactions with other proteins, and the effects of mutations on these interactions. The previously obtained three-dimensional structure was therefore simulated with the MD technique. Using the OPLS-AA as the force field, the vaccine construct's mass was determined to be 47,011.99 amu. To neutralize the net charge on the protein, four Cl ions were added at atoms 65,580, 17,853, 21,750 and 13,863. The water molecules remaining behind were 43,756. Following that, 50,000 energy minimization steps were performed where steepest descents converged to Fmax < 1,000 in 1,455 steps. The potential energy was found to be −2.4462095e + 06 kJ/mol while the average potential energy was found to be −2.38951e + 06 kJ/mol with a total drift of −195,161 kJ/mol. The average temperature after 50,000 steps was 299.757 K with a total drift of 1.11929 K (**Figure 7A**), the average pressure was −0.466784 bar with a total drift of −0.403029 bar (**Figure 7B**), whereas the average density was 1,012.79 kg/m^3 with a total drift of 0.55504 kg/m^3 (**Figure 7C**). Trajectory examination was performed for 50 nanoseconds. Throughout the experiment, the radius of gyration graph indicated that the designed construct was quite stable (**Figure 7D**). In terms of RMSD backbone, there seems to be very few fluctuations, indicating the vaccine's stability and structural integrity over time. The RMSD plot has been depicted in **Figure 7E** which showed that RMSD levels had gone up to ∼1.12 nm. On the other hand, high points in the RMSF plot suggested a high level of flexibility in the vaccine design (**Figure 7F**).

TABLE 4 | Non-linear B-cell epitopes as forecasted by the Ellipro webserver.

	Amino acid residue	Total residues	Score
1	A:A136, A:A137, A:Y138, A:V139, A:K140, A:D141, A:L142, A:P143, A:G144, A:V145	10	0.798
2	A:D392, A:A394, A:S395, A:N396, A:V397, A:Q398, A:K399	7	0.765
3	A:P362, A:Y400, A:P401, A:K402, A:K403, A:T404, A:G405, A:G406, A:S407, A:L408, A:P409, A:G410, A:T411, A:G412, A:S413, A:G414, A:G415, A:G416, A:G417, A:G418, A:P419, A:G420, A:G421, A:A422, A:G423, A:K424, A:D427, A:N428, A:G429, A:F430, A:D431, A:A432, A:D433, A:P434, A:A435, A:T436, A:D437, A:H438, A:H439, A:K440	40	0.747
4	A:M15, A:S16, A:V17, A:M18, A:D19, A:V20, A:V21, A:E22, A:L23, A:I24, A:S25, A:A26, A:M27, A:E28, A:E29, A:K30, A:F31, A:G32, A:V33, A:S34, A:A35, A:A36, A:A37, A:A38, A:V39, A:A40, A:V41, A:A42, A:A43, A:G44, A:P45, A:A46, A:E47, A:A48, A:E50, A:E51, A:L59, A:K60, A:A61, A:A62, A:G63, A:A64, A:N65, A:K66, A:V67, A:A68, A:I70, A:K71, A:A72, A:R74, A:G75, A:T77, A:G78, A:L79, A:G80, A:L81, A:K82, A:E83, A:A84, A:K85, A:D86, A:E89, A:K96, A:G98, A:V99, A:S100, A:K101, A:D102, A:D103, A:A104, A:E105, A:A106, A:L107, A:K108, A:K109, A:S110, A:L111, A:E112, A:E113, A:A114, A:G115, A:A116, A:E117, A:V118, A:E119, A:V120, A:K121	87	0.744
5	A:P319, A:K425, A:I426, A:L441, A:S442, A:V443, A:A444, A:G445, A:H446, A:H447	10	0.72
6	A:T293, A:L294, A:T295, A:G296, A:P297, A:G298, A:P299, A:G300, A:K301, A:R302, A:H303, A:A338, A:D339, A:I340, A:T341, A:D342, A:K343, A:G344, A:I345, A:A346, A:L347, A:T348, A:G349, A:G350, A:G351, A:A352, A:L353, A:L354, A:R355, A:P379, A:K380, A:N382, A:A383, A:D384, A:V385, A:L387, A:D388, A:L389, A:L390, A:L391	40	0.711
7	A:T200, A:D201, A:R202, A:Q203, A:G204, A:N205, A:A206	7	0.609
8	A:G223, A:G225, A:G226, A:E227, A:V228, A:G247, A:S248, A:C249, A:A250, A:N251, A:G252, A:G253, A:G254, A:Y255, A:G256, A:P257, A:G258, A:A261, A:Q262, A:L265, A:R266, A:K269	22	0.591
9	A:D356, A:L357, A:K358, A:K359, A:P360, A:L361	6	0.557
10	A:Q268, A:Q272, A:A275, A:G276, A:P277, A:G278, A:G280, A:R315	8	0.51

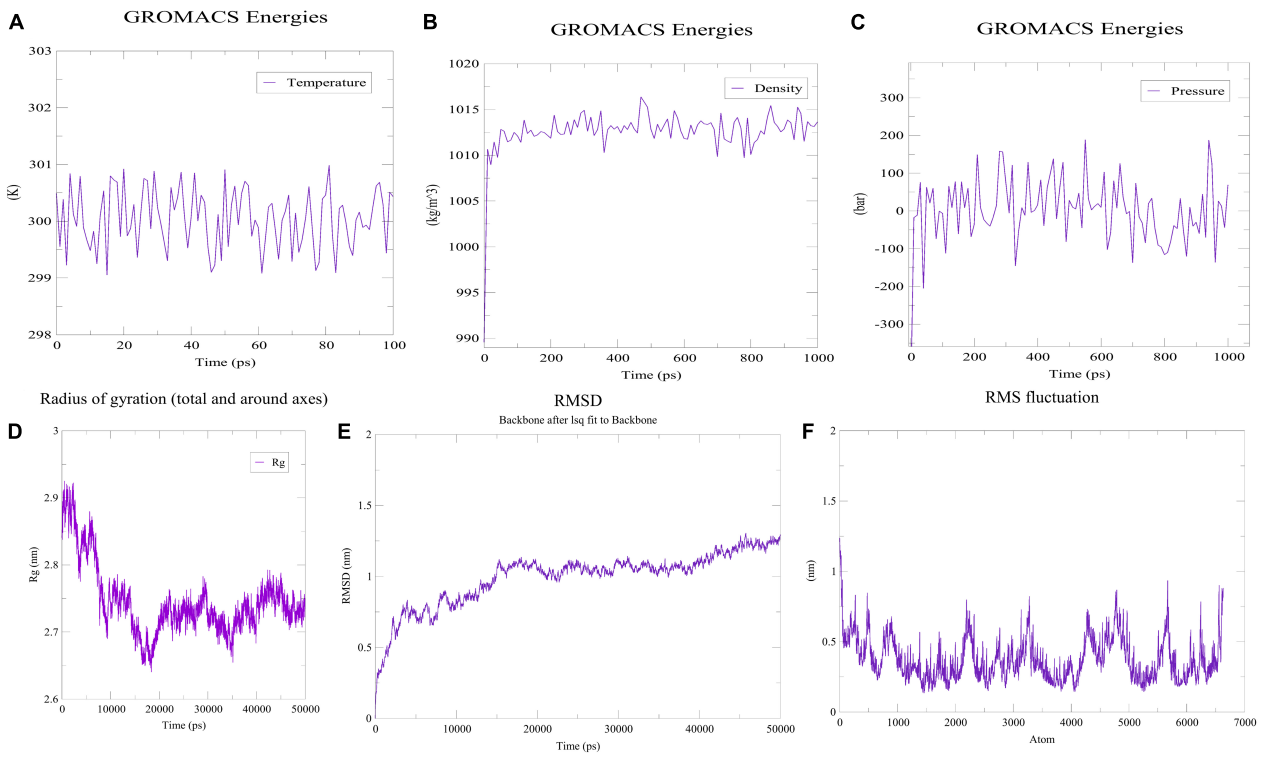


FIGURE 7 | MDS analysis of the designed construct. **(A)** Temperature of the simulated system. **(B)** Pressure of the system. **(C)** Density attained by the system. **(D)** Radius of gyration plot indicating compactness of the construct around its axis. **(E)** RMSD graph depicting stability of the designed construct. **(F)** RMSF plot showing flexibility of the construct.

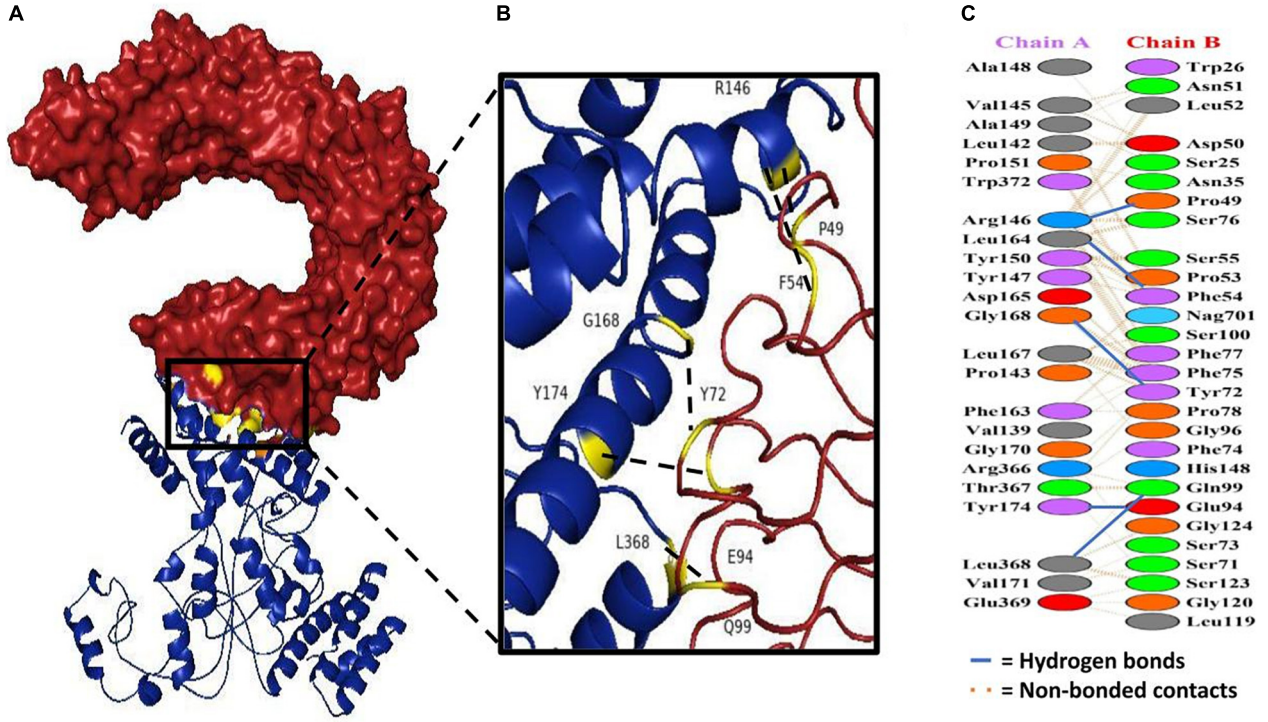


FIGURE 8 | Docking study of subunit vaccine with immune receptors. **(A)** TLR4 chain B can be seen in red, while the vaccine construct can be seen in blue. **(B)** The interacting interface of receptor and designed construct showing all the residues involved in bond formation. **(C)** Protein-protein interaction residues across the interface.

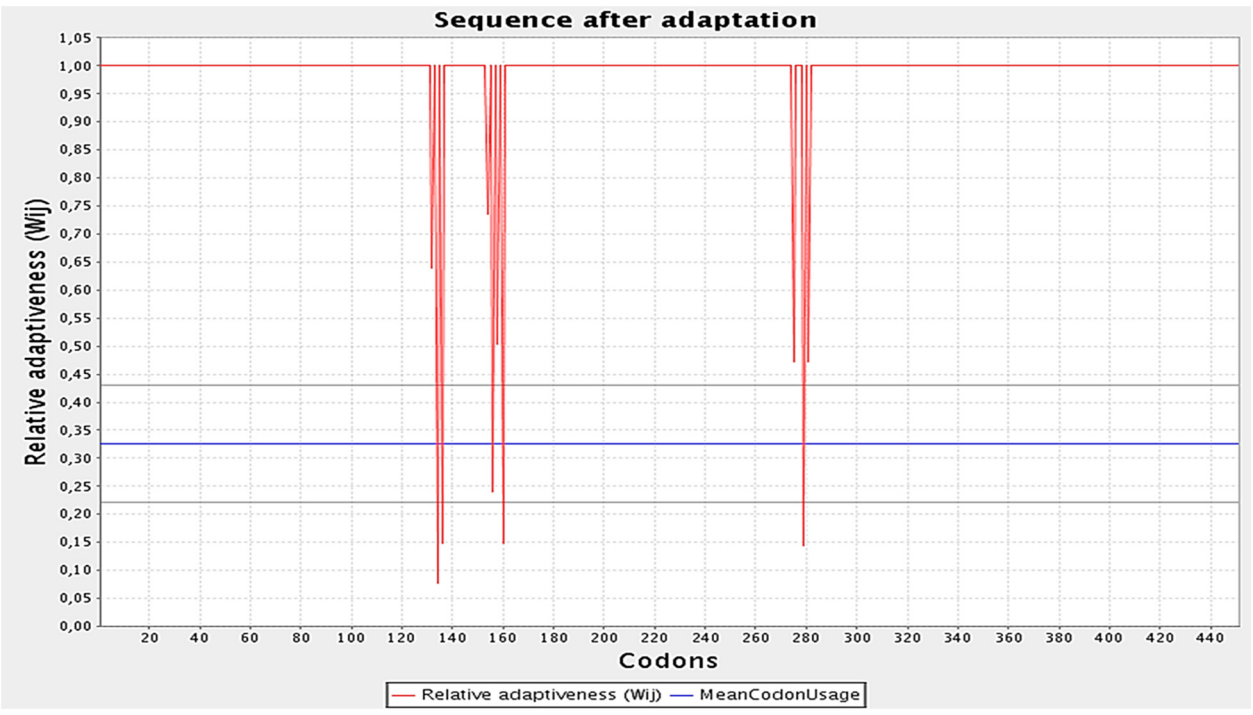


FIGURE 9 | Optimization of the codons of the final design. Graph illustrates the codon adaptation index of the optimized construct to be 0.97. The GC levels were found to be 53.28%.

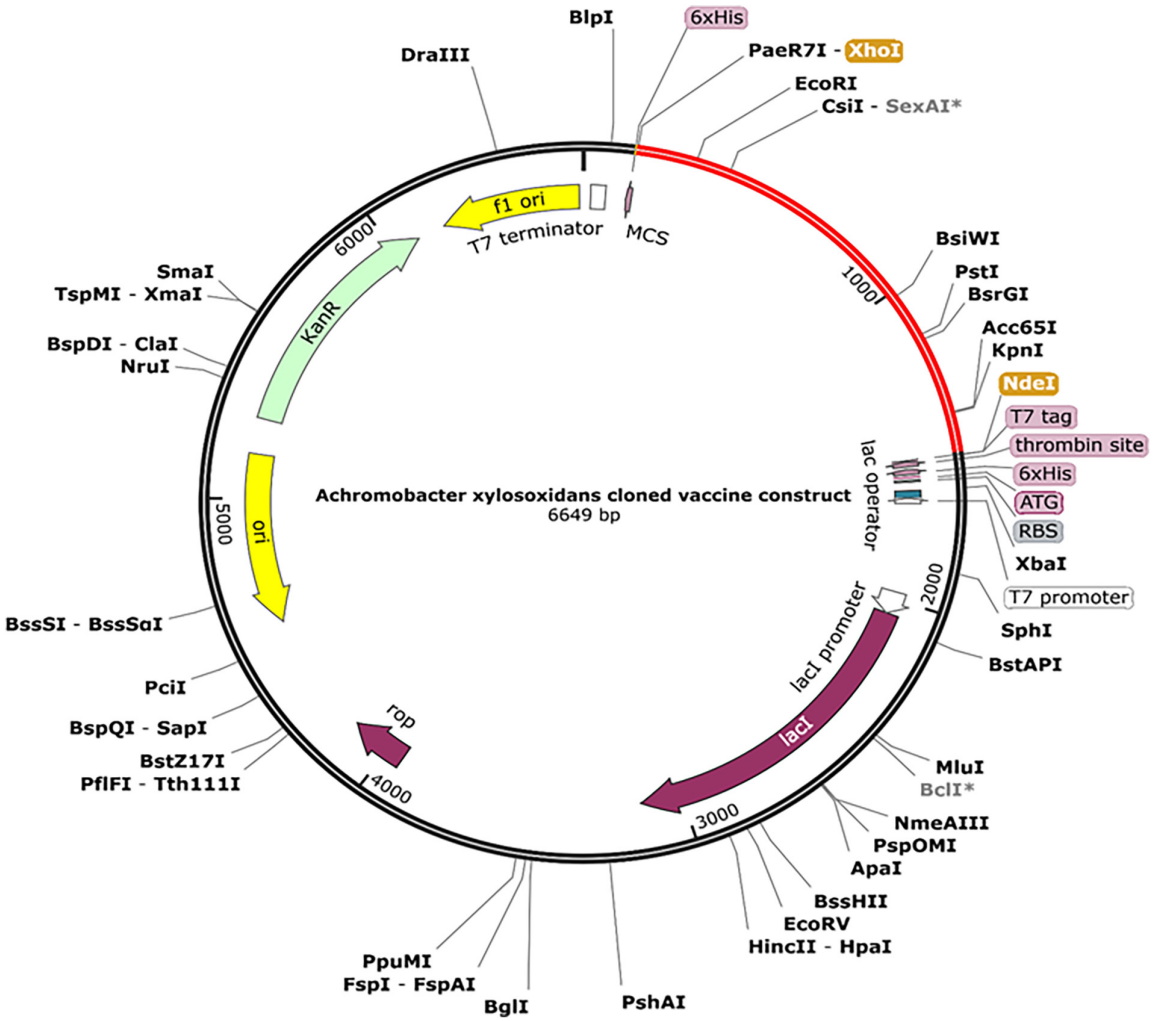


FIGURE 10 | A final clone of the protein in *E. coli* plasmid [pET28a (+)] *via in silico* cloning. The red portion depicts designed vaccine, whereas the black portion is structure of *E. coli* vector backbone. The C-terminus of cloned construct contains a 6xHis tag.

Docking Analysis of the Ensemble With Immune Receptors

Docking analysis between TLR-4 and the multi-epitope vaccine was accomplished *via* an online server, ZDOCK (54). Overall, ten vaccine-TLR combinations were generated, but only the optimal combination was selected. Fourteen hydrogen bonds were reported to be present between the vaccine construct and the receptor according to the PDBsum server (**Figure 8** and **Supplementary Table 3**).

Codons Optimization and Computational Cloning

The designed construct was tailored to use codons consistent with those of *E. coli* (strain K12) to maximize protein expression (**Figure 9**). Optimal codon sequence length was 1,353 nucleotides. GC content of 53.28% and a codon adaptation index score of 0.97 indicated the vaccine's high expression in host cells. Finally, the optimized sequence was cloned into the pET28a (+) vector (**Figure 10**).

Immune Simulations

By simulating the *in silico* human immune responses to numerous doses of antigen using the C-ImmSim web server (59), we evaluated the immune system response. The simulation produced significantly more secondary and tertiary responses after injections than primary reactions. Moreover, the antigenic concentrations declined after doses, but antibody titers (IgG1 + IgG2, IgM, and IgG + IgM) increased significantly, indicating that the immune system had been greatly stimulated (**Figure 11A**). Additionally, numerous long-lasting B cell isotypes were identified, signifying potential switch-over between isotypes and possible memory development (**Figures 11B,C**). Similarly, a proliferating level of helper and cytotoxic T cells suggested the development of secondary and tertiary immune responses (**Figures 11D,E**). Likewise, we tested the Interleukin (IL) and

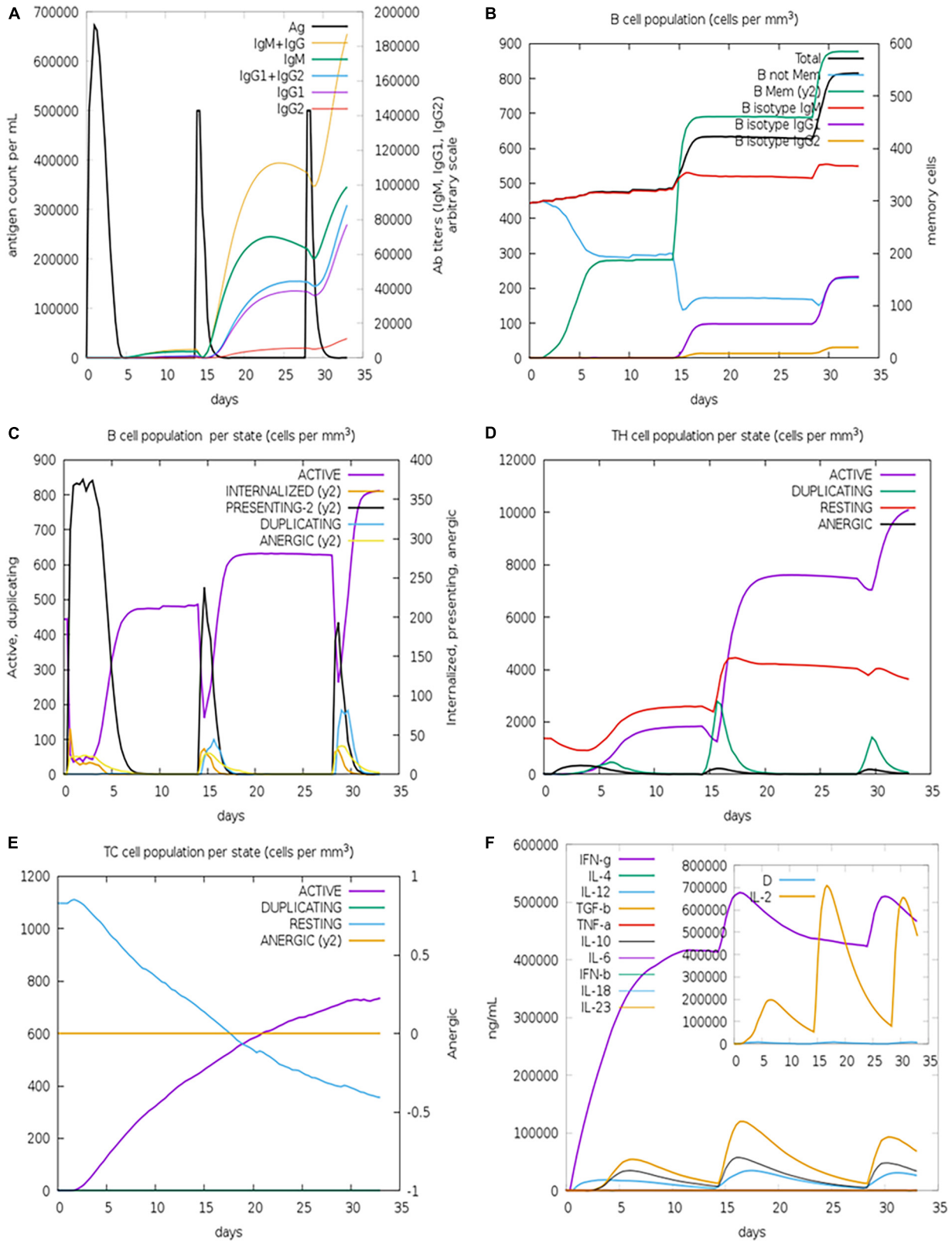


FIGURE 11 | Immune simulations of the designed construct. **(A)** Immune responses to three doses of vaccination. **(B)** Increase in antibody generating plasma cells. **(C)** B-cells/state. **(D)** TH population/state. **(E)** TC population/state. **(F)** Cytokine production levels.

cytokine levels. The high levels of IFN-γ and IL-12 prompted in the simulation further contributed to generating a consistent and robust immune response (**Figure 11F**). As a result, an effective immune response was shown to be elicited by the designed construct. **Figures 11A–F** illustrates the results as generated by the immune server.

DISCUSSION

This genome era brought with it a large amount of proteome and genome related data of nearly every clinically relevant organism. By using this knowledge, we are able to choose putative candidates for drug development, identify strains of bacteria or antibiotic resistance and develop diagnostic kits. Furthermore, *in silico* approaches improved the development of vaccines; vaccinomics is an emerging field that allows for *in silico* vaccine design. The *A. xylosoxidans*, like other Gram-negatives, elicit host immune responses *via* interaction with toll-like receptors (60). In this study, few immunoinformatic approaches have been employed to detect epitopes and design a construct that interacts efficaciously with toll-like receptors.

The whole proteome of reference strain of *A. xylosoxidans* was retrieved to carry out the study on epitopes prediction. So far, no studies have been performed to design a T cell and B cell containing multi-epitope vaccine. Here, we have performed a study to design a vaccine that is not only safe and potent but also stable in nature. Studies have reported that epitopes or antigens induce cellular as well as humoral immunity that is instigated by T-cells and antibodies, respectively (61).

The combination of multiple tools led to the selection of twenty epitopes, eleven CTLs, four HTLs, and five B cell epitopes that were non-allergenic, non-toxic, and immune potentiating. The selected epitopes were adjoined using linkers to finally construct a vaccine ensemble. Proteins typically require an adjuvant at the N-terminus in order to stimulate the immune response; we added ribosomal protein L7/L12 to enhance immunostimulatory properties (62).

After we proposed 3D structure of the construct, we sent it for additional refinement following which the Ramachandran plot informed us that more than 86% of residues occupied the favored region, emphasizing the modeled vaccine's high quality.

The presence of TLRs (particularly TLR 4) is linked to the important role of recognition of antigens from Gram-negative pathogens (63). Therefore, our research investigated how designed chimera interact with these receptors. Using ZDOCK, we determined that the vaccine was a good ligand for immunological receptors, because when compared to the binding ability of the KDO crystal ligand (−622 kcal/mol), the binding energy of the vaccine construct was much greater (1559.49 kcal/mol) (**Supplementary Table 3**). In protein-receptor interactions, hydrogen bonds were found to be present that contribute significantly in influencing stability of complexes. The simulated immune responses elucidated the effectiveness of the designed construct in eliciting the immune response. Injections of vaccine increased memory cells whose levels persisted even after a third injection. The most preferred host is *E. coli* for producing more recombinant vaccine (64). A GC content of 53.28% and a CAI of 0.97 and was found using the JCAT tool, which supported the favorability of vaccine design for high-level expression in the chosen host i.e., *E. coli*.

CONCLUSION

In the wake of recent emergence of *A. xylosoxidans*, our research focused on the possibility of developing a multi-epitope vaccine *via* computational modeling approach. As of now, no proper medical preventive measures such as vaccines are available in the market. Through *in silico* techniques, an effective vaccine can be developed in a shorter amount of time while maintaining low costs. Using immunoinformatic tools, we constructed a vaccine containing HTL, CTL and B cell epitopes that could instigate robust immune responses against *A. xylosoxidans*. Antigenic and immunogenic qualities were found to be present in the proposed construct. The MD simulations confirmed compactness of the designed chimera, and molecular docking studies showed stable interactions with immune receptor proteins. Finally, *in silico* cloning showed optimum expression and effectiveness of the construct in triggering the immune system based on immune simulation studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

KK and YW: conceptualization and writing—review editing. KK, MA, and YW: data curation. KK, US, MA, MIA, and AA: formal analysis. KK, US, MA, and AA: investigation. KK, MA, MIA, AA, and YW: methodology. MA and YW: resources. KK: software. YW: supervision. KK, US, and MA: validation. KK and MIA: visualization. KK: writing—original draft. All authors contributed to the article and approved the submitted version.

FUNDING

MA received a fund from the Researcher Supporting Project number (RSP2022R481), King Saud University, Riyadh, Saudi Arabia, to support publishing of this article. YW work is funded by Foundation University Islamabad's internal grants by FUI ORIC. No funding agencies were involved in the study design, analysis, interpretation of data, or writing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.902611/full#supplementary-material

REFERENCES

1. Wittmann J, Dreiseikelmann B, Rohde M, Meier-Kolthoff JP, Bunk B, Rohde C. First genome sequences of *Achromobacter* phages reveal new members of the N4 family. *Virol J.* (2014) 11:1–15. doi: 10.1186/1743-422X-11-14/FIGURES/9
2. Awadh H, Mansour M, Aqtash O, Shweihat Y. Pneumonia due to a rare pathogen: *Achromobacter xylosoxidans*, subspecies denitrificans. *Case Rep Infect Dis.* (2017) 2017:1–4. doi: 10.1155/2017/3969682
3. Yabuuchi E, Ohyama A. *Achromobacter xylosoxidans* n. sp. from human ear discharge. *Jpn J Microbiol.* (1971) 15:477–81. doi: 10.1111/J.1348-0421.1971.TB00607.X
4. Choi H, Jinadatha C, Chatterjee P, Allton Y, Navarathna DH. Draft genome sequence of an unusually multidrug-resistant strain of *Achromobacter xylosoxidans* from a blood isolate. *Microbiol Resour Announc.* (2020) 9:e194–120. doi: 10.1128/MRA.00194-20
5. Tena D, González-Praetorius A, Pérez-Balsalobre M, Sancho O, Bisquert J. Urinary tract infection due to *Achromobacter xylosoxidans*: report of 9 cases. *Scand J Infect Dis.* (2008) 40:84–7. doi: 10.1080/00365540701558714
6. Tokuyasu H, Fukushima T, Nakazaki H, Shimizu E. Infective endocarditis caused by *Achromobacter xylosoxidans*: a case report and review of the literature. *Intern Med.* (2012) 51:1133–8. doi: 10.2169/INTERNALMEDICINE.51.6930
7. Imani S, Wijetunga A, Shumborski S, O'Leary E. Chronic osteomyelitis caused by *Achromobacter xylosoxidans* following orthopaedic trauma: a case report and review of the literature. *IDCases.* (2021) 25:e01211. doi: 10.1016/J.IDCR.2021.E01211
8. Eshwara VK, Mukhopadhyay C, Mohan S, Prakash R, Pai G. Two unique presentations of *Achromobacter xylosoxidans* infections in clinical settings. *J Infect Dev Ctries.* (2011) 5:138–41. doi: 10.3855/JIDC.1258
9. Gabrielaite M, Bartell JA, Nørskov-Lauritsen N, Pressler T, Nielsen FC, Johansen HK, et al. Transmission and antibiotic resistance of *Achromobacter* in cystic fibrosis. *J Clin Microbiol.* (2021) 59:e2911–20. doi: 10.1128/JCM.02911-20/SUPPL_FILE/JCM.02911-20-S0001.PDF
10. Asano K, Tada S, Matsumoto T, Miyase S, Kamio T, Sakurai K, et al. A novel bacterium *Achromobacter xylosoxidans* as a cause of liver abscess: three case reports. *J Hepatol.* (2005) 43:362–5. doi: 10.1016/j.jhep.2005.03.031
11. Zhu Z, Xu J, He F. Genomic and phylogenetic analysis of multidrug-resistant *Achromobacter xylosoxidans* ST273 strain MTYH1 co-carrying blaOXA-114g and blaCARB-2 recovered from a wound infection in China. *J Glob Antimicrob Resist.* (2021) 25:110–3. doi: 10.1016/J.JGAR.2021.03.008
12. Isler B, Kidd TJ, Stewart AG, Harris P, Paterson DL. *Achromobacter* infections and treatment options. *Antimicrob Agents Chemother.* (2020) 64:e1025–1020. doi: 10.1128/AAC.01025-20
13. Caverly LJ, Spilker T, Kalikin LM, Stillwell T, Young C, Huang DB, et al. In vitro activities of β-lactam-β-lactamase inhibitor antimicrobial agents against cystic fibrosis respiratory pathogens. *Antimicrob Agents Chemother.* (2019) 64:e1595–1519. doi: 10.1128/AAC.01595-19
14. Díez-Aguilar M, Ekkelenkamp M, Morosini MI, Merino I, de Dios Caballero J, Jones M, et al. Antimicrobial susceptibility of non-fermenting Gram-negative pathogens isolated from cystic fibrosis patients. *Int J Antimicrob Agents.* (2019) 53:84–8. doi: 10.1016/J.IJANTIMICAG.2018.09.001
15. Chen Z, Fang H, Wang L, Sun F, Wang Y, Yin Z, et al. IMP-1 encoded by a novel Tn402-like class 1 integron in clinical *Achromobacter xylosoxidans*, China. *Sci Rep.* (2014) 4:1–7. doi: 10.1038/srep07212
16. Rosini R, Nicchi S, Pizza M, Rappuoli R. Vaccines against antimicrobial resistance. *Front Immunol.* (2020) 11:1048. doi: 10.3389/FIMMU.2020.01048/BIBTEX
17. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol.* (2018) 15:182–4. doi: 10.1038/cmi.2017.92
18. Scarselli M, Aricò B, Brunelli B, Savino S, Di Marcello F, Palumbo E, et al. Rational design of a meningococcal antigen inducing broad protective immunity. *Sci Transl Med.* (2011) 3:91ra62. doi: 10.1126/SCITRANSLMED.3002234/SUPPL_FILE/3-91RA62_SM.PDF
19. Dar HA, Zaheer T, Shehroz M, Ullah N, Naz K, Muhammad SA, et al. Immunoinformatics-aided design and evaluation of a potential multi-epitope vaccine against *Klebsiella Pneumoniae*. *Vaccines.* (2019) 7:88. doi: 10.3390/vaccines7030088
20. Khalid K, Irum S, Ullah SR, Andleeb S. In-silico vaccine design based on a novel vaccine candidate against infections caused by *Acinetobacter baumannii*. *Int J Pept Res Ther.* (2022) 28:16. doi: 10.1007/S10989-021-10316-7
21. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of putative vaccine candidates against *Helicobacter* pylori exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol.* (2015) 32:280–91. doi: 10.1016/J.MEEGID.2015.03.027
22. Haq AU, Khan A, Khan J, Irum S, Waheed Y, Ahmad S, et al. Annotation of potential vaccine targets and design of a multi-epitope subunit vaccine against *Yersinia pestis* through reverse vaccinology and validation through an agent-based modeling approach. *Vaccines.* (2021) 9:1327. doi: 10.3390/VACCINES9111327
23. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics.* (2007) 8:424. doi: 10.1186/1471-2105-8-424
24. 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. doi: 10.1093/BIOINFORMATICS/BTQ249
25. IEDB. *org: Free Epitope Database and Prediction Resource*. (2022). Available online at: https://www.iedb.org/ (accessed January 12, 2022).
26. El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit.* (2008) 21:243–55. doi: 10.1002/JMR.893
27. Suleman M, Qamar MTU, Kiran, Rasool S, Rasool A, Albutti A, et al. Immunoinformatics and immunogenetics-based design of immunogenic peptides vaccine against the emerging tick-borne encephalitis virus (tbev) and its validation through in silico cloning and immune simulation. *Vaccines.* (2021) 9:1210. doi: 10.3390/VACCINES9111210
28. Dong R, Chu Z, Yu F, Zha Y. Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front Immunol.* (2020) 11:1784. doi: 10.3389/fimmu.2020.01784
29. 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) 12:490. doi: 10.3389/FIMMU.2021.621706/BIBTEX
30. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* (2006) 34:W202–9. doi: 10.1093/NAR/GKL343
31. 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. doi: 10.1007/s00894-014-2278-5
32. Flower DR, Doytchinova I, Zaharieva N, Dimitrov I. Immunogenicity prediction by vaxijen: a ten year overview. *J Proteomics Bioinform.* (2017) 10:11. doi: 10.4172/jpb.1000454
33. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics.* (2009) 25:2200–7. doi: 10.1093/BIOINFORMATICS/BTP386
34. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* (2017) 33:3098–100. doi: 10.1093/bioinformatics/btx345
35. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* (2012) 40:W597–603. doi: 10.1093/nar/gks400
36. Mcguffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* (2000) 16:404–5.
37. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods.* (2014) 12:7–8. doi: 10.1038/nmeth.3213
38. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct Funct Genet.* (2004) 57:702–10. doi: 10.1002/PROT.20264
39. Bibi S, Ullah I, Zhu B, Adnan M, Liaqat R, Kong WB, et al. In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci Rep.* (2021) 11:1–16. doi: 10.1038/s41598-020-80899-6
40. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J.* (2011) 101:2525–34. doi: 10.1016/j.bpj.2011.10.024

41. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* (2013) 41:W384–8. doi: 10.1093/nar/gkt458
42. Zhang J, Zhao X, Sun P, Gao B, Ma Z. Conformational B-cell epitopes prediction from sequences using cost-sensitive ensemble classifiers and spatial clustering. *Biomed Res Int.* (2014) 2014:689219. doi: 10.1155/2014/689219
43. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* (2007) 35:W407–10. doi: 10.1093/nar/gkm290
44. Dym O, Eisenberg D, Yeates T. *ERRAT*. (2012). Available online at: https://onlinelibrary.wiley.com/iucr/itc/Fb/ch21o3v0001/sec21o3o3o4/ (accessed May 6, 2021).
45. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins Struct Funct Bioinforma.* (2003) 50:437–50. doi: 10.1002/PROT.10286
46. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics.* (2003) 19:2500–1. doi: 10.1093/BIOINFORMATICS/BTG362
47. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* (2005) 15:1451–5. doi: 10.1101/gr.4086505
48. Dombkowski AA. Disulfide by DesignTM: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics.* (2003) 19:1852–3. doi: 10.1093/BIOINFORMATICS/BTG231
49. Potocnakova L, Bhide M, Pulzova LB. An introduction to B-cell epitope mapping and in silico epitope prediction. *J Immunol Res.* (2016) 2016:6760830. doi: 10.1155/2016/6760830
50. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics.* (2008) 9:514. doi: 10.1186/1471-2105-9-514
51. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* (2015) 1–2:19–25. doi: 10.1016/j.softx.2015.06.001
52. Al-Karmalawy AA, Dahab MA, Metwaly AM, Elhady SS, Elkaeed EB, Eissa IH, et al. Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the hACE2 receptor. *Front Chem.* (2021) 9:661230. doi: 10.3389/FCHEM.2021.661230/FULL
53. Amemiya K, Dankmeyer JL, Bernhards RC, Fetterer DP, Waag DM, Worsham PL, et al. Activation of toll-like receptors by live gram-negative bacterial pathogens reveals mitigation of TLR4 responses and activation of TLR5 by flagella. *Front Cell Infect Microbiol.* (2021) 11:1089. doi: 10.3389/FCIMB.2021.745325/BIBTEX
54. Chen R, Li L, Weng ZZDOCK. An initial-stage protein-docking algorithm. *Proteins Struct Funct Bioinforma.* (2003) 52:80–7. doi: 10.1002/PROT.10389
55. Shey R, Ghogomu S, Esoh K. In-silico design of a multi-epitope vaccine candidate against Onchocerciasis and related filarial diseases. *Sci Rep.* (2019) 9:4409.
56. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. *Proteins Struct Funct Genet.* (2007) 69:139–59. doi: 10.1002/PROT.21495
57. Laskowski RA, Jabło nska J, Pravda L, Vaøeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. *Wiley Online Libr.* (2017) 27:129–34. doi: 10.1002/pro.3289
58. 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. doi: 10.1093/nar/gki376
59. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One.* (2010) 5:9862. doi: 10.1371/journal.pone.0009862
60. Swenson CE, Sadikot RT. *Achromobacter* respiratory infections. *Ann Am Thorac Soc.* (2015) 12:252–8. doi: 10.1513/ANNALSATS.201406-288FR/SUPPL_FILE/DISCLOSURES.PDF
61. Kogay R, Schönbach C. Epitope predictions. *Encycl Bioinforma Comput Biol ABC Bioinforma.* (2019) 1–3:952–71. doi: 10.1016/B978-0-12-809633-8.20248-3
62. Samad A, Ahammad F, Nain Z, Alam R, Imon RR, Hasan M, et al. Designing a multi-epitope vaccine against SARS-CoV-2: an immunoinformatics approach. *J Biomol Struct Dyn.* (2020) 1–17:14–30. doi: 10.1080/07391102.2020.1792347
63. Tsujimoto H, Ono S, Efron PA, Scumpia PO, Moldawer LL, Mochizuki H. Role of toll-like receptors in the development of sepsis. *Shock.* (2008) 29:315–21. doi: 10.1097/SHK.0B013E318157EE55
64. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* (2014) 5:172. doi: 10.3389/FMICB.2014.00172/FULL

Conflict of Interest: AA was employed by the company PerkinElmer Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fmed.2022.923715

Unexpected Detection of Anti-SARS-CoV-2 Antibodies Before the Declaration of the COVID-19 Pandemic

Waleed Mahallawi and Nadir Ibrahim*

Medical Laboratory Technology Department, College of Applied Medical Sciences, Taibah University, Medina, Saudi Arabia

Background: Limited information is currently available regarding the global incidence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections prior to the declaration of the coronavirus disease 2019 (COVID-19) pandemic, which may result in improper conclusions regarding the timing of viral transmission.

Methods: We investigated the presence of specific antibodies against the receptor-binding domain (RBD) of SARS-CoV-2 in archived serum samples that were collected from 478 healthy blood donors and patients in Madinah, Saudi Arabia, between October 2019 and January 2020. Enzyme-linked immunosorbent assay (ELISA) was performed to measure SARS-CoV-2 IgM and IgG antibodies. In addition, rheumatoid factor (RF) and urea dissociation tests were performed in all samples, which showed seropositivity for the SARS-CoV-2 IgM antibody. Additionally, Chemiluminescence immunoassays (CLIA) targeting the RBD of the SARS-CoV-2 spike (S) protein were performed to confirm the seropositivity of the samples.

Results: Overall, 20 (4.18%) serum samples were detected by ELISA to have SARS-CoV-2 IgG or IgM antibodies. Of these, 12 (2.51%) samples were positive for IgM antibody, and 8 (1.67%) were positive for IgG antibody. The 12 samples positive for SARS-CoV-2 IgM antibody were subjected to RF and urea dissociation tests, and all samples were RF-negative. The ELISA results were negative for 7 (58.33%) samples when subjected to urea dissociation prior to ELISA, whereas the other 5 (41.67%) samples remained positive. These 5 samples remained positive for the anti-S RBD IgG antibody in the CLIA. In addition, 3 of the 8 samples with IgG positivity according to the ELISA remained positive in the CLIA. After reviewing their data, we discovered that the 8 CLIA-confirmed positive samples were obtained from returned travellers who had visited China during the 4-week period immediately preceding blood donation.

Conclusion: In conclusion, we found evidence to support the early circulation of SARS-CoV-2 among persons who visited China a few months prior to the pandemic declaration. These results can be used to better define the spread of SARS-CoV-2 infections before the COVID-19 pandemic declaration. The detection of SARS-CoV-2 antibodies in individuals before the pandemic was declared in China could rewrite the pre-pandemic timeline.

Keywords: COVID-19, archived samples, anti-SARS-CoV-2 IgM/IgG, pre-pandemic, SARS-CoV-2

OPEN ACCESS

Edited by:
Adriana Harbuzariu, Emory University, United States

Reviewed by:
Farhat Afrin, JIS Institute of Advanced Studies and Research, India
Abdelwahid Saeed Ali, King Khalid University, Saudi Arabia
Vivian Rotuno, Federal University of Paraná, Brazil

****Correspondence:***
Waleed Mahallawi
wmahallawi@gmail.com

Specialty section:
This article was submitted to Infectious Diseases – Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *19 April 2022*
Accepted: *20 June 2022*
Published: *11 July 2022*

Citation:
Mahallawi W and Ibrahim N (2022) Unexpected Detection of Anti-SARS-CoV-2 Antibodies Before the Declaration of the COVID-19 Pandemic. Front. Med. 9:923715. doi: 10.3389/fmed.2022.923715

INTRODUCTION

By the end of December 2019, the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus was identified as the underlying cause of coronavirus disease 2019 (COVID-19), which was first detected in China and rapidly spread worldwide (1). The first case of laboratory-confirmed COVID-19 in the Kingdom of Saudi Arabia was recorded on March 2, 2020. Since then, infections have continued to increase on a daily basis (2). Serologic tests represent a useful tool for assessing the prevalence of SARS-CoV-2 antibodies, which can be used to track viral infection and evaluate its frequency among the population (3). Specific SARS-CoV-2-targeted antibodies have been identified, providing a convenient method for both the clinical surveillance of public health and the monitoring of infected patients (4). However, it has been reported that significant RF reactivity interferes with tested antibodies when immunoassays such as ELISA performed. Additionally, RF has been found to be linked with higher false positive results used in clinical settings as well as researches. (5). Performing population-wide serosurveys for SARS-CoV-2 antibodies can allow for the precise evaluation of the infection rate and can also be used to detect asymptomatic cases (6, 7).

Differences in the humoral immune responses observed among hospitalized patients have revealed a relationship between disease presentation and the immune response mounted to combat the virus (8). Analyses of the antiviral antibody responses in patients diagnosed with COVID-19 indicate that the seroconversion of SARS-CoV-2-targeted IgM and IgG antibodies can be detected within 13 days of symptom onset in nearly all patients with COVID-19 (9).

Abundant information is currently available describing the spread of SARS-CoV-2; however, inadequate evidence has been presented regarding the early spread of the virus leading to the first laboratory-confirmed case. Several studies have been conducted globally to investigate available evidence to help trace the exact or approximate timing of SARS-CoV-2 spread and the viral origins. For example, a recent study reported the evaluation of stored samples obtained from Vietnamese children and adults collected during a pre-pandemic period in Vietnam (2015–2019), which found no evidence of antibodies targeting the SARS-CoV-2 nucleocapsid and spike (S) proteins antibodies in pre-pandemic samples (10). Another study conducted in Italy reported the unexpected early circulation of SARS-CoV-2 among asymptomatic persons in Italy 3 months prior to the detection of the first recognized case, allowing for the more precise mapping of early infections and the spread of the COVID-19 pandemic (11).

Our study investigated the presence of SARS-CoV-2 antibodies in archived serum samples obtained during studies conducted prior to the declaration of the COVID-19 pandemic in Saudi Arabia.

MATERIALS AND METHODS

Serum Samples and Controls

This study examined 478 archived, stored serum samples obtained from blood bank donors and patients in Madinah, Saudi Arabia, during serosurvey projects conducted between October 2019 and January 2020 and stored at −80°C. All blood donors were required to be healthy and disease-free at the time of blood donation. Two serum samples were collected at Madinah General Hospital in February 2022 from symptomatic COVID-19-infected patients <14 days after confirmation of SARS-CoV-2 infection by real-time reverse-transcriptase-polymerase chain reaction (rRT-PCR) and stored at −80°C for use as positive controls. All the positive samples in the current study were from individuals who had visited China prior to the pandemic.

Signed consent forms were obtained from all participants and patients included in this study. Ethical approval to conduct this study was acquired from the Research Ethics Committee of the Institutional Review Board, General Directorate of Health Affairs in Madinah (IRB no: H-03-M-084).

Serological Assays

SARS-CoV-2 IgM and IgG Enzyme-Linked Immunosorbent Assay

Commercially available SARS-CoV-2 Virus IgM and IgG Antibody Detection enzyme-linked immunosorbent assay (ELISA) kits (BGI Europe, Copenhagen, Denmark) were used to detect the presence of SARS-CoV-2 IgM and IgG antibodies in human serum, according to the manufacturer's instructions. The test specificity for IgM antibody detection is 96.76%, the specificity for IgG antibody detection is >98.38%, and the sensitivity for total IgM and IgG detection is >98.71% (https://www.bgi.com/global/molecular-genetics/COVID-19-antibody-detection-kit-elisa/, accessed on March 21, 2022). The results were interpreted as follows: an optical density (OD) greater than the cutoff value was categorized as positive; an OD below the cutoff value was categorized as negative. All washing and reading steps were performed using a semi-automated ELISA washer and reader (Biotek, Winooski, US).

Rheumatoid Factor Test

It has been reported that significant RF reactivity interferes with tested antibodies when immunoassays such as ELISA performed. Additionally, RF has been found to be linked with higher false positive result when used in clinical settings as well as researches. To assess the presence of possible interference due to rheumatoid factor (RF) in the SARS-CoV-2 Virus IgM ELISA, sera that tested positive for SARS-CoV-2 IgM were subjected to a semi-quantitative RF latex agglutination slide test (HumaTex RF, Human Gesellschaft fur Biochemica und Diagnostica mbH, Wiesbaden, Germany, CN: 40050), according to the manufacturer's instructions. Test performance characteristics can be found at http://www.human-de.com/data/gb/vr/lx-rf.pdf (accessed on March 21, 2022).

Urea Dissociation Test

The urea dissociation test was conducted on samples that tested positive for SARS-CoV-2 IgM by ELISA to assess the reliability of positive SARS-CoV-2 IgM results. The results of affinity index (AI) analyses are expressed as the ratios of the OD values obtained for the 4 mol/L concentration of dissociated urea to the OD values obtained from serum samples. The AI threshold value was established as the median value between the highest AI

value obtained for all tested serum samples without outliers and the lowest AI value obtained for the rRT-PCR–positive control samples. Sera with AI values greater than or equal to the AI threshold were defined as positive, whereas sera with AI values less than the AI threshold were defined as negative (12).

SARS-CoV-2 IgG Chemiluminescence Immunoassays

To confirm positive results detected using the SARS-CoV-2 IgM and IgG ELISA (13), the Elecsys AntiSARS-CoV-2 S immunoassay (La Roche Ltd, Basel, Switzerland), a Chemiluminescence immunoassay (CLIA), was used to provide *in vitro* quantification of antibodies targeting the SARS-CoV-2 S protein receptor-binding domain (RBD). The assay utilizes the RBD of recombinant SARS-CoV-2 S protein in a double-sandwich assay format, with a clinical specificity of >99.98% (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/989460/Evaluation_of_Roche_Elecsys_anti_SARS_CoV_2_S_assay_PHE.pdf). The Roche cobas e411 immunoassay analyzer was used according to the manufacturer's instructions. The results are reported as the concentration of anti-SARS-CoV-2 S antibodies detected in each sample (in U/mL), with values <0.8 U/mL categorized as negative and values ≥0.8 U/mL categorized as positive (Roche Diagnostics GmbH. 2022. Elecsys® Anti-SARS-CoV-2 S Assay. Available from: https://diagnostics.roche.com/global/en/products/params/elecsys-anti-SARS-CoV-2.html. Accessed on March 21, 2022). Concentrations in U/mL are directly equivalent to binding antibody units (BAU)/mL defined in the first World Health Organization International Standard for anti-SARS-CoV-2 immunoglobulin detection (NIBSC code 20/136), and results in U/mL may be directly compared to values reported in BAU/mL (14).

Statistical Analysis

Qualitative variables are presented as the frequency and percentage, whereas quantitative variables are presented as the mean ± standard deviation. Data were analyzed using the Statistical Package for Social Sciences (IBM SPSS Statistics version 25; Armonk, NY, USA).

RESULTS

Demographic Characteristics of the Study Population

The demographic characteristics for the study population are shown in **Table 1**. Of the 478 samples, 301 (62.97%) were obtained from men, 367 (76.78%) were obtained from Saudis, and the population had a mean age of 32.01 ± 10.12 years (range: 18–66 years).

TABLE 1 | Demographic characteristics of study population ($n = 478$).

Parameter		
Sex	Male [*n* (%)]	301 (62.97)
	Female	177 (37.03)
Nationality	Saudi [*n* (%)]	367 (76.78)
	Non-Saudi	111 (23.22)
Age	Age range	18–66
	Age mean (mean ± SD)	32.01 ± 10.12

Serological Assays for the Detection and Confirmation of SARS-CoV-2 Antibodies

In **Table 2**, we describe the results of serological assays used to detect and confirm the presence of SARS-CoV-2 antibodies in the study samples. Overall, 20 (4.18%) serum samples were positive for the presence of SARS-CoV-2 IgM or IgG antibodies by ELISA. IgM antibodies were detected in 12 (2.51%) patients, IgG antibodies were found in 8 (1.67%) patients, and both IgM and IgG antibodies were found in only 1 (0.21%) patient. The rRT-PCR–positive control samples were positive for SARS-CoV-2 IgM antibody.

The 13 samples identified as positive for SARS-CoV-2 IgM antibody by ELISA were subjected to RF and urea dissociation tests. Only one sample tested positive for RF at a concentration of 384 IU/mL. Remarkably, this sample tested positive for both IgM and IgG using the SARS-CoV-2 IgM antibody detection ELISA; however, when the sample was subjected to the urea dissociation test at a concentration of 4 mol/L urea using an AI threshold value of 0.835, the sample was categorized as negative. Additionally, when the sample was subjected to further testing for SARS-CoV-2 antibodies using Elecsys Anti-SARS-CoV-2 S immunoassay, the result was negative for antibodies against the SARS-CoV-2 S protein RBD. The remaining 12 SARS-CoV-2 IgM antibody-positive samples were RF-negative. When the urea dissociation test was performed on these samples and the 2 rRT-PCR-positive controls, the results were negative for 7 samples and remained positive for 5 samples and the rRT-PCR-positive controls.

Notably, antibodies against the SARS-CoV-2 S protein RBD were detected in all 5 samples that remained positive following urea treatment. After reviewing the data for these 5 patients, we discovered that these samples were obtained from non-Saudi foreign students who had traveled to China in the weeks prior to blood donation.

SARS-CoV-2 IgG-positive samples identified by ELISA were subjected to further testing for antibodies against the SARS-CoV-2 S protein RBD using the Elecsys Anti-SARS-CoV-2 S immunoassay. Only 3 samples remained positive following the CLIA.

DISCUSSION

Scientists continue to seek new information to better understand the earliest periods of the pandemic prior to the global spread of the infection. Various reports have offered additional proof to support the hypothesis that SARS-CoV-2 was initially contracted from animals by the humans who raised, butchered, or bought them. However, currently available information has not yet determined the precise course of events (15–17).

By contrast, a study concerning the origins of the virus indicated that viral origin inquiries commonly take several years

TABLE 2 | Serological assays for detection and confirmation of SARS-CoV-2 antibodies in the study samples.

Samples	ELISA		RF (IU/mL)	ELISA after urea dissociation	CLIA (U/mL)	
	IgM	IgG				
Sample1	−	+	<12	ND	0.165	−
Sample 2	−	+	<12	ND	0.173	−
Sample 3	−	+	<12	ND	10.071	+
Sample 4	−	+	<12	ND	12.899	+
Sample 5	−	+	<12	ND	9.045	+
Sample 6	−	+	<12	ND	0.103	−
Sample 7	+	+	384	−	0.270	−
Sample 8	−	+	<12	ND	0.146	−
Sample 9	+	−	<12	+	3.118	+
Sample 10	+	−	<12	+	2.438	+
Sample 11	+	−	<12	−	0.113	−
Sample 12	+	−	<12	−	0.312	−
Sample 13	+	−	<12	−	0.421	−
Sample 14	+	−	<12	+	2.990	+
Sample 15	+	−	<12	−	0.395	−
Sample 16	+	−	<12	−	0.521	−
Sample 17	+	−	<12	−	0.234	−
Sample 18	+	−	<12	+	3.806	+
Sample 19	+	−	<12	−	0.341	−
Sample 20	+	−	<12	+	4.036	+
Positive controls ($n = 2$)	+	ND	ND	+	ND	

ND, not done. (–), negative and (+), positive.

and reported that Chinese researchers have already conducted a number of relevant studies (18–20).

In our study, any evidence that supports or opposes published data concerning the viral origins or data that has been withheld from authorities will not be discussed. Our only goal is to disseminate our findings to the scientific community.

We assume that many factors may confound the determination of the exact timing of the pandemic onset and viral spread, such as the under-reporting of SARS-CoV-2 cases. Additionally, a major factor that may have contributed to the difficulties tracing SARS-CoV-2 was the surge of a novel virus commonly associated with asymptomatic cases, which resulted in the spread of the virus among populations without diagnosis. Reports have indicated that at least 50% of transmissions were likely due to asymptomatic individuals (21).

Although useful, serological analyses such as those used in our study to detect the SARS-CoV-2 virus are not error-free, therefore, we applied several assays to verify our findings. To exclude the possibility of RF interference in the SARS-CoV-2 IgM ELISA, we tested our samples for the presence of RF. A majority of the certified serological tests were found to be sensitive to interfering antibodies, such as RF, which are present in the serum of patients who suffer from chronic inflammatory diseases (22).

Our results suggest that the presence of RF IgM may result in false-positive reactivity in the SARS-CoV-2 IgM ELISA (12). Only one sample among the identified IgM-positive samples was also RF-positive. When we performed the ELISA after a urea dissociation step, only 5 of the 12 initially identified positive samples remained positive, indicating that the remaining 7 samples were false positives. We also confirmed our ELISA results by CLIA using the S protein RBD.

One informative study in China verified that the first zoonotic spread of the virus was estimated to have occurred in late November/early December of 2019 and no earlier than the start of November 2019 (23). Thus, the potential spread of the virus among humans at that time is also possible, particularly among asymptomatic individuals. The timeline regarding the emergence of the virus is currently poorly defined; therefore, we assume that the positive samples identified in our study are likely due to sporadic infections among individuals that were not recognized by the government at the time.

Our results detecting anti-SARS-CoV-2 antibodies among individuals who had recently visited China suggest that the virus was present and circulating before the declaration of a pandemic. The blood bank in Madinah performs donation campaigns each year to supply hospitals with blood and blood products for use in patients for medical reasons. In our study, we discovered that the 8 CLIA-confirmed positive samples were obtained from returned travelers who had been in China 2–4 weeks after November 2019. Therefore, the chances that these individuals were infected while in China are very high. However, when they presented for blood donation, they appeared to be healthy and were deemed to be eligible for donation, with no reports of any unusual symptoms, indicating that they experienced asymptomatic infections. Recent data from China indicate that the vast majority of coronavirus infections do not lead to the development of symptoms (24). This

report supports our hypothesis that asymptomatic infections are a likely factor contributing to the spread of COVID-19 (25).

Seroconversion following infection varies, ranging from 50% at 11 days after infection to 100% at 39 days post-infection (8, 26). Asymptomatic individuals appear to present with lower levels of seroconversion and antibody persistence (27); however, additional research in larger cohort investigations remains necessary to confirm this observation (28). In our study, we found that 5 samples were IgG-negative and IgM-positive by ELISA, whereas all 5 of these samples were determined to be IgG-positive by CLIA. The seroconversion for IgG and IgM may occur either simultaneously or consecutively, and both IgG and IgM levels plateau within 6 days after seroconversion (9).

In conclusion, we provide evidence to support the unexpected early circulation of SARS-CoV-2 among persons who had visited China a few months prior to the pandemic declaration. These results support the emergence and spread of SARS-CoV-2 before the COVID-19 pandemic declaration. The detection of SARS-CoV-2 antibodies in individuals prior to the reported pandemic eruption in China could rewrite the currently accepted timeline of the pandemic.

Finally, we recommend that scientists in other countries consider analyzing and reporting the results of archived pre-pandemic samples to contribute to clarifying the timeline of the emergence and spread of SARS-CoV-2.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to wmahallawi@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, General Directorate of Health Affairs in Madinah (IRB No: H-03-M-084). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

The authors extend their appreciation to Taibah University, represented by the Deanship of Scientific Research, for funding this project (No. RC-442/4).

REFERENCES

1. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses*. (2020) 12:372. doi: 10.3390/V12040372
2. Stat G. *General Authority of Statistics. Kingdom of Saudi Arabia, Riyadh, Saudi Arabia*. Riyadh: General Authority for Statistics (2019).
3. Mahallawi WH, A. Serological Assay to Detect Human SARS-CoV-2 Antibodies. *J Taibah Univ Med Sci.* (2021) 16:57–62. doi: 10.1016/j.Jtumed.2020.11.011
4. Bryant JE, Azman AS, Ferrari MJ, Arnold BF, Boni MF, Boum Y, et al. Serology for SARS-CoV-2: Apprehensions, Opportunities, and the Path Forward. *Sci Immunol*. (2020) 5:Eabc6347. doi: 10.1126/Sciimmunol.Abc6347
5. Gehin JE, Klaasen RA, Norli ES, Warren DJ, Syversen SW, Goll GL, et al. Rheumatoid Factor and Falsely Elevated Results in Commercial Immunoassays: Data From an Early Arthritis Cohort. *Rheumatol Int*. (2021) 41:1657–65. doi: 10.1007/S00296-021-04865-9
6. Mahallawi WH, Al-Zalabani AH. The Seroprevalence of SARS-CoV-2 IgG Antibodies Among Asymptomatic Blood Donors in Saudi Arabia. *Saudi J Biol Sci.* (2020) 28:1697–701. doi: 10.1016/j.Sjbs.2020.12.009
7. Mahallawi WH, Ibrahim NA, Aljohani AS, Shaikh EA, Nafe RH, Khan AM, et al. Assessment of SARS-CoV-2 Anti-Spike IgG Antibody in Women and Children in Madinah, Saudi Arabia: A Single-Center Study. *Int J Environ Res Public Health*. (2021) 18:9971. doi: 10.3390/Ijerph18199971
8. Mahallawi WH. Humoral Immune Responses in Hospitalized COVID-19 Patients. *Saudi J Biol Sci*. (2021) 28:4055–61. doi: 10.1016/j.Sjbs.2021.04.032
9. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody Responses to SARS-CoV-2 in Patients With COVID-19. *Nat Med*. (2020) 26:845–8. doi: 10.1038/S41591-020-0897-1
10. Chau NVV, Nhan LNT, Nguyet LA, Tu NTK, Hong NTT, Man DNH, et al. Absence of SARS-CoV-2 Antibodies in pre-Pandemic Plasma From Children and Adults in Vietnam. *Int J Infect Dis*. (2021) 111:127–9. doi: 10.1016/j.Ijid.2021.07.072
11. Apolone G, Montomoli E, Manenti A, Boeri M, Sabia F, Hyseni I, et al. Unexpected Detection of SARS-CoV-2 Antibodies in the Prepandemic Period in Italy. *Tumori*. (2021) 107:446–51. doi: 10.1177/0300891620974755
12. Wang Q, Du Q, Guo B, Mu D, Lu X, Ma Q, et al. A Method to Prevent SARS-CoV-2 IgM False Positives in Gold Immunochromatography and Enzyme-Linked Immunosorbent Assays. *J Clin Microbiol*. (2020) 58:E00375–20. doi: 10.1128/JCM.00375-20
13. Riester E, Findeisen P, Hegel JK, Kabesch M, Ambrosch A, Rank CM, et al. Performance Evaluation of the Roche Elecsys Anti-SARS-CoV-2 S Immunoassay. *J Virol Methods*. (2021) 297:114271. doi: 10.1016/j.Jviromet.2021.114271
14. Jochum S, Kirste I, Hortsch S, Grunert VP, Legault H, Eichenlaub U, et al. Clinical Utility of Elecsys Anti-SARS-CoV-2 S Assay in COVID-19 Vaccination: An Exploratory Analysis of the MRNA-1273 Phase 1 Trial. *Front Immunol*. (2022) 12:798117. doi: 10.3389/Fimmu.2021.798117
15. Gao Y, Cai C, Grifoni A, Müller TR, Niessl J, Olofsson A, et al. Ancestral SARS-CoV-2-specific T cells cross-recognize the Omicron variant. *Nat Med*. (2022) 28:472–6.
16. Worobey M, Levy JI, Malpica Serrano LM, Crits-Christoph A, Pekar JE, Goldstein SA, et al. *The Huanan market was the epicenter of SARS-CoV-2 emergence*. (2022). Available online at: https://zenodo.org/record/6299600#.YrXfzHZBw2w
17. Pekar JE, Magee A, Parker E, Moshiri N, Izhikevich K, Havens JL, et al. *SARS-CoV-2 emergence very likely resulted from at least two zoonotic events*. (2022). Available online at: https://zenodo.org/record/6291628#.YrXxsHZBw2w
18. He W-T, Hou X, Zhao J, Sun J, He H, Si W, et al. Virome Characterization of Game Animals in China Reveals a Spectrum of Emerging Pathogens. *Cell*. (2022) 185:1117–1129.e8. doi: 10.1101/2021.11.10.467646
19. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The Proximal Origin of SARS-CoV-2. *Nat Med*. (2020) 26:450–2. doi: 10.1038/S41591-020-0820-9

20. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, et al. The Origins of SARS-CoV-2: A Critical Review. *Cell.* (2021) 184:4848–56. doi: 10.1016/j.Cell.2021.08.017
21. Johansson MA, Quandelacy TM, Kada S, Prasad PV, Steele M, Brooks JT, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Netw Open*. (2021) 4:E2035057-E. doi: 10.1001/Jamanetworkopen.2020.35057
22. Kharlamova N, Dunn N, Bedri SK, Jerling S, Almgren M, Faustini F, et al. SARS-CoV-2 Serological Tests can Generate False Positive Results for Samples From Patients With Chronic Inflammatory Diseases. *MedRxiv [Preprint]*. (2020). doi: 10.1101/2020.11.13.20231076
23. Pekar JE, Magee A, Parker E, Moshiri N, Izhikevich K, Havens JL, et al. *SARS-CoV-2 Emergence Very Likely Resulted From at Least Two Zoonotic Events* (2022). Available online at: https://zenodo.org/record/6342616/export/hx#.YrXa23ZBw2w
24. Day M. Covid-19: four Fifths of Cases Are Asymptomatic, China Figures Indicate. *BMJ.* (2020) 369:M1375. doi: 10.1136/bmj.M1375
25. Ma Q, Liu J, Liu Q, Kang L, Liu R, Jing W, et al. Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-Analysis. *JAMA Network Open*. (2021) 4:E2137257-E. doi: 10.1001/Jamanetworkopen.2021.37257
26. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody Responses to SARS-CoV-2 in Patients With Novel Coronavirus Disease 2019. *Clin Infect Dis.* (2020) 71:2027–34. doi: 10.1093/cid/Ciaa344
27. Long QX, Tang XJ, Shi QL Li Q, Deng HJ, Yuan J, et al. Clinical and Immunological Assessment of Asymptomatic SARS-CoV-2 Infections. *Nat Med.* (2020) 26:1200–4. doi: 10.1038/S41591-020-0965-6
28. Chia WN, Zhu F, Ong SWX, Young BE, Fong SW, Le Bert N, et al. Dynamics of SARS-CoV-2 Neutralising Antibody Responses and Duration of Immunity: A Longitudinal Study. *Lancet Microbe.* (2021) 2:E240–e9. doi: 10.1016/S2666-5247(21)00025-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Immunology

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/fimmu.2022.907481

Mining of Marburg Virus Proteome for Designing an Epitope-Based Vaccine

Mohamed A. Soltan[1], Waleed K. Abdulsahib[2], Mahmoud Amer[3], Ahmed M. Refaat[4], Alaa A. Bagalagel[5], Reem M. Diri[5], Sarah Albogami[6], Eman Fayad[6], Refaat A. Eid[7], Sherin M. A. Sharaf[8], Sameh S. Elhady[9], Khaled M. Darwish[10] and Muhammad Alaa Eldeen[11*]*

[1] Department of Microbiology and Immunology, Faculty of Pharmacy, Sinai University, Ismailia, Egypt, [2] Department of pharmacology and Toxicology, College of Pharmacy, Al- Farahidi University, Baghdad, Iraq, [3] Internal Medicine Department, Faculty of Medicine, Zagazig University, Zagazig, Egypt, [4] Zoology Department, Faculty of Science, Minia University, El-Minia, Egypt, [5] Department of Pharmacy Practice, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, [6] Department of Biotechnology, College of Science, Taif University, Taif, Saudi Arabia, [7] Department of Pathology, College of Medicine, King Khalid University, Abha, Saudi Arabia, [8] Department of Microbiology, Egyptian Drug Authority (EDA), Giza, Egypt, [9] Department of Natural Products, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, [10] Department of Medicinal Chemistry, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt, [11] Cell Biology, Histology and Genetics Division, Zoology Department, Faculty of Science, Zagazig University, Zagazig, Egypt

OPEN ACCESS

Edited by:
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

Reviewed by:
Talha Bin Emran,
Begum Gulchemonara Trust University, Bangladesh
Felipe Campelo,
Aston University, United Kingdom
Amir Atapour,
Shiraz University of Medical Sciences, Iran

***Correspondence:**
Mohamed A. Soltan
mohamed.mohamed@su.edu.eg
Muhammad Alaa Eldeen
dr.muhammadalaa@gmail.com

Specialty section:
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

Received: 29 March 2022
Accepted: 16 June 2022
Published: 15 July 2022

Citation:
Soltan MA, Abdulsahib WK, Amer M, Refaat AM, Bagalagel AA, Diri RM, Albogami S, Fayad E, Eid RA, Sharaf SMA, Elhady SS, Darwish KM and Eldeen MA (2022) Mining of Marburg Virus Proteome for Designing an Epitope-Based Vaccine. Front. Immunol. 13:907481. doi: 10.3389/fimmu.2022.907481

Marburg virus (MARV) is one of the most harmful zoonotic viruses with deadly effects on both humans and nonhuman primates. Because of its severe outbreaks with a high rate of fatality, the world health organization put it as a risk group 4 pathogen and focused on the urgent need for the development of effective solutions against that virus. However, up to date, there is no effective vaccine against MARV in the market. In the current study, the complete proteome of MARV (seven proteins) was analyzed for the antigenicity score and the virulence or physiological role of each protein where we nominated envelope glycoprotein (Gp), Transcriptional activator (VP30), and membrane-associated protein (VP24) as the candidates for epitope prediction. Following that, a vaccine construct was designed based on CTL, HTL, and BCL epitopes of the selected protein candidates and to finalize the vaccine construct, several amino acid linkers, β-defensin adjuvant, and PADRE peptides were incorporated. The generated potential vaccine was assessed computationally for several properties such as antigenicity, allergenicity, stability, and other structural features where the outcomes of these assessments nominated this potential vaccine to be validated for its binding affinity with two molecular targets TLR-8 and TLR-4. The binding score and the stability of the vaccine-receptor complex, which was deeply studied through molecular docking-coupled dynamics simulation, supported the selection of our designed vaccine as a putative solution for MARV that should be validated through future wet-lab experiments. Here, we describe the computational approach for designing and analysis of this potential vaccine.

Keywords: Marburg virus, immunoinformatics, epitope mapping, multitope vaccine, health care

INTRODUCTION

Marburg virus (MARV) is a notorious pathogen that belongs to *Filoviridae* family. It was first discovered in West Germany in the year of 1967 (1). It was found that the Egyptian fruit bat (*Rousettus aegyptiacus*) acts as a reservoir for MARV where primary human infection occurs through exposure to the viral reservoir, then transmission between people occurs *via* body fluids (2). MARV pathogenesis investigation showed that it causes life-threatening hemorrhagic fever which was very similar to that of the Ebola virus (EBOV) as both fevers are characterized by a severe inflammatory reaction in addition to systemic hemorrhaging (3). Since its early discovery, MARV has shown several successive waves of outbreaks. Outbreaks with a large number of cases were shown in Congo between 1998 and 2000 and in Angola between 2004 and 2005 with fatality rates of 83% (4) and 90% (5) respectively. The last few years showed continuous outbreaks of MARV where Uganda (6) and Guinea suffered from these outbreaks with a 100% fatality rate.

The MARV proteome and genome analysis showed seven basic proteins in addition to a negative-stranded linear RNA genome that sized approximately 19 kb (7). Most MARV proteins are multifunctional with major roles in viral replication and pathogenesis. The transcriptional activator VP30 was found to be involved in nucleocapsid maturation (8). The polymerase cofactor VP35 is an essential cofactor in the process of viral replication (9). Envelope glycoprotein has a major role in the interaction with the cellular receptors of the infected host (10). Membrane-associated protein VP24 plays a significant role in nucleocapsid and viral matrix formation (11). MARV nucleoprotein encapsidates the viral genome by oligomerization (12) while RNA-directed RNA polymerase is responsible for replicating this genome. Finally, matrix protein VP40 regulates the process of virion assembly and budding from infected cells (13).

The world health organization has put MARV on a list of an urgent need to find a solution for this deadly virus where the severe outbreaks and the high fatality rate have increased the importance of this call. Consequently, several techniques including vaccines based on viral vectors such as vesicular stomatitis virus (14), Adenovirus vectored vaccines (15), DNA plasmid vaccine (16), virus-like particles composed of several MARV proteins (17), and recombinant vaccine (18) have been adopted to generate an effective vaccine against this deadly virus. However, up to date, there is no approved vaccine or drug against MARV in the market.

In the last few years, the scientific community witnessed major development in the fields of bioinformatics, structural biology, and computational tools that were designed for the analysis of the growing data of several organisms' genomes. More recently, a new field that studies the immunological data and the tools that were developed to handle these data has been named immunoinformatics (19). This novel approach was applied to develop a vaccine construct against many pathogens, starting from bacteria such as *Moraxella catarrhalis* (20) and *Escherichia coli* (21) to viruses such as Nipah virus (22) and fungi such as *Candida auris* (23) and Mucormycosis causing fungi (24). Compared to classical and single-epitope vaccines, multi-epitope vaccines have unique features in their design as they consist of multiple MHC-restricted epitopes in addition to B-cell epitopes that can be recognized by TCRs of multiple clones from various T and B cell subsets. Consequently, strong cellular and humoral immune responses can be generated simultaneously. Moreover, multi-epitope vaccines incorporate some components with adjuvant capacity that can enhance the immunogenicity and long-lasting immune responses and reduce unwanted components that can trigger either pathological immune responses or adverse effects (25). Due to these several advantages, many trials have been performed to explore the efficacy of this new form of vaccine, where a significant activation at humoral and cellular arms of the immune system was observed against several tested pathogens such as *E. coli* (26), *Salmonella Typhimurium* and *Shigella flexneri* (27), and HIV-1 infection (28).

In the current study, we applied the immunoinformatics approach to generate a potential vaccine construct against MARV. The viral whole proteome was firstly analyzed for antigenicity and virulence of each protein then the candidate proteins were extracted for B and T cell epitope prediction. Finally, the top-ranking epitopes of each candidate were selected to construct a chimeric epitope potential vaccine that was assessed computationally for its structural, immunological, and chemical characteristics to be nominated as a putative solution against MARV.

MATERIALS AND METHODS

The flow of work of basic stages that were applied in the current study is illustrated in **Figure 1**.

Selecting Vaccine Candidates

UniProt database was used to retrieve Lake Victoria marburgvirus (strain Ravn-87) standard proteome which had the proteome Id (UP000008239). The analysis of the whole proteome was performed to identify proteins with an antigenicity score that exceeds 0.5 besides a necessary physiological role or virulent role for MARV to allow nominating them as the vaccine candidates of our study. Vaxigen v2.0 was used to calculate the antigenicity score (29). Upon using UniProt for MARV proteomes investigation, eleven reference proteomes were uploaded. The protein sequence of the selected candidates was extracted from these eleven proteomes in a process that was followed by running multiple sequence alignment, so that, the conservation of the epitopes to be selected later could be confirmed.

Predicting B Cell Epitopes

B cell epitopes have been predicted by bepipred linear epitope prediction method (30) which depends on a hidden Markov model approach in addition to a propensity scale method. The allergenicity and toxicity for the selected epitopes were predicted using AlgPred 2.0 and ToxinPred webservers respectively.

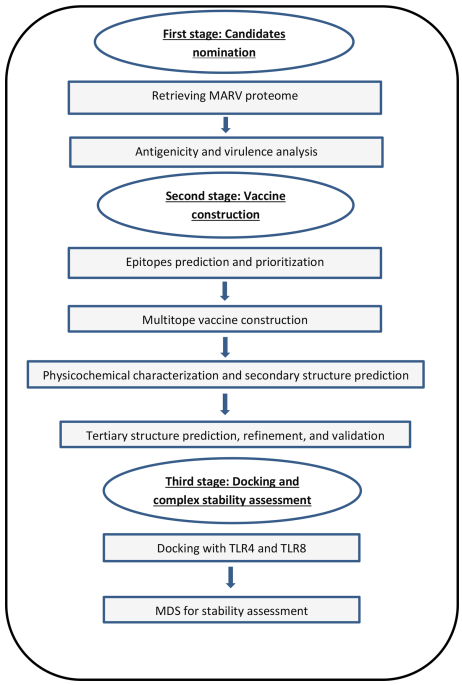


FIGURE 1 | A graph showing the applied strategy of the current study for designing and assessment of a potential vaccine against MARV.

Prediction of T Cell Epitopes

This stage started with the submission of filtered protein candidates to the Immune Epitope Database (IEBD) (31). MHC-I binding was predicted *via* NetMHCpan EL 4.0 prediction tool, the server recommended prediction method, and the binding prediction was performed against the reference set of HLA alleles as they represent the common binding specificities in addition to having population coverage of more than 97% (32). Regarding MHC-II binding prediction, IEBD gave a recommendation of 2.22 prediction method and again the binding prediction was run against the full HLA reference set that had population coverage of more than 99% (33). Then, the antigenic epitopes that acquired the best scores were analyzed to predict their ability to induct interferon-gamma *via* INF prediction server (34). In addition to that, epitopes' IL-4 induction ability was predicted through the webserver (https://webs.iiitd.edu.in/raghava/il4pred/), and IL-10 induction ability was estimated through the webserver (https://webs.iiitd.edu.in/raghava/il10pred/). Finally, the allergenicity and toxicity for the selected epitopes were predicted using AlgPred 2.0 and ToxinPred webservers respectively.

Molecular Docking of MHCI and MHCII Filtered Mono Epitopes

The final stage of epitopes prioritization before reaching the step of multitope vaccine construction was the analysis of conservancy through multiple sequence alignment and a second stage of binding assessment through molecular docking analysis. In order to run the docking study for the single epitopes, the epitope 3D structure was predicted using PEP FOLD 3 webserver (35). For MHC-II and MHC-I epitopes, crystal structures of HLA-DRB1*04:01 (PDB ID 5JLZ) and HLA-A*11:01 (PDB ID 6JP3) were selected respectively, as a molecular docking receptor. AutoDock Vina was used for Docking analysis (36).

Construction of Multitope Vaccine

Constructing the multitope vaccine was performed by merging four components. Firstly, the β-defensin adjuvant was introduced, after that, the previously selected epitopes were incorporated and linked with suitable amino acids linkers. Finally, the inclusion of PADRE sequence, that could improve the immune response for our designed vaccine (37) was performed. This final construct was assessed for several characteristics such as antigenicity *via* VaxiJen v 2.0 (29), allergenicity using AlgPred (38), toxicity through the ToxinPred server (39), and assessed for its human homology through BLASTp against human proteome.

Predicting Protein Solubility, Physicochemical Characters, and Secondary Structure

SOLpro server was used to predict the protein solubility upon overexpression in *Escherichia coli* (40). Several physicochemical characteristics of the vaccine were predicted including atomic composition, molecular weight, instability index, etc … using the ProtParam tool (41). Lastly, the secondary structure of the designed vaccine construct was predicted using the PSIPRED server (42).

Prediction and Validation of Vaccine 3D Structure

We employed 3Dpro web server for the prediction of the designed vaccine tertiary structure (43). This server depends on constructing several conformations and calculates the outcome to nominate a preferred model based on energy scores. Consequently, the selected prediction would acquire the highest stability and the lowest possible energy. The next step was the usage of the GalaxyRefine server for the refinement of the selected 3D model (44). To achieve this refinement process, the side chains were rebuilt and repacked succeeded by structure relaxation *via* molecular dynamics simulation. Finally, Ramachandran plot analysis (45) and ProSA (46) were employed to validate the predicted 3D model before and after the refinement process.

Docking Analysis between 3D Structure of Predicted Vaccine and Two Selected Targets (TLR-8 and TLR-4)

It is proposed that after multitope vaccine inoculation, it attaches to dendritic cells and macrophages through toll-like receptors (TLRs). Following that, the vaccine's epitopes are digested by antigen-presenting cells (APC) and presented to T cells (25). Hence, TLR-8 (PDB ID: 6KYA) and TLR-4 (PDB ID: 4G8A) was

brought from the protein data bank and submitted to the ClusPro 2.0 server (47) with the refined 3D structure of our vaccine, which represented the ligand, to perform a molecular docking analysis. Identifying the canonical epitope-TLR binding interactions were performed through a protein-protein interface and structural analysis using the PDBsum free web-based server (https://bio.tools/pdbsum_generate/) (48).

Normal-State Analyses *via* Torsional Coordinate-Association

For gaining more insights regarding the collective flexibilities/motion functions of the constructed multiepitope vaccine relative to its bounded TRLs, the iMODS on-line server was used (http://www.imods.chaconlab.org/) (49). This employed server is fast and accurate while being capable of assessing the collective protein complex (epitope-TLRs) motions based on normal-state analyses of their respective internal dihedral angle (torsional) coordinates (50). Furthermore, it can predict several parameters reflecting structural flexibility/deformation reflecting significant deviation from the normal distribution values obtained from thousands deposited reference sets. Within the PDB file of both epitope-TLR complexes, the atoms/residues were continuously indexed where the number ranges 1-5824 and 5825-11649 were assigned for the TLR-4 protomer A and B, respectively, (23-627 amino acids for each protomer), while as residue ranges 1-7459 and 7460-14919 were for TLR-8 protomers (32-818 amino acids for each protomer). Subsequent residue ranges 11650-14868 and 14919-18157 were assigned for the bounded multiepitope vaccine (successive 1-349 amino acids) at the TLR-4 and TLR-8 complexes, respectively.

Molecular Dynamics Simulation

Molecular dynamics simulation was perused for evaluating the relative epitope-TLR motions, conformation-time evolution, and thermodynamic stability under near-physiological conditions (51). Using GROMACS-5.1.4 and CHARMM36m forcefield, the constructed models of the multiepitope vaccine in complex with either TLR-4 or TLR-8 were individually simulated through 100 ns-all atom molecular dynamics as previously reported (24, 52). Protein complexes were solvated within TIP3P water and under periodic boundary conditions keeping the protein complex at 10 Å marginal distances from the box margins. Protein residues ionization states were assigned at physiological pH 7.4, while as the net charge of the entire constructed models was neutralized *via* sufficient chloride/potassium ions. Models were minimized for 0.05 ns under steepest-descent algorithm (53), and then equilibrated through two successive stages; initial NVT ensemble (Berendsen-temp couplings at 303.15 K for 1 ns) and then NPT ensemble (Parrinello-Rahmann barostat at 303.15 K and one atmospheric pressure for 1 ns). Production of molecular dynamics runs was proceeded for 100 ns under NPT ensemble and Particle-Mesh-Ewald algorithms for computing long-range electrostatic interaction, while as, LINCS was used for modeling covalent bond lengths. Both Coulomb's and van der Waals' non-bounded interactions were subjected to 10 Å truncations using Verlet cut-off model (54). Analysis of protein complexes were performed through estimating root-mean square deviations (RMSDs), RMS-fluctuations (RMSFs) relying on the obtained trajectory file analyses. The Molecular Mechanics/Poisson-Boltzmann (MM-PBSA) calculations, were used for estimating the free-binding energies of the simulated epitope-TLR complexes, in addition to, investigating the energy contributions of each constituting residues (55). Visual Molecular Dynamics (VMD V.1.9.3) software (University of Illinois, Urbana-Champaign, USA) was used for hydrogen bond analysis between multiepitope vaccine/TLR over the entire simulation periods. Cut-off values for hydrogen bond distance and angle were set at 3.0 Å and 20°, respectively. Conformational analysis and visualizing the simulated epitope-TLR complexes, within specified timeframes, were performed by PyMol (Schrödinger; V.2.0.6) software.

Immune Simulation of the Chimeric Peptide Vaccine

We employed C-ImmSim server to predict the stimulated immune response after the proposed vaccine injection. We visualized the estimated immune response after the administration of three multitope vaccine injections in four weeks intervals. The followed technique represents a prime-booster-booster approach to achieve a long-lasting immune response.

Reverse Translation and Codon Adaptation

Finally, we reached the step of codon adaptation analysis for the designed vaccine. It is an essential step to validate the sequence of the vaccine construct to be expressed in *E. coli* k-12 (the planned expression host in the future wet-lab experiments). For this purpose, JCAT server (56) was employed where the value of the codon adaptation index (CAI) calculated by the server would give a prediction for the suitability of the proposed vaccine sequence to be expressed in the selected host.

RESULTS

Screening Proteins for Nominating Vaccine Candidates

By applying the antigenicity score cutoff on the MARV whole proteome, three proteins were found to have an antigenicity score that exceeded 0.5. These proteins are envelope glycoprotein with an antigenicity score of 0.54, VP30 with an antigenicity score of 0.56 and VP24 with an antigenicity score of 0.55. This was accompanied by studying the function of these three candidates in order to identify those that had physiological roles or significant virulence, proteins VP30 and VP 24 played a role in the viral replication while envelope glycoprotein has a major function of attachment to its specific host cell receptors. As a result, these three proteins were selected as final targets of epitope prediction.

Predicting B Cell Epitopes

The B cell epitopes were predicted through IEBD webserver (**Supplementary Figure 1**) using a threshold value of 0.35. There were 10, 14, and 28 predicted epitopes for VP24, VP30, and

envelope glycoprotein respectively. Then, the epitopes that had a size between 8 to 18 peptides were selected which downsized our epitopes list (**Table 1**), After that, we planned to identify the top 6 epitopes of these 3 proteins (2 epitopes for each protein) based on the antigenicity score and conservancy analysis in order to incorporate them to the multitope vaccine construct. For VP24 protein, we found only one predicted B cell epitope sized between 8:18 and this epitope had an antigenicity score less than 0.4. As a result, we modified our plan and selected 3 B cell epitopes from the other proteins (envelope glycoprotein and VP30). It is worth mentioning that all B cell filtered epitopes were predicted to be non-allergenic ad non-toxic.

Predicting T Cell Epitopes

The prediction of MHC-I epitopes resulted in 13203, 14715, and 36315 epitopes for VP24, VP30, and envelope glycoprotein respectively with a percentile rank ranging from 0.01 to 100. We selected the epitopes that showed a small percentile rank (as this small percentile rank indicates good binding properties for the epitopes) and a large antigenicity score. **Table 2** shows the top five epitopes identified for each protein. Meanwhile, the prediction of MHC-II epitopes resulted in 6453, 7209, and 18009 predictions for VP24, VP30, and envelope glycoprotein respectively. Again top-ranking epitopes are the ones that demonstrated a small percentile rank and a large antigenicity score. These top-ranked epitopes were assessed for their capability to induce INF-γ, IL-4, and IL-10. **Table 3** shows the top five epitopes identified for each protein.

Molecular Docking for T Cell Epitopes

We selected HLA-A*11:01 and HLA-DRB1*04:01 as representative alleles to analyze the binding affinity of the filtered MHC-I and MHC-II epitopes respectively. **Figures 2**, **3** demonstrate the docked complexes of MHC-I and MHC-II epitopes respectively while **Table 4** displays the binding energy of these complexes. The binding energy scores for both types of peptides ranged between -7.2 and -9.1 (**Table 4**), and to validate these scores we investigated each of the mentioned receptors which were deposited in the protein databank with an attached ligand. we employed these ligands to act as control by removing and re-docking to their respective receptor using the same conditions of predicted epitopes docking. The docking score for these controls were -6.3 and -7.7 for HLA-A*11:01 and HLA-DRB1*04:01, respectively (the docking scores of filtered peptides were more negative than the control; therefore, they were estimated to be good binders).

Construction of Multitope Vaccine, Assessment of Physicochemical Characteristics and Prediction of Secondary Structure

After selecting the most promising B and T cell epitopes from VP24, VP30, and envelope glycoproteins, we moved to design the multitope vaccine depending on six CTL and six HTL epitopes (two epitopes for each protein) in addition to six BCL epitopes (three from VP30 and three from envelope glycoprotein), these epitopes were linked together using GGGS, GPGPG, and KK linkers respectively. In order to finalize the construct, we incorporated PADRE peptide sequence and beta-defensin adjuvant to reach the final sequence of our multitope vaccine which was composed of 349 amino acids and sequenced as the following:

"EAAAKGIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKC
STRGRKCCRRKKEAAAKAKFVAAWTLKAAAGGGSIITR
VNMGFGGGSTVKWGNFIFGGGSSSISVQASYGGGSSVQ
ASYDHFGGGSRTFSLINRHGGGSTTRPPIYFRGPGPGE
WLLLEVTSAIHISPGPGPGSEWLLLEVTSAIHISGPGPG
TRPPIYFRKKRSIFWGPGPGTIYFLISLILIQSIKGPGPGTN
RELLLLMARKMLPGPGPGLDNLTNRELLLLMARKKNLS
KPPPPPKDMCKKPCTDPACNRDHDLDKKASNSQPQDV
DSVKKTGVPPKNVEYTEGEEAKKKAKFVAAWTLKAAA
GGGS"

This was followed by analyzing this final design for allergenicity, the SVM method that depends on amino acid composition predicted our designed vaccine to be non-allergen, Moreover, toxicity analysis predicted our designed vaccine to be non-toxic. Furthermore, the analysis of the antigenicity predicted our multitope construct to be antigenic with an antigenicity score of 0.61. The final design was found to have a SOLpro score of 0.993 indicating it to be soluble upon overexpression (a score that exceeds 0.5 indicating solubility upon overexpression) and did not significantly resemble human protein sequences when analyzed through Blastp (hence, the predicted vaccines would not elicit autoimmune reactions in

TABLE 1 | Predicted B cell epitopes from VP30 and envelope glycoprotein.

VP30			Envelope glycoprotein		
Epitope	**Start-End**	**Antigenicity Score**	**Epitope**	**Start-End**	**Antigenicity Score**
MQQPRGRSRNRS	1-12	1.11	ASNSQPQDVDSV	25-36	0.85
NLSKPPPPPKDMC	66-78	0.42	QKVADSPLEAS	57-67	0.63
PCTDPACNRDHDLD	86-99	0.54	TGVPPKNVEYTEGEEAK	74-90	1.15
NLPQDQNGVI	205-214	-0.47	PSNIRDYPKC	110-119	0.92
			KYWTSSNETQRNDT	196-209	0.025
			VTDPSGKSLLLD	97-108	-0.029
			SGSGSGEQGPHT	269-280	0.48
			EQKQSSTIL	287-295	0.66
			DKIRKDEQKEETGWGL	625-640	0.17

TABLE 2 | Top-ranked T-cell epitopes (MHC-I peptides) of VP24, VP30, and envelope glycoprotein.

No	Protein	Epitope	Antigenicity score	Allergenicity	Toxicity
1	VP24	KPSSIEIKL	1.77	Non-allergenic	Non-toxic
2	VP24	TVKWGNFIF	1.6	Non-allergenic	Non-toxic
3	VP24	IITRVNMGF	1.64	Non-allergenic	Non-toxic
4	VP24	NITEKSINL	1.62	Non-allergenic	Non-toxic
5	VP24	HISPNLLGI	1.52	Non-allergenic	Non-toxic
6	VP30	SSISVQASY	1.27	Non-allergenic	Non-toxic
7	VP30	QLPSKPQYI	0.86	Non-allergenic	Non-toxic
8	VP30	SVQASYDHF	0.82	Non-allergenic	Non-toxic
9	VP30	ENQLPSKPQY	0.67	Non-allergenic	Non-toxic
10	VP30	RSHQVALSTY	0.54	Non-allergenic	Non-toxic
11	Envelope glycoprotein	HTPPNISLTF	1.7	Non-allergenic	Non-toxic
12	Envelope glycoprotein	RTFSLINRH	1.16	Non-allergenic	Non-toxic
13	Envelope glycoprotein	EQHTPPNISL	1.13	Non-allergenic	Non-toxic
14	Envelope glycoprotein	GCFGILQEY	1.11	Non-allergenic	Non-toxic
15	Envelope glycoprotein	TTRPPIYFR	1.1	Non-allergenic	Non-toxic

TABLE 3 | Top-ranked T-cell epitopes (MHC-II peptides) of VP24, VP30, and envelope glycoprotein.

No	Protein	Epitope	Antigenicity	IFN-γ inducer	IL4 inducer	IL10 inducer	Antigenecity	Toxicity
1	VP24	EWLLLEVTSAIHISP	1.28	Yes	Yes	Yes	Non-allergenic	Non-toxic
2	VP24	PFLALRILLGVALKD	1.13	Yes	No	No	Non-allergenic	Non-toxic
3	VP24	FLALRILLGVALKDQ	1.12	Yes	No	No	Non-allergenic	Non-toxic
4	VP24	SEWLLLEVTSAIHIS	0.93	Yes	Yes	Yes	Non-allergenic	Non-toxic
5	VP24	EPFLALRILLGVALK	0.91	Yes	No	No	Non-allergenic	Non-toxic
6	VP30	TSLRAALSLTCAGIR	1.09	Yes	No	No	Non-allergenic	Non-toxic
7	VP30	TNRELLLLMARKMLP	0.7	Yes	Yes	No	Non-allergenic	Non-toxic
8	VP30	TCAGIRKTNRSLINT	0.89	Yes	No	No	Non-allergenic	Non-toxic
9	VP30	CAGIRKTNRSLINTM	0.74	Yes	No	No	Non-allergenic	Non-toxic
10	VP30	LDNLTNRELLLLMAR	0.73	Yes	Yes	Yes	Non-allergenic	Non-toxic
11	Envelope glycoprotein	TRPPIYFRKKRSIFW	1.4	Yes	No	Yes	Non-allergenic	Non-toxic
12	Envelope glycoprotein	TIYFLISLILIQSIK	0.59	No	Yes	Yes	Non-allergenic	Non-toxic
13	Envelope glycoprotein	KRSIFWKEGDIFPFL	0.62	Yes	Yes	No	Non-allergenic	Non-toxic
14	Envelope glycoprotein	IYFRKKRSIFWKEGD	1.02	Yes	Yes	No	Non-allergenic	Non-toxic
15	Envelope glycoprotein	PPIYFRKKRSIFWKE	1.22	Yes	Yes	No	Non-allergenic	Non-toxic

the host). In addition, ProtParam online tools were used to further analyze the other physicochemical characteristics of this multitope vaccine as shown in **Table S1**. Finally, the secondary structure assessment predicted 23.2% helix, 26.1% strand, and 50.7% coil of the vaccine construct secondary structure (**Supplementary Figure 2**).

Tertiary Structure Prediction, Validation and Refinement

Firstly, the 3D model of the proposed vaccine was predicted by 3Dpro webserver. Following that, the predicted model was validated before and after structure refinement on GlaxyRefine webserver. In order to perform this structural validation, we run a Ramachandran plot analysis and calculated the Z-score for the models. Regarding the primary structure, 88.5%, 11.1%, and 0.4% of residues were located in favored, allowed, and outlier regions, respectively with a Z-score of -3.68. Moving to the refined model, 93.3%, 6.3%, and 0.4% of residues were located in favored, allowed and outlier regions, respectively and the Z-score was -4.29. The refined model and its structural validation scores are shown in **supplementary Figure 3**.

Molecular Docking of our Vaccine With TLR4 and TLR8

ClusPro 2.0 server was employed to validate the binding process between our designed potential vaccine and its relevant receptors. The predicted binding energy values were -1458 and -1386 kCal/mol for TLR-4 and TLR-8, respectively. In order to validate these values, *Brucella* Lumazine Synthase, which acts as an agonist for TLR-4 (57), was docked with the same server to TLR-4 and the smallest generated docking score was -1178.5. Based on that, the binding energy of the currently designed vaccine on TLR-4 was smaller than that of the control; therefore, a good binding is predicted for the multitope vaccine. For both investigated dimeric TLRs, the docked ligand epitope exhibited preferential binding to one TLR monomer unit over the other. The overall conformation of the epitope/TLR-4 complex showed a transverse orientation of its multiepitope vaccine across the binding cavity of TLR-4 protomer B (**Figure 4A**). Notably, the backside *N*-terminal of the docked vaccine depicted contacts with the target's surface interface of the other monomeric unit (TLR-4 protomer A). The latter depicted 1773 $Å^2$ and 774 $Å^2$ interface area for the epitope in regard to 1709 $Å^2$ and 748 $Å^2$ for

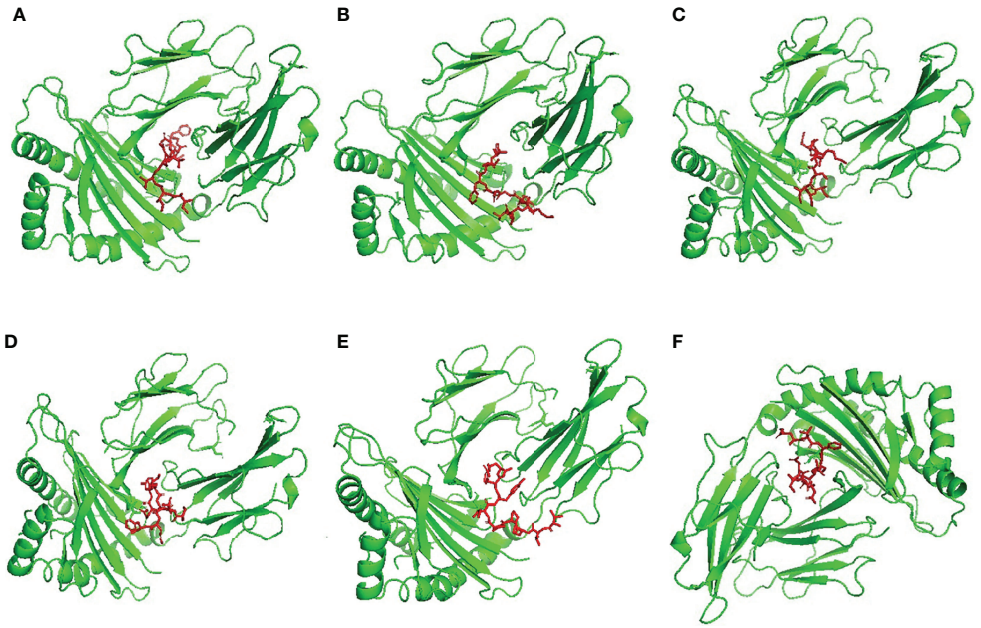


FIGURE 2 | Structural positions of MHC-I epitopes (red color) in 3-dimensional structure of HLA-A*11:01 (green color), structures **(A–F)** are for epitopes number 1,2,3,4,5 and 6 respectively from **Table 4**.

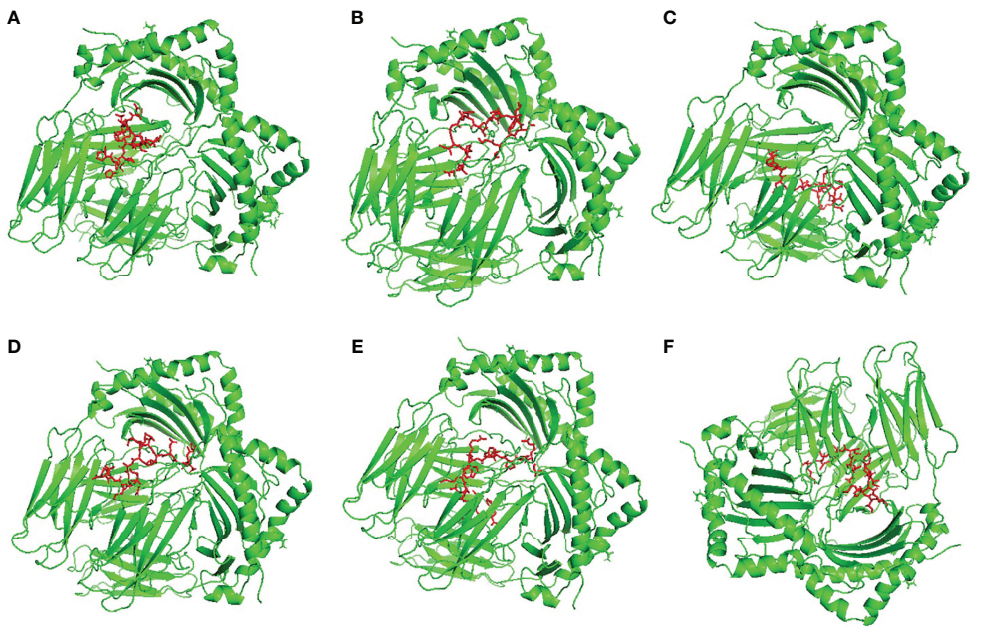


FIGURE 3 | Structural positions of MHC-II epitopes (red color) in 3-dimensional structure of HLA-DRB1*04:01 (green color), structures **(A–F)** are for epitopes number 1,2,3,4,5 and 6 respectively from **Table 4**.

TLR-4 protomer B and protomer A, respectively. Concerning the epitope/TLR-8 docked complex, a differential ligand orientation was depicted in relation to the binding pocket of the TLR-8 target. The *N*-terminal of the docked epitope showed relevant insertion into the TRL-8 protomer B binding pocket with depicted proximity towards two target's loops (260-268 and 433-482 residue ranges) protruding inside the binding cavity (**Figure 4B**). The inserted epitope's section reached the

TABLE 4 | The binding energy of T cell epitopes with their respective allele.

No.	Epitope	MHC-I Allele	Binding Energy (kcal/mol)	Epitope	MHC-II Allele	Binding Energy (kcal/mol)
1	IITRVNMGF		-8.4	EWLLLEVTSAIHISP		-7.5
2	TVKWGNFIF		-8.4	SEWLLLEVTSAIHIS		-7.4
3	SSISVQASY	HLA-A*11:01	-8.1	LDNLTNRELLLLMAR	HLA-DRB1*04:01	-7.4
4	SVQASYDHF		-8.4	TNRELLLLMARKMLP		-7.5
5	TTRPPIYFR		-9.1	TRPPIYFRKKRSIFW		-8.7
6	RTFSLINRH		-8.0	TIYFLISLILIQSIK		-7.2

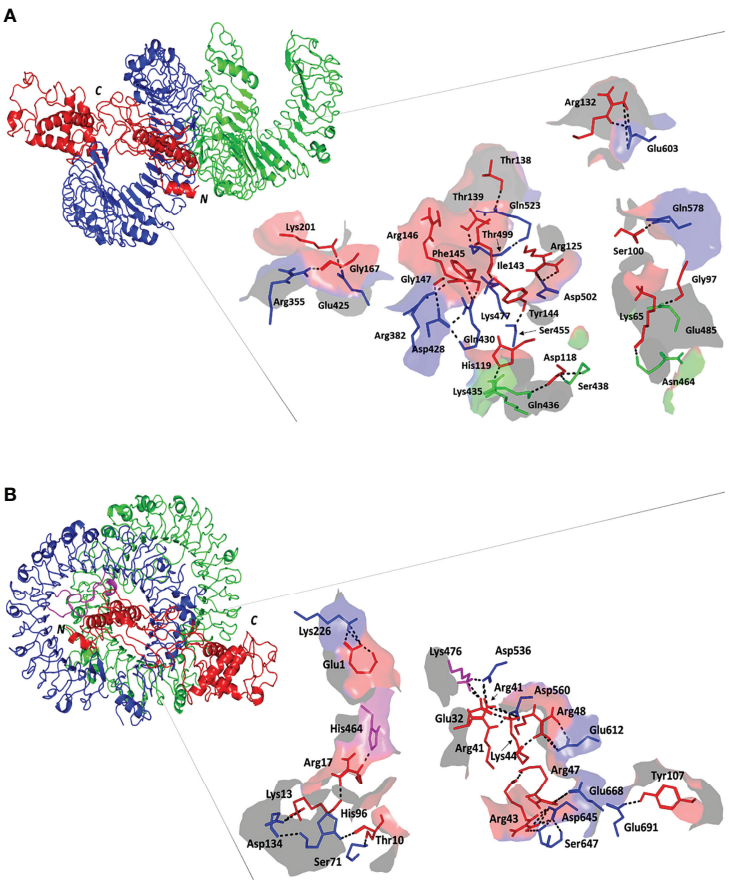


FIGURE 4 | Docked complex and interface binding interactions for the designed vaccine and the two investigated TLRs; **(A)** multiepitope vaccine/TLR-4 and **(B)** multiepitope vaccine/TLR-8 complexes. 3D cartoon representation of the docked vaccine (red) at the TLR binding site (green and blue for protomer A and B, respectively). The protruding loops into the TLR-8 binding site are colored magenta. The carboxy and amine terminals of the simulated epitopes are assigned with *C* and *N*-letters, respectively. Zoomed images illustrate the surface representation of multiepitope vaccine/TLR binding polar interactions showing the residue pairs as lines, colored depending on their location within the proteins, and numbered according to their respective residue sequence. The bonds of polar interactions are shown as black dashed lines.

dimerization interface at other TLR-8 monomeric unit (protomer A) since the latter is at a slightly drifted parallel orientation with TLR-8 protomer B. The rest of the epitope structure including its middle and carboxy-terminal rested at the surface of the target's protomer B. Such ligand's conformation/orientation correlated to lower interface surface areas (1677:1478 $Å^2$ for epitope: protomer B and 256:234 $Å^2$ for epitope: protomer A) as compared to the TLR-4 system. However, the respective TLR-8 protomer B and A interface areas were higher for the earlier protomer within a similar differential fashion to those depicted for the epitope/TLR-4 model.

Evaluation of the nature of epitope/TLR binding interactions showed interesting findings. Higher polar interaction patterns were depicted with the epitope/TLR-4 interface as compared to the TLR-8 complex. A total of three salt bridges and 21 hydrogen bond pairs were assigned for the docked epitope and TLR-4

binding site (**Figure 4A**). Depicted salt bridges were confined with TLR-4 protomer B between the residue pairs; Arg125-Asp502, Arg132-Glu603, and Lys201-Glu425 for the docked epitope and target site, respectively. Corresponding to the above-described protein-protein interface areas, higher hydrogen bonds, and other non-bonded contacts were assigned for protomer B (15 and 144) in regard to protomer A (15 and 144 versus 6 and 73, respectively). Moving towards the epitope/TLR-8 predicted complex, lower extent of binding interactions were illustrated through the protein-protein interface analysis. Nine salt bridges and 17 hydrogen bonds were depicted between epitope and TLR-8 protomer B, yet no polar interaction was illustrated at the epitope/protomer A interface (**Figure 4B**). Salt bridges were between the residue pairs; Glu1-Lys226, Lys13-Asp134, Glu32-Lys476, Arg41-Asp536, Arg41-Asp560, Arg43-Asp645, Lys44-Asp560, Arg47-Glu668, and Arg48-Glu612 for epitope/TLR-8 protomer B, respectively. On the other hand, limited non-bounded contacts (135 and 7) were illustrated at the interface between docked epitope and TLR-8 protomer B and A, respectively. The detailed atom-atom interactions and bonding distances, as well as representative diagrams for the depicted epitope/TLR protomer interfaces, are thoroughly described in **Tables S2, S3**.

Normal-State Analyses *via* iMODS Server

The inherited stability and conformational mobility of both the docked vaccine and TLRs were analyzed based on the torsion angle-related normal state analysis using the iMODS server. Interestingly, the estimated B-factors were significantly higher for the docked vaccine at both TLR models with higher values being associated with the ligand's *C*-terminal residues (**Figures 5A1, B1**). On the other hand, the target TLR-4 protomers showed higher inherited flexibility in relation to those of the TLR-8 target. Generally, the B-factor correlates to the relative magnitude of atom displacement around conformational equilibria. Findings of B-factor analyses were recapitulated by the obtained complex deformability index presented in **Figures 5A2, B2** where higher individual distortions were assigned for vaccine amino acids, particularly at terminal residues, as compared to those of the bound TLR target proteins. The estimated eigenvalues representing the motion stiffness of each vaccine/TLR complex were 7.02 x 10^{-06} and 7.16 x 10^{-06}, for TLR-4 and TLR-8 complexes, respectively. These values are inversely proportional to variance predicting the significantly higher mobility of the vaccine as compared to the TLR ones across collective functional motions (**Figures 5A3,4, B3,4**). The server provided the covariance matrix illustrating the coupled residue pairs demonstrating uncorrelated, correlated, or anti-correlated motions as white, red, and blue colors, respectively. Both docked vaccines had less anti-correlated motions as well as higher correlated residue-pair motions than those of the bound TLRs (**Figures 5A5, B5**). Finally, the obtained elastic-network model explains the differential flexibility patterns among both the vaccine and bound TLR (**Figures 5A6, B6**). The elastic-network model illustrates the atom pairs linked *via* springs based on stiffness degree between them relying on different color representations. Typically, stiffer strings correlate to dark gray colors. Along the normal distribution of stiffer string, the docked vaccines showed discontinuous dark-gray bands where it was of more discontinuous strings within the TLR-8 model. On the contrary, the target TLR residues more continuous gray bands around the same immobility normal string.

Analysis of Molecular Dynamics Simulation Runs

The simulated multiepitope vaccine models showed differential thermodynamic stability profiles in regard to the kind of the bounded TLR target. The alpha-carbon RMSD trajectories (*Cα*

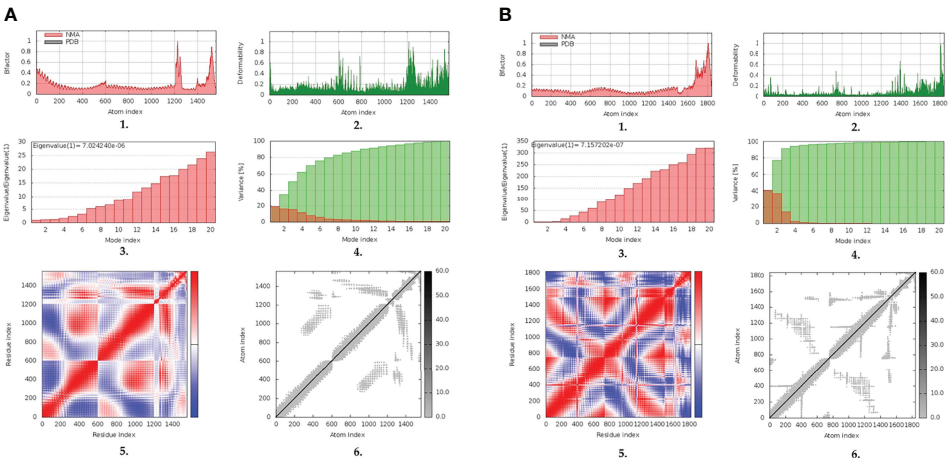


FIGURE 5 | Normal-state analyses *via* iMODS server for the docked multitope vaccine/TLR complexes; **(A)** multitope vaccine/TLR-4; **(B)** multitope vaccine/TLR-8 models. Ligand-receptor interaction was assessed throughout comparative **1.** B-factor indices, **2.** deformability, **3.** variances, **4.** eigenvalues, **5.** covariance of residue indices, and **6.** elastic network analyses.

RMSDs) for the simulated epitope/TLR-4 model showed typical dynamic behaviors (**Figure 6A**). Throughout the initial simulation frames, the *Cα* RMSD values increased gradually owing to the release of the previously applied constraints at the minimization and equilibration stages. Relaxation of the system model was proceeded till around the first 20 ns where afterward the protein *Cα* RMSDs attained respective equilibration plateau till the end of the runs (100 ns). Rapid equilibration and steady *Cα* RMSDs were maintained for more than half the simulation runs (> 70 ns) regarding the epitope, TLR-4, and their respective combined complex. Across each respective equilibration plateau, the epitope's average *Cα* RMSD values were higher than those of the corresponding bound TLR-4 protein target (24.27 ± 1.11 Å versus 8.73 ± 0.71 Å). Regarding the other simulated model, epitope/TLR-8, significant *Cα* RMSD fluctuations were depicted for both the simulated epitope and bound target (**Figure 6B**). Fluctuations were much more profound around the 30-40 ns and within initial simulation timelines. Nevertheless, near equilibration plateaus were attained following the late simulation times (beyond 65 ns) showing limited fluctuations. Notably, much higher *Cα* RMSD fluctuations were depicted for the ligand epitope as compared to its bound TLR-8 target (average 26.24 ± 2.09 Å versus 8.71 ± 1.42 Å) following respective thermodynamic equilibrations. Interestingly, both the ligand and target proteins were converged down to *Cα* RMSD values being comparable to their corresponding protein within the other epitope/TLR model near the end of the molecular dynamics simulation (100 ns).

For gaining more insights regarding the simulated protein global stability and residue-wise flexibility contributions, the *Cα* RMSF trajectories were monitored across the entire 100 ns simulation runs (**Figure 7**). Simulated proteins of both systems showed typical dynamic behaviors where terminal residues, as well as their vicinal amino acids, depicted higher motion patterns (elevated *Cα* RMSFs) in regard to their central core ones. The latter *Cα* RMSF fluctuations were of higher values for the simulated epitopes in regards to their bounded TLR protein targets. Interestingly, both TLR-4 protomers were assigned with comparatively lower mobility/fluctuation profiles as compared to those of TLR-8 ones (3.62 ± 0.80 Å versus 4.64 ± 1.32 Å). Similarly, higher stability/immobility patterns were also depicted for the TLR-4 bounded epitope in relation to that of the TLR-8 model (6.44 ± 3.10 Å versus 10.91 ± 4.94 Å). It is worth mentioning that the simulated epitopes of both models showed the most recognized flexibility/mobility trends for their respective *C*-terminal residues as well as vicinal amino acids (high numbered amino acid sequencing).

Subsequent conformational analysis was proceeded to grasp the main time-evolution conformational changes within the different epitope/TLR models. This was performed through comparative conformational analysis of simulated models at the initial and final dynamic trajectories (58). Extracted frames (0 ns and 100 ns) were subsequently minimized down to $1x10^{-3}$ Kcal/mol.A^2 gradient *via* MOE2019.01 package before being visually analyzed *via* PyMol software. Interestingly, the simulated epitope managed to be confined within the binding sites of both investigated TRLs, while exhibiting stable binding complex states (**Figure 8**). More significant conformational alterations were depicted with the simulated epitopes in comparison to their respective bound TLR targets. The latter epitope alterations allowed these ligands to adopt more compacted conformations with orientations being directed towards the TLR lateral interfaces and binding sites. It is worth noting that the higher comparative dynamic alterations were illustrated for the *C*-terminal region of the simulated epitopes as compared to their amine ends. Regarding both simulated epitope/TLR-4 versus TLR-8 models, the earlier depicted less profound dynamic conformational alteration across the simulated runs. This was highly observed within the multiepitope vaccine/TLR-4 model where lower aligned *Cα* RMSD values between respective 0 ns and 100 ns trajectories were depicted (5.590 Å and 9.923 Å, for multiepitope vaccine/TLR-4 and TLR-8 models, respectively).

Evaluation of the total ligand/TLR free-binding energies (ΔG_{total}) showed significant affinity of the designed vaccine with higher preferentiality (higher negative values) towards TLR-4 as well as its protomer B in regard to TLR-8 and cognate protein, respectively (**Table 5**). For identifying the nature of vaccine/TLR binding, the obtained total free-binding energy was dissected into its constituting energy terms in regards to the hydrophobic van der Waal potentials ($\Delta G_{van\ der\ Waal}$), Coulomb's electrostatic energies ($\Delta G_{electrostatic}$), polar solvation ($\Delta G_{solvation}$), and non-polar solvation (ΔG_{SASA}) energy contributions. Notably, the $\Delta G_{electrostatic}$ within both vaccine/TLR models depicted superior energy contribution for the

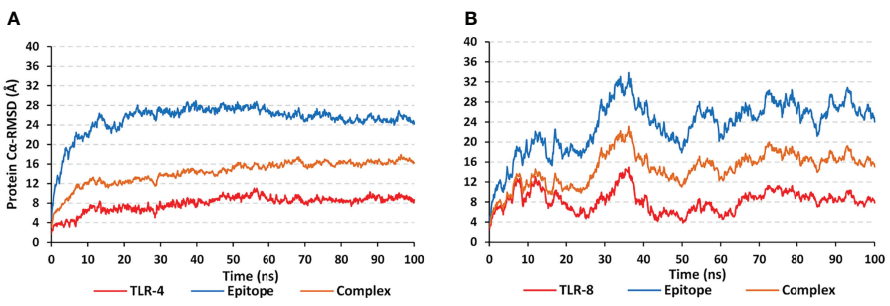


FIGURE 6 | Trajectory-based thermodynamic stability analysis of the simulated multiepitope vaccine in complex with different TLRs along 100 ns all-atom simulation runs. The estimated *Cα* RMSDs (Å) of simulated **(A)** epitope/TLR-4 model; **(B)** epitope/TLR-8 model, were represented as a function of molecular dynamics timelines (ns).

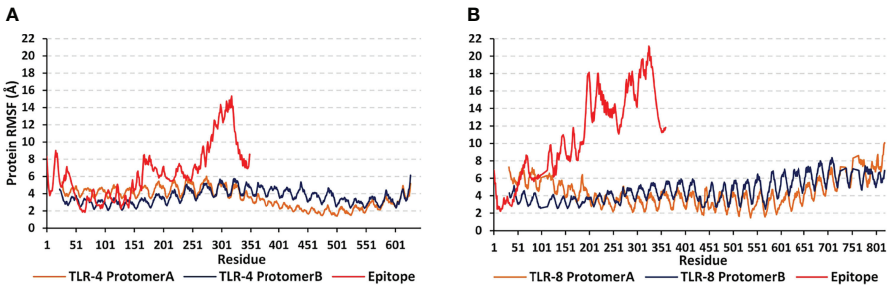


FIGURE 7 | Global stability analysis and residue-wise mobility trends of the simulated multiepitope vaccine in complex with different TLRs along 100 ns all-atom simulation runs. The protein's $C\alpha$ RMSFs (Å) of simulated **(A)** epitope/TLR-4 model; **(B)** epitope/TLR-8 model, were represented as a function of constituting residue sequence numbering; TLR-4 protomer A = 27-627; TLR-4 protomer B = 23-627; TLR-8 protomer A = 32-818; TLR-8 protomer B = 33-818; multiepitope vaccine = 1-350 amino acid numbering.

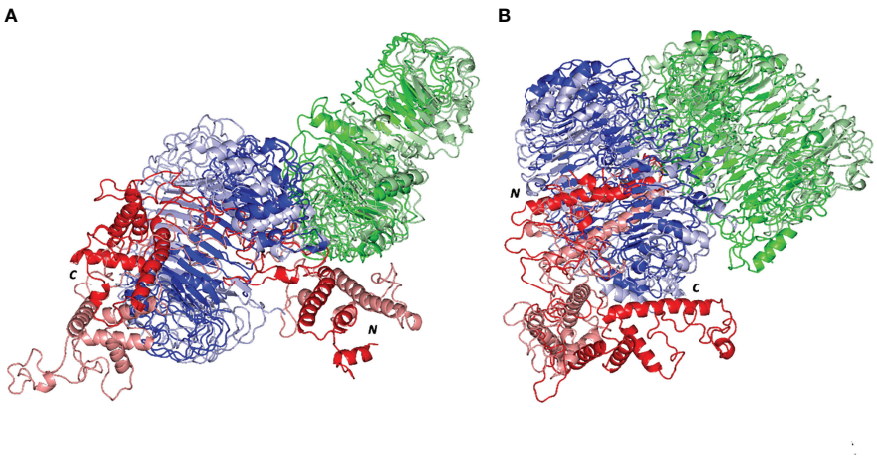


FIGURE 8 | Comparative dynamic conformation/orientation for the simulated multiepitope vaccine/TLR models. Aligned initial and final simulated trajectories of each model; **(A)** multiepitope vaccine/TLR-4; **(B)** multiepitope vaccine/TLR-8. Proteins are represented as 3D cartoons and differently colored in green, blue, or red for respective TLR protomer-A, TLR protomer-B, or the epitope proteins, as well as, in dark or light colors in regard to their respective extracted 0 ns or 100 ns trajectories. The carboxy and amine terminals of the simulated multiepitope vaccines are assigned with *C* and *N*-letters, respectively.

TABLE 5 | Free energies of binding and their individual energy contribution terms for multiepitope vaccine/TLR complexes.

	Multiepitope vaccine/TLR-4 complex			Multiepitope vaccine/TLR-8 complex		
Energy (kJ/mol ± S.D.)	**Protomer A**	**Protomer B**	**Combined**	**Protomer A**	**Protomer B**	**Combined**
$\Delta G_{van\ der\ Waal}$	-315.619 ± 51.163	-629.274 ± 41.117	-944.893 ± 56.14	-114.891 ± 21.397	-390.993 ± 50.268	-505.884 ± 45.833
$\Delta G_{electrostatic}$	-5732.477 ± 164.269	-9604.370 ± 101.274	-15336.847 ± 132.7715	-2576.698 ± 79.934	-7126.549 ± 147.729	-9703.247 ± 43.833
$\Delta Gs_{olvation}$	572.009 ± 98.711	1577.708 ± 112.365	2149.717 ± 105.538	17.459 ± 140.784	2216.553 ± 163.088	2234.012 ± 101.936
ΔG_{SASA}	-37.303 ± 5.591	-83.648 ± 1.513	-120.951 ± 3.552	-13.651 ± 5.475	-69.720 ± 13.213	-83.371 ± 11.204
ΔG_{total}	-5513.390 ± 89.208	-8739.584 ± 61.622	-14252.974 ± 77.325	-2687.781 ± 103.224	-5370.709 ± 75.238	-8058.490 ± 81.431

$\Delta G_{electrostatic}$ interactions over the hydrophobic potentials reaching up to several folds. Within both simulated models, minimal polar solvation energy was assigned for protomer A in regard to protomer B, being of particular differential values for the vaccine in bound with TLR-8 target. On the contrarily, higher apolar solvation energy term contributions were assigned for the TLR's protomer B over those of protomer A, showing the greatest value for the TRL-4 complex system.

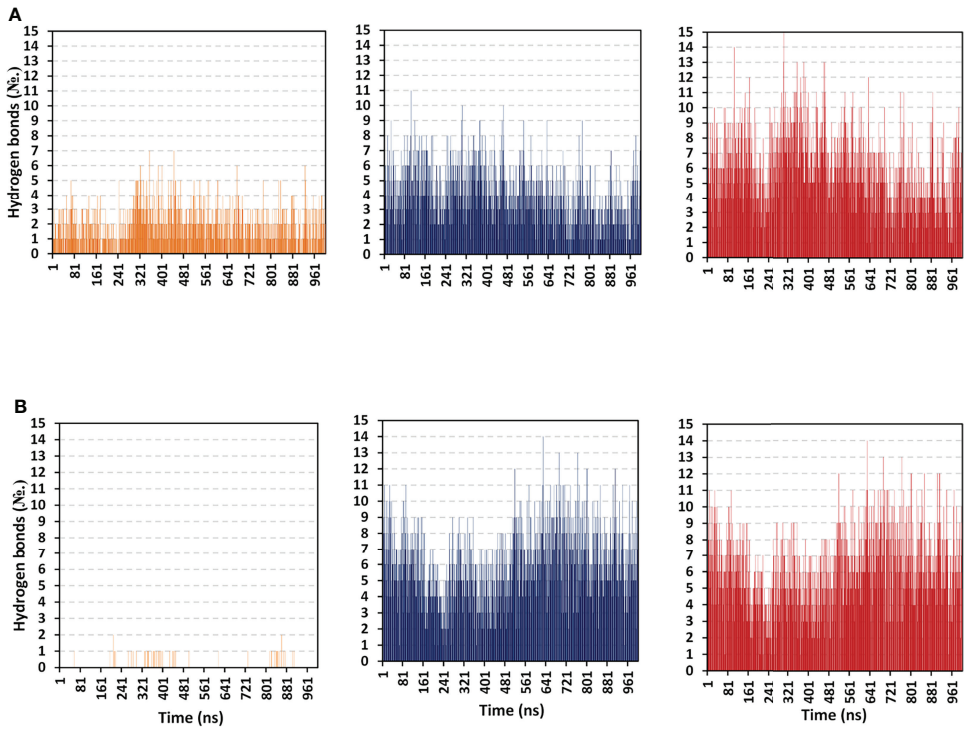


FIGURE 9 | Hydrogen bond number-time evolution within the simulated multiepitope vaccine/TLR models. **(A)** multiepitope vaccine/TLR-4; **(B)** multiepitope vaccine/TLR-8 models. Left, middle and right panels are respective for the vaccine/TLR protomer A, protomer B, and whole target protein complexes.

Further exploration of the differential $\Delta G_{electrostatic}$protein-protein binding interactions between both TLR models as well as cognate protomers, hydrogen bond analysis was conducted across the simulated trajectories (**Figure 9**). As a general observation, vaccine binding towards the target's protomer B illustrated the greater range of h and TLR-8 bound complexes (average hydrogen bond №; ~ 4 ± 1.87 and 6 ± 2.19, respectively). On the contrarily, a moderate to a limited number of the hydrophilic hydrogen bonds were depicted for the TLR-4 protomer A (~ 2 ± 1.27 H-bonds) as well as its corresponding protein at TLR-8 system (< 1 ± 0.25 H-bonds). Interestingly, patterns of hydrogen bonding at the vaccine/TLR-4 complex were higher at the first half of the simulation runs as compared to the late simulation timeframes. These polar protein-protein binding patterns were inversed for the TLR-8 bounded complex where higher numbers of hydrogen bonding were depicted beyond the protein system convergence and across the respective equilibration plateau (> 70 ns timeframes).

Across the 100 ns simulated run of vaccine/TLR-4 system, significant hydrogen bond frequencies (% occurrence of 100 ns time) were assigned for the vaccine-TLR residue pairs; Gly21-Arg496 (29.11%), Arg22-Asp490 (49.01%), Cys23-Ser518 (38.12%), Lys88-Glu608 (46.14%), Gly97-Glu485 (28.22%), Tyr107-Glu605 (24.75%), Tyr114-Asp502 (35.46%), Ser116-Gln484 (26.24%), Arg125-Asp550 (60.89%), Asn131-Glu603 (27.23%), Arg132-Glu603 (55.45%), and Arg140-His426 (60.89%). Concerning the inbound TLR-8 model, the following vaccine-TLR residue's hydrogen bond pairs were correlated to significant occupancies; Glu1-Glu129 (26.63%), Glu1-Glu150 (22.57%), Lys5-Glu133 (47.52%), Arg19-Asp462 (40.69%), Glu32-Lys476 (33.17%), Thr40-Asp645 (20.99%), Arg41-Asp536 (73.76%), Arg41-Asp560 (48.02%), Lys44-Glu612 (50.10%), Arg47-Glu668 (67.33%), Arg48-Glu612 (99.56%), Lys50-Glu768 (42.08%), and Lys88-Glu691 (64.65%). It worth noting that the residue pairs of TLR-8 model were exclusively related to the protomer B rather than protomer A target protein.

To further highlight the significance of the above-described hydrogen bond-residue pairs within the multiepitope vaccine/TLR complex stability, the $\Delta G_{Total\ binding}$ was further decomposed into the protein's residue-wise energy contributions (**Figure 10**). Findings within the latter Figure showed higher positive-values energy contributions (repulsive forces) for the TLR-8 protomers as compared to those of TLR-4 ones. Energy contributions were more distributed across the residues of both TLR-4 protomers. On the contrarily, the TLR-8 protomer A residue-wise energy contributions were almost concentrated for the *C*-terminal residues. Regarding the bounded multiepitope vaccines, higher residue-wise energy contributions were assigned for that in bound to TLR-4 and particularly towards the epitope's *N*-terminal side. the highest negative-value energy contributions were assigned for the above-described vaccine's residue which was identified to mediate relevant hydrogen bonding pairs. The highest energy binding residues included; Arg22 (-302.37 kJ/mol), Lys88 (-373.25 kJ/mol), Arg 125 (-359.65 kJ/mol), Arg132 (-353.73 kJ/mol), and Arg140 (-315.34 kJ/mol) for epitope/TLR-4 model. on the other

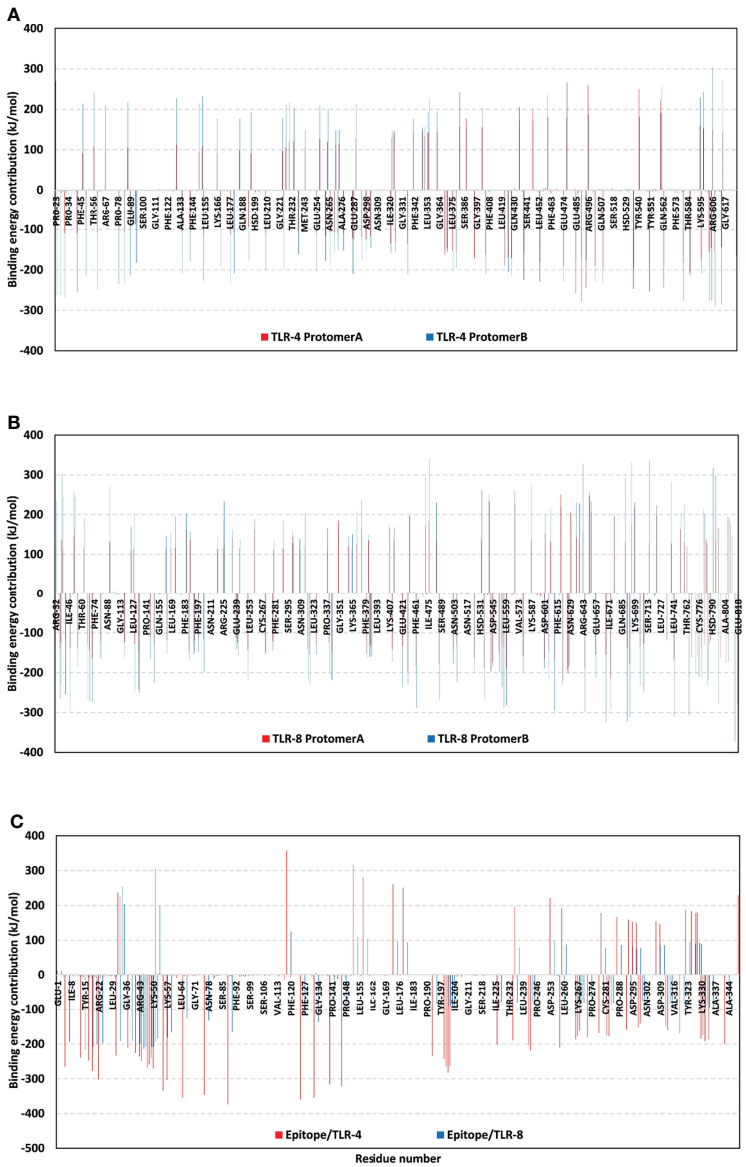


FIGURE 10 | Residue-wise free binding energy contributions for simulated multiepitope vaccine/TLR models. **(A)** bounded TLR-4 protomers A and B; **(B)** bounded TLR-8 protomers A and B; **(C)** bounded multiepitope vaccine at TLR-4 and 8 binding sites. Regions of negative-valued **Δ*G*** kJ/mol confer highly favored binding forces of attraction, while as those of positive-valued **Δ*G*** kJ/mol confer unfavored binding forces of high-repulsions.

hand, Lys5 (-192.88 kJ/mol), Arg19 (-200.52 kJ/mol), Arg41 (-198.69 kJ/mol), Lys44 (-206.76 kJ/mol), Arg47 (-207.69 kJ/mol), Arg48 (-207.42 kJ/mol), Lys50 (-183.53 kJ/mol), and Lys88 (-165.52 kJ/mol) were assigned of highest energy binding contributions within epitope/TLR-8 complex system.

Immune Simulation of the Designed Vaccine

The generated immune response as a result of the proposed vaccine successive injections is shown in **Supplementary Figure 4**. Firstly, the multitope vaccine stimulated the formation of high levels of IgM + IgG, where these antibodies demonstrated a growing level trend with successive injections. Secondly, several cytokines were predicted to be stimulated following the proposed vaccine injection where INF-γ showed the highest level of induced cytokines. lastly, both B and Th cells experienced a high increase with successive vaccine doses and the highest level of both cells was obtained following the second booster dose injection.

Vaccine Reverse Translation in Addition to Codon Optimization

Finally, we reached the last stage where the computational analysis included a reverse translation and codon optimization of our multitope vaccine using JCat server for this purpose. The

analysis of the amino acid sequence of our vaccine showed GC content of 52.9% indicating an accepted value as the accepted range ranged from 30 to 70%. Moreover, calculating Codon Adaptation Index (CAI) indicated a high probability of protein expression with a value of 1, when transferring these experiments to the wet lab, the CAI value ranges from 0 to 1 with the accepted range ranges between 0.8 and 1.

DISCUSSION

MARV is a lethal zoonotic virus that can lead to a series of symptoms starting from abdominal or chest pain and develops into a massive hemorrhage with multi-organ dysfunction (59). Currently, there is no specific treatment for this viral hemorrhagic fever and patients receive only supportive treatment with electrolytes, pain killers, and antibiotics to inhibit the secondary infection (60). Due to its high fatality rate and successive outbreaks, the development of an effective vaccine against this deadly virus is a major health priority (61). In addition to the traditional approaches for vaccine development, the great revolution in the sequencing methods and the availability of huge data regarding pathogens' genomes have led to a growing methodology for vaccine designing that would save both cost and time (62). The continuous development of the computational tools and servers that deal with the genomic and proteomic data have revolutionized the approach of the *in silico* vaccine prediction that in turn has been extended to involve bacteria, viruses, fungi, and even cancer (63). Examples of targeted viruses for designing an epitope-based vaccine through a computational approach includes Zika virus (64), Influenza A virus (65), African Swine Fever Virus (66), and COVID- 19 (67)

The application of immunoinformatics for predicting vaccine candidates against MARV has been applied in previous trials. A recent study (68) revealed an epitope-based vaccine based on predicted B and T cell epitopes of RNA Dependent RNA Polymerase Protein. Another study (69), targeted the same enzyme for the prediction of B and T cell epitopes and nominated the best-selected epitopes as vaccine candidates against MARV. On the other hand, unlike previous trials that targeted one protein candidate without specific rationale, we filtered the MARV whole proteome based on the antigenicity and the virulence of each protein and came out with three proteins namely VP24, VP30, and envelope glycoprotein. VP24 has a major role in the formation of viral infectious particles. It was reported that silencing of VP24 through molecular biology techniques would not affect the viral transcription or replication. Instead, the release of viral particles was significantly impaired which supports the role of VP24 in the interaction between nucleocapsid and budding site at the plasma membrane (70). Moving to VP30, it was reported that knockdown of VP30 in MARVinfected cells results in a significant reduction in the total viral proteins production in addition to the inhibition of viral particles release (71). The last candidate, envelope glycoprotein, mediates the process of MARV attachment and entry into its target receptors of infected cells (72), and for this reason, it was selected as a potential MARV vaccine target (73). The next stage after the detection of vaccine candidates is the prediction of B and T cell epitopes. Usage of epitopes instead of the whole proteins has the advantage of targeting the antigenic parts of proteins and reducing the probability of adverse allergic reactions (74). The limited immunogenicity of single-epitope peptides moves the prediction process to design a multitope vaccine instead where vaccine construct with several epitopes would have an improved antigenicity and immunogenicity especially when an appropriate adjuvant sequence is incorporated (75). This was the strategy of the current study which demonstrated an advantage over studies (68) and (69) that applied a similar approach for designing a MARV vaccine.

The current study has either applied the default thresholds suggested by the prediction severs or more strict scores for more stringent filtered results. For example, VaxiJen 2.0 server, which was utilized for antigenicity prediction, assumes that the peptides with a score of more than 0.4 are probable antigens but in the current study we put our threshold at 0.5. On the other hand, the prediction of B cell epitopes was performed based on the default threshold suggested by the prediction server. Moving to the prediction servers and applied algorithms, we applied NetMHCpan EL 4.0 prediction tool for the estimation of MHCI peptides, and this tool's performance was assessed on two independent external data sets; one consisting of 15,965 eluted ligands covering 27 HLA molecules, and another consisting of 1,251 validated CTL epitopes covering 80 HLA molecules reported in the IEDB (76). We also used bepipred linear epitope prediction method for the estimation of B cell epitopes, a tool that when tested on the validation data set showed a significantly better performance than any of the other tested methods (30). For the antigenicity assessment, VaxiJen 2.0 was used. This server applies an alignment-free approach for antigen prediction, which is based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. The models performed well in validation and showed a prediction accuracy of 70% to 89% (77). In addition to the validation stage that was applied during the development of these tools, another source of validation is the protection of the epitopes, that were predicted through these tools, against several pathogens when they were assessed for their activity in wet-lab experiments (28, 78). It is important to mention that the prediction of these tools is not 100% true and here comes one of the major advantages of the multi-epitope vaccines over the ones that are based on mono epitope which is the involvement of more than one epitope that would reduce the probability of the production of a non-effective final vaccine due to a prediction error.

In the current study, the single predicted epitopes were filtered based on several criteria such as the percentile rank, the antigenicity score, the degree of binding to a representative allele, and the conservancy of these epitopes. Moreover, the prediction was initially performed against a reference list of alleles to provide a high percentage in terms of population coverage. For the multitope construction, amino acid linkers were used to link top-ranked single epitopes and assure the effective separation of the assembled mono epitopes *in vivo* (79).

The first linker, EAAAK, was used to enhance the bi-functional catalytic activity, and give stiffness in addition to enhancing fusion protein stability. The second linker, GPGPG, was selected for its ability to induce HTL immune response and the ability to break the junctional immunogenicity, resulting in individual epitopes' restoration of immunogenicity. The final linker, KK, was employed because of its ability to bring the pH value close to the physiological range (21). Moreover, the beta-defensin adjuvant and PADRE peptide were added to finalize the proposed vaccine construct with their added roles of potentiating the immune response and minimizing the HLA polymorphism in the population (80). Selected T cell single epitopes were docked in a representative receptor as a primary step to validate their reactivity (81) where the binding energy of the docked complex supported the nomination of top predicted epitopes. Following the multitope vaccine construction and before predicting its three-dimensional structure, the proposed vaccine construct was analyzed for its characteristics and was found to be soluble upon over-expression, antigenic, non-allergen, and non-toxic. Other assessed physicochemical characteristics demonstrated that the instability index was 38.9 indicating that the construct is stable, the aliphatic index was 73.3 indicating that the construct is thermostable, and the negative GRAVY score (-0.33) indicated that the construct is hydrophilic. These promising results moved our study to the next steps of tertiary structure prediction and docking analysis. The proposed vaccine tertiary structure was predicted and refined computationally and the validation scores showed a high quality of the predicted model.

The final stage of the current study was a molecular docking analysis between the designed vaccine and TLR-4 and 8 where the low binding energy scores gave a primary indication that a good binding is predicted to occur between the proposed vaccine model and its targets and to get a closer view of the docked complex behavior, we employed normal mode analysis that was integrated into the iMODS server and the output data, that described the collective functional motions of the complex, demonstrated promising stability of the complex that was deeply analyzed in a molecular dynamics stimulation study.

The thermodynamic stability was illustrated for the designed multiepitope vaccine towards two TLR targets within the conducted 100 ns all atom dynamics simulations. Both systems were successfully converged since comparable $C\alpha$ RMSD values were depicted for each corresponding protein of opposite simulated models at the end of the simulation runs. Moreover, the differential $C\alpha$ RMSD values between the multiepitope vaccine and its bound TLR target were within a 3-fold difference conferring successful protein convergence needing no further simulation extensions. Generally, $C\alpha$ RMSD is the stability-indicating parameter that estimates the molecular deviations from the respective reference molecule at the initial frame. This tool has been applied to validate molecular dynamics simulation as well as ensure the significant ligand/target thermodynamic stability and confinement through furnishing low $C\alpha$ RMSD values and achieving rapid equilibration (82, 83). The here depicted ligand/TLR $C\alpha$ RMSD-based stability profiles were highly comparable to those obtained from several reported studies investigating computationally designed multitope vaccines targeting human TLRs (84–86).

The significant preferentiality of our designed multiepitope vaccine towards the TLR-4 was also illustrated where the latter showed earlier $C\alpha$ RMSD equilibration and steadier $C\alpha$-RMSD tones levelling up for more than 70 ns of the simulated runs. This could be reasoned for the differential orientation/conformation of the simulated multiepitope vaccine within each TLR binding site. The presence of two protruding loops (260-268 and 433-482 residue rages) into the TLR-8 binding pocket could have hindered the epitope/TLR-4-like transverse orientation across the TLR-8's inner pocket. The binding of the designed vaccine at the lateral interface of TLR-8 was also depicted through several reported studies. Three research groups showed significant affinity of a novel engineered anti-SARS-CoV-2 vaccines towards TLR-8 target (84–86). Despite the differential vaccine size at each study, both vaccines predicted a relevant anchoring at the TLR-8 target's dimerization lateral interface. A study by Sana et al. also depicted a lateral docking orientation of anti-Crimean Congo hemorrhagic fever virus multi-valent vaccine towards three human TLRs, including the TLR-8 (86). Thus, our depicted docking findings were considered valid and highly reliable being in great condcordance with several reported data.

Validation of the adopted docking poses of the epitope/TLR-4 and -8 complex was further proceeded through investigating the RMSF tones across the molecular dynamics simulated trajectories. Typically, the RMSF-based analysis represents a residue-wise flexibility assessment tool that permits an estimation of the protein amino acids' average deviation from their reference position. Such analytical parameter would provide information regarding the inherited flexibility/mobility of the simulated protein's down to their own constituting amino acid levels (87). In our study, the simulated TLR proteins depicted typical fluctuation/thermodynamic mobility patterns being comparable with those observed within several previously reported *in-silico* studies (24, 85). However, the differentially higher RMSF values for the simulated multiepitope vaccines in regard to their bound TLRs were mostly related to the protein's respective higher structural folding and/or packing. Having the simulated TLRs at their higher oligomeric states (Homo-2-mer-A2) would rather infer minimal flexibility and higher immobility (88) as compared to the simulated monomeric multiepitope vaccines. Additionally, both TLRs exhibit a highly-dense packing shoe-like ternary protein structure with several high-ordered β-sheets. The latter would impose lower RMSF values for TLR as compared to the multiepitope vaccines where the latter possesses extended α-helices with interconnecting flexible β-loops of long-to-medium lengths. These vaccine/TLR differential flexibility profiles were also illustrated at the higher RMSDs for the overlaid initial/final timeframes as well as at the normal-state analyses developed *via* the iMODS server. The latter approach showed more uniform stiffness trends, as well as lower mobility indices, deformability, and B-factor values for TLRs in regard to the bounded vaccines.

Both the investigated MM/PBSA-based binding-free energy calculations and hydrogen bond analysis emphasized the preferential affinity of the bound vaccine towards the TLR-4 binding site over that of TLR-8 target protein. Showing higher negative-value total energies and pronounced electrostatic energy term contributions have validated the significance and comparative patterns of the vaccine's polar interactions with bound TLR-4/B as well as their cognate protomers A/B previously described within our preliminary docking findings. Finally, the conducted molecular dynamics simulations provided further validation of the conducted docking study based on energy contributions and bonding interactions. Both furnished residue-wise energy contributions and high-frequency hydrogen bond residue pairs were representative to the preliminary docking protein-protein interaction findings reported within this study.

Although limited polar/hydrogen bonding patterns were shown for the bound vaccine/TLR-8 protomer, the extended contacts with the other cognate (protomer B) have managed to overcompensate. This benched assistance allowed the vaccine/TLR-8 complex to exhibit an overall binding profile being comparable to that of the epitope/TLR-4 model. The latter highlights the significance of the vaccine's *N*-terminal conformation/orientation shift towards the TLR-8 dimerization interface for stabilizing the epitope/TLR complex beyond 70 ns and till the simulation end. Again, such fundamental shift was reasonably depicted through the increased hydrogen bond profile beyond the epitope/TLR-8 protein convergence and throughout their respective dynamic equilibration plateau. It is worth noted that such conformational/orientation shift would have greatly counterbalanced the predicted electrostatic penalties and polar solvation energies ($\Delta G_{\text{Solvation}}$) arose during the multiepitope vaccine ligand binding at TLR-8 binding site. Generally, solvation energies confer significant forces of repulsive against the ligand-binding process since such processes rely on solvent-displacement. Despite the depicted shift, the polar solvation energy terms were shown higher within the multiepitope vaccine/TLR-8 model as compared to those of the TLR-4 one. This could confer the preferentiality of the vaccine's transverse orientation at TLR-4 as compared to TLR-8, while as further signify the TLR-8 protruding loops that would impose great challenge against vaccine/TLR-8 proper anchoring. This was highly rationalized since several reported *in-silico* studies, including our presented data, showed a significant docking of their designed vaccines towards the TLR-8 lateral interface while depicting relevant thermodynamic stability throughout variable molecular dynamics simulation time runs (84–86).

Another interesting finding was presented at our residue-wise energy contribution analysis where the depicted solvation energies were majorly mediated *via* the TLR residues rather than those of the vaccines. This could be reasoned for the high-ordered water molecules along the hydrophobic surface of the multiepitope vaccine/TLR binding site. On the other hand, the total non-polar interactions ($\Delta G_{\text{van der Waal}} + \Delta G_{\text{SASA}}$) were shown to be higher at the TLR-4, particularly for the protomer B, conferring it respective larger binding surface area as well as higher hydrophobic potentialities towards vaccine anchoring. This speculation could be rationalized since accumulated evidence has considered the investigated TLRs' binding sites to be extended and of more hydrophobic nature (89). Based on all above evidence, it was satisfactory to say that the designed multiepitope vaccine showed significant binding affinity towards the two human TLR binding pockets at their biologically active higher oligomeric states (homo-2-mer-A2), yet with significant preferentiality towards that of TLR-4 target protein.

CONCLUSIONS

Application of computational approaches for vaccine design before validation through wet lab techniques is a modern path that has been applied extensively in the last few years with the advantage of being a great economical solution that saves both cost and time. Proteome exploration of MARV recommended three antigenic proteins (VP24, VP30, envelope glycoprotein) with essential physiological and pathological roles as vaccine candidates. Utilization of *in silico* tools for the prediction of B and T cell epitopes then assembly of a multitope vaccine came up with a potential vaccine construct having promising physicochemical and immunological characteristics. In addition to that, validation of both mono and multiple epitopes through molecular docking-coupled dynamics simulation analysis would support the nomination of the currently designed vaccine as a putative solution against MARV. We recommend directing this vaccine to the next stage of biological assessment for validating our findings.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

MAS, WKA, AMR, MA, MAE, and KMD: conceptualization, methodology, and original draft preparation. SA, EF, SMAS, SSE., and RAE: writing— review, and editing. AAB, SSE, and RMD: Funding. MAS and MAE: supervision and project administration. All authors contributed to the article and approved the submitted version

FUNDING

This research was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under grant number (RG-1-166-43).

ACKNOWLEDGMENTS

The Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, has funded this project under grant no. (RG-1-166-43). Therefore, all the authors acknowledge, with thanks, DSR for technical and financial support. The simulations in this work were performed at King Abdulaziz University's High Performance Computing Center (Aziz Supercomputer) (http://hpc.kau.edu.sa) The authors, therefore, acknowledge with thanks the center for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.907481/full#supplementary-material

REFERENCES

1. Kajihara M, Hang'Ombe BM, Changula K, Harima H, Isono M, Okuya K, et al. Marburgvirus in Egyptian Fruit Bats, Zambia. *Emerg Infect Dis* (2019) 25 (8):1577. doi: 10.3201/eid2508.190268
2. Towner JS, Amman BR, Sealy TK, Reeder Carroll SA, Comer JA, Kemp A, et al. Isolation of Genetically Diverse Marburg Viruses From Egyptian Fruit Bats. *PLoS Pathog* (2009) 5(7):e1000536. doi: 10.1371/journal.ppat.1000536
3. Bauer MP, Timen A, Vossen ACTM, van Dissel JT. Marburg Haemorrhagic Fever in Returning Travellers: An Overview Aimed at Clinicians. *Clin Microbiol Infect* (2019) 28:100821. doi: 10.1111/1469-0691.12673
4. Bausch DG, Borchert M, Grein T, Roth C, Swanepoel R, Libande ML, et al. Risk Factors for Marburg Hemorrhagic Fever, Democratic Republic of the Congo. *Emerg Infect Dis* (2003) 9(12):1531. doi: 10.3201/eid0912.030355
5. Roddy P, Marchiol A, Jeffs B, Palma PP, Bernal O, de la Rosa O, et al. Decreased Peripheral Health Service Utilisation During an Outbreak of Marburg Haemorrhagic Fever, Uíge, Angola, 2005. *Trans R Soc Trop Med Hyg* (2009) 103(2):200–02. doi: 10.1016/j.trstmh.2008.09.001
6. Nyakarahuka L, Ojwang J, Tumusiime A, Balinandi S, Whitmer S, Kyazze S, et al. Isolated Case of Marburg Virus Disease, Kampala, Uganda, 2014. *Emerg Infect Dis* (2017) 23(6):1001. doi: 10.3201/eid2306.170047
7. Hardick J, Woelfel R, Gardner W, Ibrahim S. Sequencing Ebola and Marburg Viruses Genomes Using Microarrays. *J Med Virol* (2016) 88(8):1303–08. doi: 10.1002/jmv.24487
8. Modrof J, Möritz C, Kolesnikova L, Konakova T, Hartlieb B, Randolf A, et al. Phosphorylation of Marburg Virus VP30 at Serines 40 and 42 is Critical for its Interaction With NP Inclusions. *Virology* (2001) 287(1):171–82. doi: 10.1006/viro.2001.1027
9. Amatya P, Wagner N, Chen G, Luthra P, Shi L, Borek D, et al. Inhibition of Marburg Virus RNA Synthesis by a Synthetic Anti-VP35 Antibody. *ACS Infect Dis* (2019) 5(8):1385–96. doi: 10.1021/acsinfecdis.9b00091
10. Mittler E, Schudt G, Halwe S, Rohde C. Becker S. A Fluorescently Labeled Marburg Virus Glycoprotein as a New Tool to Study Viral Transport and Assembly. *J Infect Dis* (2018) 218(Suppl 5):S318–32. doi: 10.1093/infdis/jiy424
11. Bhattarai N, Gerstman BS, Chapagain PP. Role of K-Loop Cysteine Residues in the Marburg Virus Protein VP24-Human Keap1 Complex. *ACS Omega* (2018) 3(12):18639–45. doi: 10.1021/acsomega.8b02386
12. Liu B, Dong S, Li G, Wang W, Liu X, Wang Y, et al. Structural Insight Into Nucleoprotein Conformation Change Chaperoned by VP35 Peptide in Marburg Virus. *J Virol* (2017) 91(16):e00825–17. doi: 10.1128/jvi.00825-17
13. Madara JJ, Han Z, Ruthel G, Freedman BD, Harty RN. The Multifunctional Ebola Virus VP40 Matrix Protein is a Promising Therapeutic Target. *Future Virol* (2015) 10(5):537–46. doi: 10.2217/fvl.15.6
14. Daddario-DiCaprio KM, Geisbert TW, Ströher U, Geisbert JB, Grolla A, Fritz EA, et al. Postexposure Protection Against Marburg Haemorrhagic Fever With Recombinant Vesicular Stomatitis Virus Vectors in non-Human Primates: An Efficacy Assessment. *Lancet* (2006) 367(9520):1399–404. doi: 10.1016/S0140-6736(06)68546-2
15. Callendret B, Vellinga J, Wunderlich K, Rodriguez A, Steigerwald R, Dirmeier U, et al. A Prophylactic Multivalent Vaccine Against Different Filovirus Species is Immunogenic and Provides Protection From Lethal Infections With Ebolavirus and Marburgvirus Species in non-Human Primates. *PLoS One* (2018) 13(2):e0192312. doi: 10.1371/journal.pone.0192312
16. Grant-Klein RJ, Altamura LA, Badger CV, Bounds CE, Van Deusen NM, Kwilas SA, et al. Codon-Optimized Filovirus DNA Vaccines Delivered by Intramuscular Electroporation Protect Cynomolgus Macaques From Lethal Ebola and Marburg Virus Challenges. *Hum Vaccines Immunother* (2015) 11 (8):1991–2004. doi: 10.1080/21645515.2015.1039757
17. Swenson DL, Warfield KL, Larsen T, Alves DA, Coberley SS, Bavari S. Monovalent Virus-Like Particle Vaccine Protects Guinea Pigs and Nonhuman Primates Against Infection With Multiple Marburg Viruses. *Expert Rev Vaccines* (2008) 7(4):417–29. doi: 10.1586/14760584.7.4.417
18. Marzi A, Menicucci AR, Engelmann F, Callison J, Horne EJ, Feldmann F, et al. Protection Against Marburg Virus Using a Recombinant VSV-Vaccine Depends on T and B Cell Activation. *Front Immunol* (2019) 9. doi: 10.3389/fimmu.2018.03071
19. Baruah V, Bose S. Immunoinformatics-Aided Identification of T Cell and B Cell Epitopes in the Surface Glycoprotein of 2019-Ncov. *J Med Virol* (2020) 92 (5):495–500. doi: 10.1002/jmv.25698
20. Soltan MA, Elbassiouny N, Gamal H, Elkaeed EB, Eid RA, Eldeen MA, et al. *In Silico* Prediction of a Multitope Vaccine Against Moraxella Catarrhalis: Reverse Vaccinology and Immunoinformatics. *Vaccines* (2021) 9(6):669. doi: 10.3390/vaccines9060669
21. Soltan MA, Behairy MY, Abdelkader MS, Albogami S, Fayad E, Eid RA, et al. *In Silico* Designing of an Epitope-Based Vaccine Against Common E. Coli Pathotypes. *Front Med* (2022) 9:829467. doi: 10.3389/fmed.2022.829467
22. Soltan MA, Eldeen MA, Elbassiouny N, Mohamed I, El-damasy DA, Fayad E, et al. Proteome Based Approach Defines Candidates for Designing a Multitope Vaccine Against the Nipah Virus. *Int J Mol Sci* (2021) 22 (17):9330. doi: 10.3390/ijms22179330
23. Akhtar N, Joshi A, Kaushik V, Kumar M, Mannan MA ul. *In-Silico* Design of a Multivalent Epitope-Based Vaccine Against Candida Auris. *Microb Pathog* (2021) 155:104879. doi: 10.1016/j.micpath.2021.104879
24. Soltan MA, Eldeen MA, Elbassiouny N, Kamel HL, Abdelraheem KM, El-Gayyed HA, et al. *In Silico* Designing of a Multitope Vaccine Against Rhizopus Microsporus With Potential Activity Against Other Mucormycosis Causing Fungi. *Cells* (2021) 10(11):3014. doi: 10.3390/cells10113014
25. Zhang L. Multi-Epitope Vaccines: A Promising Strategy Against Tumors and Viral Infections. *Cell Mol Immunol* (2018) 15(2):182–4. doi: 10.1038/cmi.2017.92
26. Hasanzadeh S, Habibi M, Shokrgozar MA, Ahangari Cohan R, Ahmadi K, Asadi Karam MR, et al. *In Silico* Analysis and *In Vivo* Assessment of a Novel Epitope-Based Vaccine Candidate Against Uropathogenic Escherichia Coli. *Sci Rep* (2020) 10(1):1–16. doi: 10.1038/s41598-020-73179-w
27. Ardestani H, Nazarian S, Hajizadeh A, Sadeghi D, Kordbacheh E. *In Silico* and *In Vivo* Approaches to Recombinant Multi-Epitope Immunogen of GroEL Provides Efficient Cross Protection Against S. Typhimurium, S. Flexneri and S. Dysenteriae. *Mol Immunol* (2022) 144:96–105. doi: 10.1016/j.molimm.2022.02.013
28. Kardani K, Hashemi A, Bolhassani A. Comparative Analysis of Two HIV-1 Multiepitope Polypeptides for Stimulation of Immune Responses in BALB/c Mice. *Mol Immunol* (2020) 119:106–22. doi: 10.1016/j.molimm.2020.01.013
29. Doytchinova IA, Flower DR. Bioinformatic Approach for Identifying Parasite and Fungal Candidate Subunit Vaccines. *Open Vaccine J* (2010) 1(1):4. doi: 10.2174/1875035400801010022
30. Larsen JEP, Lund O, Nielsen M. Improved Method for Predicting Linear B-Cell Epitopes. *Immunome Res* (2006) 2(1):1–7. doi: 10.1186/1745-7580-2-2
31. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: Immune Epitope Database - Analysis Resource in 2019. *Nucleic Acids Res* (2019) 47(W1):W502–6. doi: 10.1093/nar/gkz452

32. Weiskopf D, Angelo MA, De Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive Analysis of Dengue Virus-Specific Responses Supports an HLA-Linked Protective Role for CD8+ T Cells. *Proc Natl Acad Sci USA* (2013) 110(22):E2046–53. doi: 10.1073/pnas.1305227110
33. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional Classification of Class II Human Leukocyte Antigen (HLA) Molecules Reveals Seven Different Supertypes and a Surprising Degree of Repertoire Sharing Across Supertypes. *Immunogenetics* (2011) 63(6):325–35. doi: 10.1007/s00251-011-0513-0
34. Dhanda SK, Vir P, Raghava GPS. Designing of Interferon-Gamma Inducing MHC Class-II Binders. *Biol Direct* (2013) 8(1):1–15. doi: 10.1186/1745-6150-8-30
35. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J Chem Theory Comput* (2014) 10(10):4745–58. doi: 10.1021/ct500592m
36. Oleg T, Arthur JO. AutoDock Vina: Improving the Speed and Accuracy of Docking With a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* (2010) 31(2):455–61. doi: 10.1002/jcc.21334
37. Franke ED, Hoffman SL, Sacci JB, Wang R, Charoenvit Y, Appella E, et al. Pan DR Binding Sequence Provides T-Cell Help for Induction of Protective Antibodies Against Plasmodium Yoelii Sporozoites. *Vaccine* (1999) 17(9-10):1201–5. doi: 10.1016/S0264-410X(98)00341-7
38. Saha S, Raghava GPS. AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes. *Nucleic Acids Res* (2006) 34(Suppl 2):W202–9. doi: 10.1093/nar/gkl343
39. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. *In Silico* Approach for Predicting Toxicity of Peptides and Proteins. *PLoS One* (2013) 8(9):e73957. doi: 10.1371/journal.pone.0073957
40. Magnan CN, Randall A, Baldi P. SOLpro: Accurate Sequence-Based Prediction of Protein Solubility. *Bioinformatics* (2009) 25(17):2200–07. doi: 10.1093/bioinformatics/btp386
41. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. "Protein Identification and Analysis Tools on the ExPASy Server,". In: *The Proteomics Protocols Handbook* (2005) 571–607. doi: 10.1385/1-59259-890-0:571
42. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 Years on. *Nucleic Acids Res* (2019) 47(W1):W402–7. doi: 10.1093/nar/gkz297
43. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: A Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Res* (2005) 33(Suppl 2):W72–6. doi: 10.1093/nar/gki396
44. Heo L, Park H, Seok C. GalaxyRefine: Protein Structure Refinement Driven by Side-Chain Repacking. *Nucleic Acids Res* (2013) 41(W1):W384–8. doi: 10.1093/nar/gkt458
45. Colovos C, Yeates TO. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci* (1993) 2(9):1511–19. doi: 10.1002/pro.5560020916
46. Wiederstein M, Sippl MJ. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res* (2007) 35(Suppl 2):W407–10. doi: 10.1093/nar/gkm290
47. 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(2):255–78. doi: 10.1038/nprot.2016.169
48. Laskowski RA. PDBsum: Summaries and Analyses of PDB Structures. *Nucleic Acids Res* (2001) 29(1):221–2. doi: 10.1093/nar/29.1.221
49. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: Internal Coordinates Normal Mode Analysis Server. *Nucleic Acids Res* (2014) 42(W1):W271–6. doi: 10.1093/nar/gku339
50. Lopéz-Blanco JR, Garzón JI, Chacón P. Imod: Multipurpose Normal Mode Analysis in Internal Coordinates. *Bioinformatics* (2011) 27(20):2843–50. doi: 10.1093/bioinformatics/btr497
51. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations With GROMACS. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* Cham: Springer (2014), 3–27. doi: 10.1007/978-3-319-15976-8_1
52. Helal MA, Shouman S, Abdelwaly A, Elmehrath AO, Essawy M, Sayed SM, et al. Molecular Basis of the Potential Interaction of SARS-CoV-2 Spike Protein to CD147 in COVID-19 Associated-Lymphopenia. *J Biomol Struct Dyn* (2020) 40(3):1109–19. doi: 10.1080/07391102.2020.1822208
53. Zaki AA, Ashour A, Elhady SS, Darwish KM, Al-Karmalawy AA. Calendulaglycoside A Showing Potential Activity Against SARS-CoV-2 Main Protease: Molecular Docking, Molecular Dynamics, and SAR Studies. *J Tradit Complement Med* (2021) 12(1):16–34. doi: 10.1016/j.jtcme.2021.05.001
54. Páll S, Hess B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Comput Phys Commun* (2013) 184(12):2641–50. doi: 10.1016/j.cpc.2013.06.003
55. Kumari R, Kumar R, Lynn A. G-Mmpbsa -A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J Chem Inf Model* (2014) 54(7):1951–62. doi: 10.1021/ci500020m
56. 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(suppl_2):W526–31. doi: 10.1093/nar/gki376
57. Rossi AH, Farias A, Fernández JE, Bonomi HR, Goldbaum FA, Berguer PM. Brucella Spp. Lumazine Synthase Induces a TLR4-Mediated Protective Response Against B16 Melanoma in Mice. *PLoS One* (2015) 10(5):e0126827. doi: 10.1371/journal.pone.0126827
58. Almalki AJ, Ibrahim TS, Elhady SS, Hegazy WAH, Darwish KM. Computational and Biological Evaluation of β-Adrenoreceptor Blockers as Promising Bacterial Anti-Virulence Agents. *Pharmaceuticals* (2022) 15(2):110. doi: 10.3390/ph15020110
59. Nyakarahuka L, Shoemaker TR, Balinandi S, Chemos G, Kwesiga B, Mulei S, et al. Marburg Virus Disease Outbreak in Kween District Uganda, 2017: Epidemiological and Laboratory Findings. *PLoS Negl Trop Dis* (2018) 13(3):e0007257. doi: 10.1371/journal.pntd.0007257
60. Bausch DG, Feldmann H, Geisbert TW, Bray M, Sprecher AG, Boumandouki P, et al. Outbreaks of Filovirus Hemorrhagic Fever: Time to Refocus on the Patient. *J Infect Dis* (2007) 196(Suppl 2):S136–41. doi: 10.1086/520542
61. Dulin N, Spanier A, Merino K, Hutter JN, Waterman PE, Lee C, et al. Systematic Review of Marburg Virus Vaccine Nonhuman Primate Studies and Human Clinical Trials. *Vaccine* (2021) 39(2):202–8. doi: 10.1016/j.vaccine.2020.11.042
62. Kardani K, Bolhassani A, Namvar A. An Overview of *in Silico* Vaccine Design Against Different Pathogens and Cancer. *Expert Rev Vaccines* (2020) 19(8):699–726. doi: 10.1080/14760584.2020.1794832
63. Raoufi E, Hemmati M, Eftekhari S, Khaksaran K, Mahmodi Z, Farajollahi MM, et al. Epitope Prediction by Novel Immunoinformatics Approach: A State-Of-the-Art Review. *Int J Pept Res Ther* (2020) 26(2):1155–63. doi: 10.1007/s10989-019-09918-z
64. Prasasty VD, Grazzolie K, Rosmalena R, Yazid F, Ivan FX, Sinaga E. Peptide-Based Subunit Vaccine Design of T-And B-Cells Multi-Epitopes Against Zika Virus Using Immunoinformatics Approaches. *Microorganisms* (2019) 7(8):226. doi: 10.3390/microorganisms7080226
65. Sharma S, Kumari V, Kumbhar BV, Mukherjee A, Pandey R, Kondabagil K. Immunoinformatics Approach for a Novel Multi-Epitope Subunit Vaccine Design Against Various Subtypes of Influenza A Virus. *Immunobiology* (2021) 226(2):152053. doi: 10.1016/j.imbio.2021.152053
66. Fagbohun O. Immunoinformatics Based Development of A Multi-Epitope Vaccine Against African Swine Fever Virus. *FASEB J* (2021) 35. doi: 10.1096/fasebj.2021.35.s1.04383
67. Dong R, Chu Z, Yu F, Zha Y. Contriving Multi-Epitope Subunit of Vaccine for COVID-19: Immunoinformatics Approaches. *Front Immunol* (2020) 11. doi: 10.3389/fimmu.2020.01784
68. Mahmud SMN, Rahman M, Kar A, Jahan N, Khan A. Designing of an Epitope- Based Universal Peptide Vaccine Against Highly Conserved Regions in RNA Dependent RNA Polymerase Protein of Human Marburg Virus: A Computational Assay. *Anti-Infective Agents* (2019) 18(3):294–305. doi: 10.2174/2211352517666190717143949
69. Pervin T, Oany AR. Vaccinomics Approach for Scheming Potential Epitope-Based Peptide Vaccine by Targeting L-Protein of Marburg Virus. *Silico Pharmacol* (2021) 9(1):1–18. doi: 10.1007/s40203-021-00080-3
70. Bamberg S, Kolesnikova L, Möller P, Klenk H-D, Becker S. VP24 of Marburg Virus Influences Formation of Infectious Particles. *J Virol* (2005) 79(21):13421–33. doi: 10.1128/jvi.79.21.13421-13433.2005

71. Tigabu B, Ramanathan P, Ivanov A, Lin X, Ilinykh PA, Parry CS, et al. Phosphorylated VP30 of Marburg Virus Is a Repressor of Transcription. *J Virol* (2018) 92(21):e00426–18. doi: 10.1128/jvi.00426-18
72. Keshwara R, Hagen KR, Abreu-Mota T, Papaneri AB, Liu D, Wirblich C, et al. A Recombinant Rabies Virus Expressing the Marburg Virus Glycoprotein Is Dependent Upon Antibody-Mediated Cellular Cytotoxicity for Protection Against Marburg Virus Disease in a Murine Model. *J Virol* (2019) 93(6): e01865–18. doi: 10.1128/jvi.01865-18
73. Saito T, Maruyama J, Nagata N, Isono M, Okuya K, Takadate Y, et al. A Surrogate Animal Model for Screening of Ebola and Marburg Glycoprotein-Targeting Drugs Using Pseudotyped Vesicular Stomatitis Viruses. *Viruses* (2020) 12(9):923. doi: 10.3390/v12090923
74. Oyarzun P, Kashyap M, Fica V, Salas-Burgos A, Gonzalez-Galarza FF, McCabe A, et al. A Proteome-Wide Immunoinformatics Tool to Accelerate T-Cell Epitope Discovery and Vaccine Design in the Context of Emerging Infectious Diseases: An Ethnicity-Oriented Approach. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.598778
75. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y. Epitope-Based Vaccine Design: A Comprehensive Overview of Bioinformatics Approaches. *Drug Discovery Today* (2020) 25(6):1034–42. doi: 10.1016/j.drudis.2020.03.006
76. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* (2017) 199 (9):3360–68. doi: 10.4049/jimmunol.1700893
77. Doytchinova IA, Flower DR. VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinf* (2007) 8(1):1–7. doi: 10.1186/1471-2105-8-4
78. Soltan MA, Magdy D, Solyman SM, Hanora A. Design of Staphylococcus Aureus New Vaccine Candidates With B and T Cell Epitope Mapping, Reverse Vaccinology, and Immunoinformatics. *Omi A J Integr Biol* (2020) 24(4):195–204. doi: 10.1089/omi.2019.0183
79. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics Analysis and *in Silico* Designing of a Novel Multi-Epitope Peptide Vaccine Against Staphylococcus Aureus. *Infect Genet Evol* (2017) 48:83–94. doi: 10.1016/j.meegid.2016.12.010
80. 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(11):e0142563. doi: 10.1371/journal.pone.0142563
81. Mohammed AA, Shantier SW, Mustafa MI, Osman HK, Elmansi HE, Osman IAA, et al. Epitope-Based Peptide Vaccine Against Glycoprotein G of Nipah Henipavirus Using Immunoinformatics Approaches. *J Immunol Res* (2020) 2020. doi: 10.1155/2020/2567957
82. Arnittali M, Rissanou AN, Harmandaris V. Structure of Biomolecules Through Molecular Dynamics Simulations. *Proc Comput Sci* (2019) 156:69–78. doi: 10.1016/j.procs.2019.08.181
83. Liu K, Watanabe E, Kokubo H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations. *J Comput Aided Mol Des* (2017) 31(2):201–11. doi: 10.1007/s10822-016-0005-2
84. Tahir ul Qamar M, Rehman A, Tusleem K, Ashfaq UA, Qasim M, Zhu X, et al. Designing of a Next Generation Multiepitope Based Vaccine (MEV) Against SARS-COV-2: Immunoinformatics and *In Silico* Approaches. *PLoS One* (2020) 15(12):e0244176. doi: 10.1371/journal.pone.0244176
85. Sarkar B, Ullah MA, Araf Y, Rahman MS. Engineering a Novel Subunit Vaccine Against SARS-CoV-2 by Exploring Immunoinformatics Approach. *Inf Med Unlocked* (2020) 21:100478. doi: 10.1016/j.imu.2020.100478
86. Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y. Immunoinformatics-Guided Designing of Epitope-Based Subunit Vaccines Against the SARS Coronavirus-2 (SARS-CoV-2). *Immunobiology* (2020) 225(3):151955. doi: 10.1016/j.imbio.2020.151955
87. Benson NC, Daggett V. A Comparison of Multiscale Methods for the Analysis of Molecular Dynamics Simulations. *J Phys Chem B* (2012) 116(29):8722–31. doi: 10.1021/jp302103t
88. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. The Structural Basis of Lipopolysaccharide Recognition by the TLR4-MD-2 Complex. *Nature* (2009) 458(7242):1191–95. doi: 10.1038/nature07830
89. Ohto U, Yamakawa N, Akashi-Takamura S, Miyake K, Shimizu T. Structural Analyses of Human Toll-Like Receptor 4 Polymorphisms D299G and T399I. *J Biol Chem* (2012) 287(48):40611–7. doi: 10.1074/jbc.M112.404608

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fimmu.2022.956776

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Mohamed Soltan,
Sinai University, Egypt
Muhammad Alaa Eldeen,
Zagazig University, Egypt

*CORRESPONDENCE
Ajmal Khan
ajmalkhan@unizwa.edu.om
Ahmed Al-Harrasi
aharrasi@unizwa.edu.om

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

RECEIVED 30 May 2022
ACCEPTED 07 July 2022
PUBLISHED 03 August 2022

CITATION
Aziz S, Waqas M, Halim SA, Ali A, Iqbal A, Iqbal M, Khan A and Al-Harrasi A (2022) Exploring whole proteome to contrive multi-epitope-based vaccine for NeoCoV: An immunoinformtics and *in-silico* approach. *Front. Immunol.* 13:956776. doi: 10.3389/fimmu.2022.956776

Exploring whole proteome to contrive multi-epitope-based vaccine for NeoCoV: An immunoinformtics and *in-silico* approach

Shahkaar Aziz[1†], Muhammad Waqas[2,3†], Sobia Ahsan Halim[2†], Amjad Ali[3], Aqib Iqbal[1], Maaz Iqbal[1], Ajmal Khan[2*] and Ahmed Al-Harrasi[2*]

[1]Institute of Biotechnology and Genetic Engineering, the University of Agriculture Peshawar, Peshawar, Pakistan, [2]Natural and Medical Sciences Research Center, University of Nizwa, Birkat-ul-Mouz, Nizwa, Oman, [3]Department of Biotechnology and Genetic Engineering, Hazara University Mansehra, Mansehra, Pakistan

Neo-Coronavirus (NeoCoV) is a novel Betacoronavirus (β-CoVs or Beta-CoVs) discovered in bat specimens in South Africa during 2011. The viral sequence is highly similar to Middle East Respiratory Syndrome, particularly that of structural proteins. Thus, scientists have emphasized the threat posed by NeoCoV associated with human angiotensin-converting enzyme 2 (ACE2) usage, which could lead to a high death rate and faster transmission rate in humans. The development of a NeoCoV vaccine could provide a promising option for the future control of the virus in case of human infection. *In silico* predictions can decrease the number of experiments required, making the immunoinformatics approaches cost-effective and convenient. Herein, with the aid of immunoinformatics and reverse vaccinology, we aimed to formulate a multi-epitope vaccine that may be used to prevent and treat NeoCoV infection. Based on the NeoCoV proteins, B-cell, cytotoxic T lymphocyte (CTL), and helper T lymphocyte (HTL) epitopes were shortlisted. Four vaccines (Neo-1–4) were devised by fusing shortlisted epitopes with appropriate adjuvants and linkers. The secondary and three-dimensional structures of final vaccines were then predicted. The binding interactions of these potential vaccines with toll-like immune receptors (TLR-2, TLR-3, and TLR-4) and major histocompatibility complex molecules (MHC-I and II) reveal that they properly fit into the receptors' binding domains. Besides, Neo-1 and Neo-4 vaccines exhibited better docking energies of -101.08 kcal/mol and -114.47 kcal/mol, respectively, with TLR-3 as compared to other vaccine constructs. The constructed vaccines are highly antigenic, non-allergenic, soluble, non-toxic, and topologically assessable with good physiochemical characteristics. Codon optimization and *in-silico* cloning confirmed efficient expression of the designed vaccines in *Escherichia coli* strain K12. *In-silico* immune simulation indicated that Neo-1 and Neo-4 vaccines could induce a strong immune response against NeoCoV. Lastly, the binding stability and

10.3389/fimmu.2022.956776

strong binding affinity of Neo-1 and Neo-4 with TLR-3 receptor were validated using molecular dynamics simulations and free energy calculations (Molecular Mechanics/Generalized Born Surface Area method). The final vaccines require experimental validation to establish their safety and effectiveness in preventing NeoCoV infections.

KEYWORDS

immunoinformatics, multi-epitope vaccine, subunit vaccine, epitopes prediction, vaccine design, NeoCoV

1 Introduction

Coronaviruses (CoVs) are an enveloped positive-stranded RNA virus family divided into four genera, including α–, β–, γ–, and δ–CoV. The first two genera can infect mammals (bats and humans), whereas the latter primarily infect birds and, sometimes, mammals. Most coronaviruses that infect humans are believed to have originated in bats—the key mammalian coronavirus reservoir (1, 2). Bat-CoVs have received particular attention since several recently seen CoVs have been associated with unexpected disease outbreaks in the present century, causing high fatality rates and significant economic impact. Three such viruses suggested to be transmitted from bats to humans include Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and the latest Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) (2, 3).

MERS-CoV belongs to the C lineage of the Beta-CoV (Merbecoviruses), which offers a significant risk due to its high fatality rate of over 35%. Several animals, such as bats, hedgehogs, and camels, carry merbecoviruses indefinitely. Although camels are known as the intermediate hosts of MERS-CoV, bats, particularly those belonging to the Vespertilionidae family, are primarily thought to represent the virus's evolutionary source or intermediate ancestor (4). Neo-Coronavirus (NeoCoV), a novel Beta-CoV, was discovered in a Neoromicia cf. zuluensis bat specimen in 2011. This virus varied from MERS-CoV by single amino acid substitution (0.3%) in the RdRp gene fragment (translated 816-nt) and by amino acid sequence distance of 10.9% in the glycoprotein coding gene enables attachment and entry of CoV into the cell. Therefore, NeoCoV and MERS-CoV are more closely related to each other. Victor Max Corman and colleagues suggested 85% sequence similarity between these two CoVs, indicating that NeoCoV and MERS-CoV originated from common viral species (5).

Scientists have emphasized the threat posed by NeoCoV in bats reported in South Africa, which could have a high death rate and faster spread, amidst the control of cases resurging due to many evolving variants of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (6). It has also been proposed that if the NeoCoV attains mutations and causes human infection, it might develop into Coronavirus disease-22 (COVID-22) and cause symptoms three times as severe as COVID-19. Any outbreak caused by this viral strain, may trigger a 30% increase in fatalities, comparable to MERS. Around 17 million people could die due to COVID-22, compared to COVID-19, which has resulted in over 6 million people succumbing to death until now (6, 7). Nevertheless, it is worth noting that NeoCoV has yet to be confirmed in people and no reported human fatalities has observed with this virus. At the same time, the globe deals with the coronavirus disease-2019 pandemic. In Jan 2022, World Health Organization (WHO) warned about NeoCoV and demanded further research to determine NeoCoV's possible threat to humans (6).

MERS-CoV and several related bat-coronaviruses utilize human dipeptidyl peptidase-4 (DPP4), a functional receptor located on the airways cell surface (i.e., lungs), as an entry receptor (8–10). Nonetheless, the cell entry receptor for NeoCoV is unspecified so far (11). NeoCoV and closely related PDF-2180-CoV can use certain forms of bat angiotensin-converting enzyme 2 (ACE2) and human ACE2 for cell entry, according to the latest research in preprint (2). The NeoCoV virus utilizes its S1 subunit at carboxyl-terminal domains (S1-CTD) of spike protein to bind with ACE2 with great affinity and species specificity. Besides, a molecular determinate (Asp338) was uncovered at the binding interface of the virus that prevent NeoCoV entry *via* human ACE2 (2). Researchers found that T510F mutation at the receptor-binding motif leads to increase the efficiency of NeoCoV to infect the human cells expressing ACE2 receptor (2). Furthermore, antibodies produced by natural infection or vaccination against SARS-CoV-2 or MERS-

Abbreviations: CTL, cytotoxic T lymphocytes; FEL, free energy landscape; HTL, helper T lymphocytes; MHC, major histocompatibility complex; MM-GBSA, Molecular mechanics/generalized born surface area; NeoCoV, Neo-Coronavirus; Rg, radius of gyration; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SASA, solvent accessible surface area; TLR, toll-like immune receptors.

CoV are incapable of neutralizing NeoCoV infection. Also, the potential use of ACE-2 receptor by MERS-related viruses has been indicated for the first time, underlining a possible health threat posed by "MERS-CoV-2" with high mortality and spread rate (2).

Presently, there is limited knowledge regarding NeoCoV, and it is uncertain if the virus can be transmitted to humans or spread worldwide. Further research can help understand this new coronavirus and its immunology and vaccinations if it occurs in humans. Moreover, the COVID-19 pandemic is still ongoing. Thus, there is an urgent need to improve vaccination rates in both human and animals (12) and oversee other potential health implications, including the NeoCoV.

Vaccination is a crucial strategy for viral control and eradication (13). The development of a NeoCoV vaccine could provide a promising option for the future control of this virus if it infects humans. Conventional vaccine development procedures take a long time and require a great deal of manual effort (14). Immunoinformatics tools evaluate the host immune response to provide alternative techniques in order to formulate economical and advantageous vaccines against the diseases since predictions can curtail the number of *in vitro* tests required (15, 16). Vaccines based on the structural and non-structural proteins (NSPs) are reported to induce protective immune responses (17, 18). Scientifically rigorous approaches based on various proteins have been exploited to design multi-epitope subunits for the viral and parasite diseases, including malaria and SARS-CoV-2 (13, 19, 20). Here, we applied immunoinformatics techniques to predict numerous immunogenic proteins from the whole proteome of NeoCoV and developed a multi-epitope vaccine in this research by studying the structural and NSPs of NeoCoV.

2 Materials and method

2.1 Retrieval of protein sequence

The complete amino acid sequence of NeoCoV proteins was retrieved from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) in FASTA format. These comprise four structural proteins and 19 non-structural protein components. The retrieved proteins and their NCBI accession numbers are provided in the Supplementary Table S1.

2.2 Selection of protein for vaccine formulation

The position and residue range of the viral protein sequences and vaccine constructs were determined using the TMHMM-2.0 prediction tool (21). The localization predictions of the viral proteins were performed by DeepLoc (22) which uses a template-free algorithm that implements a deep neural network to envisage subcellular localization of proteins with acceptable accuracy using only sequence information (22).

2.3 Allergenicity, antigenicity, and toxicity prediction

Owing to their significance in food and/or food products, allergenicity prediction is an essential step in medications and biopharmaceuticals (23). The AllerTOP v2.0 (24) and AllergenFP P1.0 (25) were used to predict the allergenicity of the viral protein and epitopes. The former employs E-descriptors and an auto cross-covariance (ACC) transformation (26). The latter is a binary classifier between allergen and non-allergen that transforms protein sequences into uniform vectors of equal length using an ACC transformation, as described by Dimitrov et al. (24). The antigenicity of the viral proteins, epitopes, and vaccine construct was determined using a VaxiJen tool (27). The amino acid sequence of the query was used as an input, with the chosen organism target of 'virus' at a cut-off of 0.4. ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) was employed to confirm vaccine constructs antigenicity. Using the Toxinpred server (28), safety evaluation of viral proteins and epitopes were performed.

2.4 B-cell epitopes prediction

For the primary prediction of linear B-cell epitopes, the BCPreds server (29) was used. This server employs a kernel approach (30) with 75% sensitivity; the epitope length was fixed at 20 amino acids. It has been found that using various techniques to predict epitopes improves the probability of true positives in epitope prediction (31). Thus, the ABCpred (32) and BepiPred-2.0 (33) were employed for validation. Epitope scores of ABCpred were calculated using a threshold of 0.5 and a window length of 20. Further, leveraging the Ellipro server (34) with default parameters, the improved modelled structure of the multi-epitope vaccine was submitted to confirmational B-cell epitopes prediction.

2.5 MHC class-I binding epitopes prediction

Selected viral proteins were subjected to MHC-I epitopes prediction employing the NetCTL 1.2 server. The binding affinity of the predicted class epitopes was tested against 12 different HLA supertypes, including A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62 (35). Peptides were ranked based on a combined score; all other parameters were set as default. Only those epitopes were subjected to downstream analysis at this stage that indicated binding with at least four HLA I supertypes (36). To predict the immunogenicity of shortlisted MHC-I epitopes, the online bioinformatics server IEDB was used (36). The epitopes showing positive immunogenicity scores were chosen for further analysis. Then, predicted MHCs epitopes class were also tested to

find if they are antigenic, non-allergen and safe. The IEDB SMM method (http://tools.iedb.org/mhcii/) was used to test the binding affinity of shortlisted epitopes with their respective MHC-I alleles by predicting the IC_{50} values. The IC_{50} threshold value of 500nM as an MHC affinity specifies the significant immunogenicity for MHC-I restricted T cells (37). The prediction performance of the NetCTL 1.2 server was confirmed using the CTLPred server (38) with a combined approach and default setup (ANN and SVM thresholds of 0.51 and 0.36). The methodology of the vaccine selection and construction is illustrated in Figure 1.

2.6 MHC class-II binding epitopes prediction

NetMHCIIpan was used to predict the MHC-II binding epitopes from the full-length sequence of selected viral proteins. DRB10101, DRB10301, DRB10401, DRB10701, DRB10801, DRB11101, DRB11301, and DRB1 1501 were the HLA II supertype alleles targeted for epitope searching [30]. These HLA II supertype alleles cover 95% of the world's HLA variations (39). We employed NetMHCIIpan because it is presently the most reliable MHC-II epitope prediction (40). First, the predicted helper T lymphocyte (HTL) epitopes that interacted with at least three supertypes alleles were evaluated for antigenicity, non-allergenicity, and non-toxicity. Following that, B-cell, and MHC-I overlapping peptides were removed. The resulting epitopes were prioritized using the IEDB SMM method with an IC_{50} cut-off of 500 nM. Interferon-gamma (IFN-γ) production is critical for viral clearance and the activation of the host immune response. Using the IFN epitope server (http://crdd.osdd.net/raghava/ifnepitope/), we predicted the IFN-γ induction efficiency of selected MHC-II epitopes. Furthermore, the IL4pred server (https://webs.iiitd.edu.in/raghava/il4pred/) and IL10pred server (https://webs.iiitd.edu.in/raghava/il10pred/) were used to assess the interleukin-4 (IL-4; with a threshold of 0.2) and interleukin-10 (IL-10; with a threshold of 0.3) secretion potential of selected MHC-II epitopes, respectively. Finally, the prioritized epitopes sequences were submitted to the MHCPred server (41) in order to determine their 9-mer peptide by testing the binding affinity with the most common allele, DRB10101.

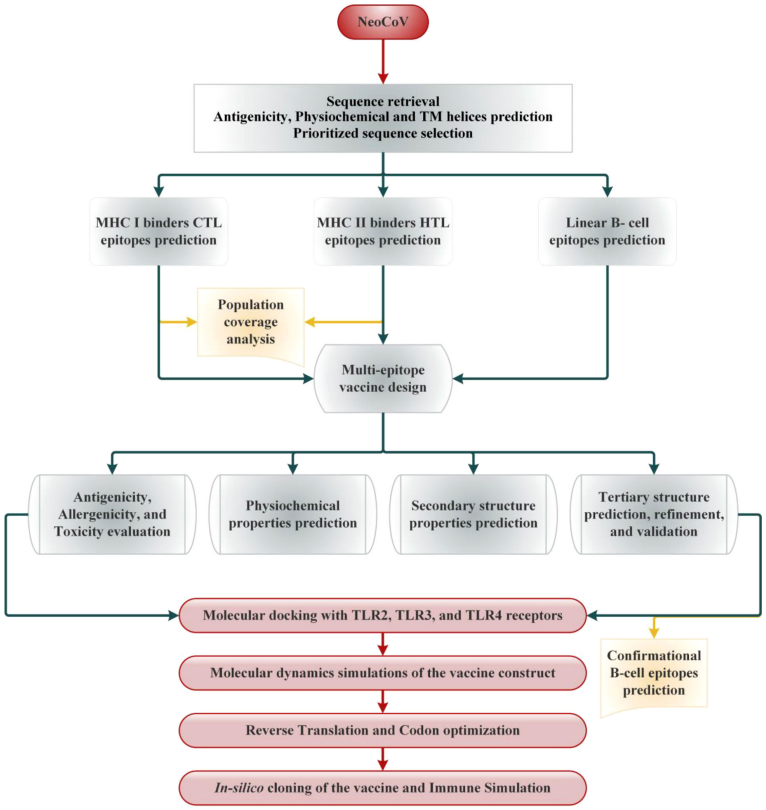


FIGURE 1
Schematic illustration of the overall strategy implemented in this study to design multi-epitope-based vaccine from NeoCoV whole proteome.

2.7 Population coverage analysis

Ethnicity and geography influence the distribution and expression of different HLA alleles (42). It can influence the formulation of an effective epitope-based vaccine. Therefore, IEDB (http://tools.iedb.org/population/) population coverage tool was used to estimate the vaccine's coverage in the target population. For that purpose, we separately and collectively assessed the picked CTL and HTL epitopes from class I and II MHC and their HLA binding alleles (43). Herein, we focused on the whole global coverage of the alleles and regions of many continents.

2.8 Evaluation of human homology and epitopes novelty

To decrease the cross-reactivity in host cells, nonoverlapping epitopes and vaccine constructs were BLASTp against human proteome (taxon id: 9606). The IEDB server (https://www.iedb.org/) was used to determine whether the prioritized epitopes had already been experimentally tested in previous investigations.

2.9 Vaccine designing

Prioritized B and T-cell epitopes from NeoCoV whole proteome were chosen for the final vaccine formulation. As safe vaccination adjuvants, compounds with immunomodulatory capabilities were added to the vaccine constructions to boost the immune system. To attach the adjuvants to the epitopes, the rigid linker EAAAK was used. The GPGPG linkers, rich in glycine and proline, were employed to space the MHC-II and B-cell epitopes. AAY, an efficient and flexible linker, was also used to connect the MHC-I epitopes. In addition, the PADRE (Pan DR T Helper Epitope) sequence was inserted to improve the vaccine construct immunogenicity. Using three adjuvants (β-defensin, heparin-binding hemagglutinin (HBHA), and 50S ribosomal protein L7/L12), four vaccines (Neo-1 to Neo-4) were prepared (44).

2.10 Physiochemical properties, solubility, and toxicity analysis

The physicochemical characteristics of the final vaccine and its subunits (epitopes) were estimated through ExPASy ProtParam server (45). The solubility of vaccine constructs was assessed using the SOLpro (46) and Protein-sol (47) servers. SOLpro uses SVM-based technique to predict the protein sequence solubility, with a tenfold cross-validation-estimated overall accuracy of about 74% (48). Protein-sol is based on the data of protein solubility in an *Escherichia coli* expression system (49). Using the ToxinPred server (28), the toxicity of the final vaccine constructs and each of its subunits was predicted. To distinguish between toxic and non-toxic, ToxinPred uses an SVM model based on a collection of 1805 toxic peptides (50).

2.11 Secondary structure prediction

The primary sequence of vaccines was deployed on NetSurfP-3.0 server (https://services.healthtech.dtu.dk/service.php?NetSurfP-3.0) to predict the secondary structure features of the constructed vaccines. To generate sequence embedding, this server exploits the ESM-1b language model, which is then processed through deep neural network. This server was also used to predict vaccine constructs solvent accessibility and disorder regions.

2.12 Three-dimensional structure modeling and refinement of vaccine

The RoseTTAFold tool (51) was employed to predict the 3-dimensional (3D) structure of the multi-epitope vaccine. Then, the predicted 3D model was submitted for structural refinement to the GalaxyRefine server (52). Using molecular dynamics simulation, GalaxyRefine carries out the correction of side chains and stabilizes the structure (53). The improved model's quality was assessed using the GDT-HA score, Molprobity score, clash score, RMSD score, and Ramachandran plot score. Finally, using ERRAT and ProSAweb servers, validation of our refined 3D model of vaccine construct was conducted (54, 55).

2.13 Molecular docking

MOE2020 (56) software was used for modeled vaccine-immune receptor docking using a protein-protein docking protocol of this software. A rigid body refinement method was applied, and the final 30 poses were retained. We chose several toll-like immune receptors (TLR) for docking, such as TLR2 (PDB ID: 2Z7X [https://www.rcsb.org/structure/2Z7X]), TLR3 (PDB ID: 2A0Z [https://www.rcsb.org/structure/2A0Z]), and TLR4 (PDB ID: 3FXI [https://www.rcsb.org/structure/3FXI]) and major histocompatibility complex (MHC) I (PDB ID: 1AKJ [https://www.rcsb.org/structure/1AKJ]) and II (PDB ID: 3L6F [https://www.rcsb.org/structure/3L6F]). Based on the lowest docking energy (high binding affinity) score, top docked complexes were subjected to further analysis. For the 2D interaction analysis, PDBsum web server was used (57) and Blender software (58) was used to illustrate the 3D structure of the docked complexes.

2.14 Optimization of codon and *in-silico* cloning

The residues sequence of the finalized vaccine was used as an input in EMBOSS Backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/) to obtain the cDNA of our constructs. The vaccine construct was then codon optimized using the Java Codon Adaptation Tool (JCat) service (59). This server provides results in terms of percent GC content and codon adaptation index (CAI), which were observed to assess the expression potential of protein. Also, at the N- and C-terminus of the vaccine codon sequence, cleavage sites for *Xho*I and *Nde*I enzymes were added. Using the SnapGene program (https://www.snapgene.com/), the optimized multi-epitope vaccine construct sequence was cloned between the *Xho*I and *Nde*I loci in the expression vector, pET28a (+).

2.15 Computational immune simulation

A computational immunological simulation was conducted *via* C-ImmSim server to assess the developed vaccine immunogenicity (60). This server employs machine learning algorithms and a position-specific scoring matrix (PSSM) to predict epitope and assess immunological interactivities (60). The immunological simulation was carried out according to the procedure described elsewhere (61). For 1050 simulation steps (about 12 months), three in-silico doses were given at suggested 4-week intervals and at time steps of 1, 84, and 170 (one time step is 8 hours of everyday living). The default settings were kept for all other triggering parameters.

2.16 Simulation study of the vaccine-TLRs complex

Molecular dynamics (MD) simulations of the multi-epitope vaccine/TLR complexes were performed using AMBER 20 (62) software with the ff19SB forcefield (63). Each system was initially solvated in a truncated octahedral box with 10Å buffer and an OPC water model before being neutralized with Na+ and Cl^- ions (64). The SHAKE method was used to limit all covalent bonds with hydrogen (65). The particle-mesh Ewald method (66) with a cut-off of 8Å was used to calculate long-range electrostatics. Using the PMEMD engine (67) on GPUs, parallel scaling in long-range electrostatics was enhanced. A continuum model was used to calculate the van der Waals long-range interactions. The LEaP module was also utilized to help find missing hydrogen atoms. Following these preparations, the complex systems were energy minimized in two stages (2000 steps steepest descent minimization; 10,000 steps conjugate gradients minimization) (68). In a microcanonical ensemble (NVE), each system was then heated from 0.1 to 300K in 400ps. A Langevin thermostat (69) and a collision frequency of 2.0ps^{-1} were used to control the kinetic energy of harmonic oscillators for dynamic propagation. The density was then changed using the same way in the 400ps run. In an NVE ensemble with no restriction and a pressure relaxation duration of 2ps, all systems were equilibrated at 300 K for 2000ps. Using the isotropic position scaling approach and a pressure relaxation duration of 1ps, the pressure was held constant during the equilibration. Finally, a 110ns MD simulation was run for each system, adopting the equilibration protocol in periodic boundary conditions.

2.16.1 Post dynamics assessment

The CPPTRAJ module of AMBER20 was used to analyze the output trajectory of vaccine-TLRs complexes. Using Cα atoms of each system, Root means square deviation (RMSD) and Root mean square fluctuation (RMSF) was computed as described elsewhere (67). In addition, the structural alteration during the simulation was quantified with the radius of gyration applying the equation (67). Further, the solvent-accessible surface area (SASA) was computed to study protein's surface characteristics. In order to compute all hydrogen bonds between vaccine and receptor complex, the threshold distance and angel were retained at 3.5Å and 120° between the Hydrogen bond donor and acceptor atoms.

2.16.2 Gibbs free energy distribution

The conformational free energy values of the complexes were studied at both the stable and transient stages. The CPPTRAJ package of AMBER 20 (67) was employed to investigate the systems' Free Energy Landscape (FEL). Employing PC1 and PC2 principal components, the trajectories data were split into 100 bins. The most fluctuating values are PC1 and PC2. As a result, bins with no population were artificially restricted to a population size of 0.5 during the free energy calculations. Free energy was calculated and expressed in kcal/mol at 300°C.

2.17 Estimation of binding free energy

Binding free energies of the complexes were estimated through Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method (70) using the below equation (71):

$$\Delta Gbind = \Delta GR + L - (\Delta GR + \Delta GR) \quad (3)$$

Protein-ligand complex energy is ΔGR + L, apo protein energy is ΔGR, and ligand energy is ΔGL. Each free energy term (ΔG) in the above equation was computed as described (72).

3 Results

3.1 Protein sequence selection

Initially the structural and non-structural proteins sequences of NeoCoV were retrieved from NCBI to formulate a multi-

epitope vaccine (Table S1). Structural proteins aid the virus in invading the host and synthesizing particles while non-structural proteins (NSPs) mediate the viral replications and structural protein synthesis. Followed by infection, these proteins trigger a distinct immunological response. Antigenicity, allergenicity, and transmembrane helices predictions were made on extracted sequences prior to their downstream analysis. All four structural proteins were found antigenic. Amongst NSPs, only ORF4a, ORF4b, ORF8b, ORF5b, viral protease, NSP10, ADP Binding module, and Coronavirus endopeptidase C30 were predicted as antigenic. With an antigenicity score of 0.61, structural protein 'Nucleocapsid' (N) was found as the most antigenic viral component of NeoCoV. The position and residue range of the viral protein sequences and vaccine constructs were determined using TMHMM Server. The selected sequence length of the envelope protein (39-50, 97-219), nucleocapsid protein (1-12), spike protein (1310-1344), and ORF4b (117-128, 171-284) were predicted inside the cell. All other selected components of the virus' structural and NSPs were localized outside the cell (Table 1).

3.2 Prediction and selection of linear B-cell epitopes

B-cell epitopes are antigen fragments that attach to antibodies or immunoglobulin, thereby cause B-cells to produce an immunological response (73). Primer for B-cell epitopes were predicted by BCPREDS for all antigenic protein components of NeoCoV, while BepiPred and ABCPred were employed to validate BepiPred's predictions. If any of the other two servers did not predict a B-cell epitope predicted by the primary server, it was discarded. The predicted epitopes were preferred based on their non-allergenicity, antigenicity, non-toxicity, and surface accessibility. Epitopes that showed antigenicity score ≥0.4 were chosen for vaccine formulation. For all viral proteins, a total of 92 epitopes were predicted, however, only 19 B-cell epitopes fulfilled the criteria which were selected for further analysis. Despite showing antigenicity, no B-cell epitopes for ORFa, ORFb, and NSP10 were selected due to not fulfilling the set criteria for prioritizing the epitopes (Table 2).

3.3 Prediction and selection of potential MHC class-I binding epitopes

The final list of identified epitopes for each viral antigenic component are given in Table 3. Briefly, 69 NeoCoV epitopes were predicted which showed significant binding interactions with a minimum of four HLA I subtypes. Prior to selection for further investigation, T cell epitopes were rated based on strong IEDB score, binding affinity (≥ 4 HLA1 subtypes), B-cell epitope overlap, considerable antigenicity, non-allergenicity, and non-toxicity. Following these criteria, eight epitopes for S-protein, four for M-protein, two for N-protein and ORF4b, and a single epitope for each E-protein and Coronavirus endopeptidase C30 were retained for vaccine design. The overall HLA score was computed for each epitope. Interestingly, these epitopes had never been investigated experimentally before, indicating that they are novel predictions. Upon applying the IEDB SMM-align method, most of the shortlisted epitopes showed high binding affinity (IC_{50} value >500nM) with their respective HLA supertype allele. The CTLPred server also confirmed these selected epitopes as potent Cytotoxic T-lymphocyte (CTL) epitopes following the combined approach as well as ANN and SVM thresholds of 0.51 and 0.36, respectively (Table S2).

3.4 Prediction and selection of potential MHC class-II binding epitopes

Using NetMHCIIpan version 3.2, a total of 66 putative MHC-II epitopes (15-mer) in the antigenic NeoCoV proteins were predicted to bind with at least three specific HLA DRB alleles (39). Before further analysis, epitopes were evaluated for strong binding affinity (≥ 3 HLA II subtypes), B-cell epitope overlap, considerably antigenicity, non-allergenicity, and non-toxicity. Following these specifications, seven epitopes for S-protein, four epitopes for N-protein, two each for N-protein, PLPro, NSP10, and ADP Binding module, and single epitopes for each ORF4a, ORF4b, and ORF5b were retained for vaccine design (Table 4). These epitopes showed IC_{50} >500 nM with their respective HLA II alleles using IEDB SMM-align method. Furthermore, these epitopes had never been investigated experimentally which confirms their predictions. The ability of the selected epitopes to stimulate IFN-γ secretion through helper T lymphocytes (HTLs or MHC-II epitopes) was determined by the IFNepitope server. The majority of shortlisted HTL epitopes were also predicted as IL4 and IL10 inducers (Table S3), thus, these epitopes can be classified as putative HLA II T-cell epitopes capable of activating CD4+ T-cells. Nine-mer epitopes of the final 15-mer peptides possess IC_{50} values >50nM with the DRB1*0101 allele (Table S4).

3.5 Population coverage analysis

The MHC-I and MHC-II epitopes have 90.73% and 82.22% worldwide coverage, respectively, according to the population coverage study. We also focused on their combined population coverage since a vaccine construct contains both types of MHC epitopes. The total combined coverage accounted for 80.42% of the world's population. For combined MHC class-I and II epitopes, Europe has highest population coverage (99.12%), followed by North America (97.19%), West Indies (96.42%), West Africa (90.91%), Southeast Asia (90.42%), North Africa

TABLE 1 NeoCoV proteins' antigenicity prediction (With a viral model, threshold = 0.4). The TMHMM Server v2.0 was used to predict transmembrane helices in proteins.

Viral Component	Position	Aminoacids	VaxiJen Score	Probability	Localization
E-Protein	outside	1-19	0.4707	Non-allergen	Cell membrane
	TMhelix	20-38			
	inside	39-50			
	TMhelix	51-70			
	outside	71-73			
	TMhelix	74-96			
	inside	97-219			
M-Protein	outside	1-414	0.4879	Non-allergen	Endoplasmic reticulum, Membrane
N-Protein	inside	1-12	0.6193	Non-allergen	Cytoplasm, Soluble
	TMhelix	13-35			
	outside	36-82			
S-Protein	outside	1-1286	0.4988	Non-allergen	Cell membrane
	TMhelix	1287-1309			
	inside	1310-1344			
ORF4a	outside	1-103	0.4405	Non-allergen	Extracellular, Soluble
ORF4b	outside	1-109	0.5082	Non-allergen	Nucleus, Soluble
ORF8b	outside	1-199	0.4997	Non-allergen	Peroxisome, Membrane
ORF5b	outside	1-93	0.4567	Non-allergen	Endoplasmic reticulum, Membrane
	TMhelix	94-116			
	inside	117-128			
	TMhelix	129-146			
	outside	147-150			
	TMhelix	151-170			
	inside	171-284			
Protease (PLPro)	outside	1-313	0.4811	Non-allergen	Extracellular, Soluble
NSP10	outside	1-120	0.6372	Non-allergen	Extracellular, Soluble
ADP Binding module	outside	1-101	0.5731	Non-allergen	Mitochondrion, Soluble
Coronavirus endopeptidase C30	outside	1-278	0.575	Non-allergen	Extracellular, Soluble

(87.71%), South Asia (87.34%), Northeast Asia (82.64%), East Africa (81.84%), East Asia (81.48%), South America and Oceania (78.41% each), Southwest Asia (77.24%), South Africa (73.48%), Central Africa (74.52%), and Central America (27.46%). The relative population coverages of individual MHC classes and combined MHC epitopes are shown in Figure 2.

3.6 Physiochemical profile of epitopes

Using the Expasy ProtParam server, the physicochemical profile of final peptides was evaluated (Tables S3, S5 and S6). The instability index scores of 12 MHC class I and 18 MHC class II peptides predicted them as stable in the test tube (>40). Similarly, the high aliphatic index value showed that 13 MHC-I and 20 MHC-II final epitopes are thermostable. Likewise, the negative GRAVY score shows that most MHC-I epitopes and all MHC-II epitopes are hydrophilic. The projected half-life in mammalian cells for CTL epitopes ranged from 1.1 hr to 30 hrs, whereas for HTL epitopes, it was 1.1h to >20hrs. Most of the final CTL and HTL peptides were computed as acidic (isoelectric point< 7), and respectively their molecular weight ranged from 880.95–1115.40 Da and 1445.5–1826.05 Da. Furthermore, 13/19 B-cell epitopes were stable, 13 were predicted thermostable, and the majority indicated the possibility of interaction with water. The estimated half-life in mammalian cells ranged from 1h-30hrs, and most were projected as acidic with molecular weight varying from 1683.71 to 2363.73 Da.

3.7 Vaccine construction

To stimulate an immune response against NeoCoV, four vaccines were developed leveraging the prioritized epitopes. Three adjuvants, β-defensin, 50S ribosomal protein L7/L12, and HBHA protein, were used to create four distinct vaccines (Neo-1–4). The PADRE sequence is a pan HLA-DR epitope peptide used to improve vaccination efficacy while minimizing toxicity. As a potential immunogen, the PADRE sequence

10.3389/fimmu.2022.956776

TABLE 2 Selection of B-cell epitopes. Twenty amino acid-long B-cell epitopes were predicted with BCPREDS and validated with BepiPred and ABCPred servers.

Viral Component	Pos.	Epitopes	B-score	AT-score	A-score	Allergenicity	Toxicity
E	61	*TG**RS**VYVKFQ**ESKPPLPP**EE*	0.99	0.54	0.83	NA	NT
M	1	**MSNM*TQLSEQQIIAIIKDWN***	0.89	0.50	0.82	NA	NT
N	192	***GNSSRGASPGPSGVGAPG*GD**	1	0.54	0.85	NA	NT
	20	**TNQPRG*RGRNPKPRAAPNTT***	1	0.50	0.89	NA	NT
	361	**PKKEKK*QKAPKEESNDQEMA***	1	0.41	0.82	NA	NT
	124	***EEGATDAPSTFGTRN*PNNDS**	0.99	0.57	0.74	NA	NT
S	1207	***NNLPPPLLSNSTGTD*FKDEL**	0.99	0.67	0.82	NA	NT
	17	*ANAKI**VTLPGNDATGY**CPSV*	0.99	0.53	0.8	NA	NT
	532	***NSPTT**GQLWAYN**FGGVPYRV***	0.97	0.53	0.82	NA	NT
	84	***DLG**TQYVYSASN**HK**STANDA*	0.97	0.56	0.73	NA	NT
	506	***NYGATNKDDVV**KPGGRASQQ*	1	0.46	0.88	NA	NT
	189	***LLQPRTESKCPGNSN**YVSYF*	0.96	0.80	0.73	NA	NT
	842	*ESV**KTPQTVPLTTGFGG**EF**N*	0.73	0.47	0.69	NA	NT
ORF8b	59	***LGIGGDRTERLTQEMELSNW***	0.99	0.98	0.67	NA	NT
ORF5b	9	***KPVQLVPVSPVDHGGE**SNDS*	0.99	0.90	0.91	NA	NT
PLPro	286	*KF**DSGTLSKASD**WKCKVTDV*	0.83	0.93	0.69	NA	NT
	32	FFNGA**DIS*DTIPDEKQHGCS***	0.8	0.45	0.9	NA	NT
ADP Binding Module	29	*GAVQQ**ESDEYILTRGPLQVG***	0.93	0.45	0.76	NA	NT
Coronavirus endopeptidase C30	30	*LVSMTNHSFS**VQKHVGAP**AN*	0.87	0.57	0.79	NA	NT

Full length epitope is the prediction of BCPREDS. Bold letters show the BepiPred predicted epitope and letters written in italics are the ABCPred predicted epitope. Pos, Position; B-score, BCPred predicted score; AT-score, VexiJen antigenicity score; A-score, ABCpred predicted score; NT, Non-toxic; NA, Non-allergen. The projected epitope with an antigenicity score of ≥ 0.4 was considered antigenic.

enables the binding of CTL epitopes with various class II MHC molecules with high affinity (44). During the design phase, the selected epitopes were also conjugated with the appropriate linkers. The final vaccine length of Neo-1–4 is 358, 438, 448, and 415 bp, respectively. The sequence and length of the four vaccine constructs are presented in (Table 5).

3.8 Assessment of antigenicity, allergenicity, toxicity, and surface accessibility

Besides various adjuvants employed in the design, all vaccine constructs were tested for antigenic property, safety, allergenicity, and surface accessibility (Table 6). The selected sequence length of Neo-1 (1-49), Neo-2 (1-150), and Neo-4 (1-177) were predicted inside the cell. All other selected components of these constructed vaccines as well as the complete sequence length of Neo-3, were localized outside the cell. The antigenicity score of the final vaccine constructs ranged from 0.44–0.56 and 0.92–0.97 according to the VexiJen and ALLERGENPro predictions, respectively, indicating that our final vaccine constructs have high antigenicity. The final vaccine constructs and its components were predicted to be non-allergenic by the AllergenFP 1.0 and the AllerTOP 2.0 server. Moreover, the ToxinPred server revealed our constructed vaccines and every subunit as non-toxic.

3.9 Vaccine physiochemical properties evaluation

The numerous vaccine components and their individual properties are listed in Table 6. The projected values indicate that our final vaccines and each subunit are well soluble. The molecular weight of four vaccine constructs [Neo-1 (37678.63 Da), Neo-2 (45770.41 Da), Neo-3 (47868.55 Da), Neo-4 (44042.97 Da)] is less than 110 kDa, which is regarded as suitable for vaccine development (74). The theoretical isoelectric point (pI) of four vaccines construct ranged from 8.57 to 9.56, which falls within the normal pH range. The instability index of Neo-1–4 was ≤40; thus, we can expect our vaccine protein is stable. Our vaccines Neo-1–4 have an aliphatic index of 60.67, 73.63, 69.62, and 75.98, respectively, which is a significant indicator of thermostability at different temperatures. The grand average of hydropathicity (GRAVY) for the four constructed vaccines was negative, indicating the vaccines are hydrophilic in nature (75). The half-life of all the designed vaccines was determined in mammalian immature red blood cells (30 hours, *in vitro*), in yeast (>20 hours, *in vivo*), and *Escherichia coli* (>10 hours, *in vivo*). The instability index of Neo-1-4 (>40) suggest their stability in test tube. Additionally, BLASTp against the human genome revealed that the sequences had no human counterpart. Proteins that are human homologs may induce autoimmunity unexpectedly

10.3389/fimmu.2022.956776

TABLE 3 List of prioritized MHC-I (CTL) binding epitopes by NetCTL 1.2 server.

Protein	MHC-I Epitopes	No. of HLASuper-types	HLA supertypes (IC_{50} nM)	Total HLAscore	AT	Allergenicity	Toxicity	Immunogenicity
E	FTVVCAITL	4	A2 (380.29), B8 (135), B39 (245.13), B62 (95.11)	3.628	1.05	NA	NT	0.17427
M	GTNSGVAIY	4	A1 (77.11), A3 (259), A26 (133), B62 (142.97)	5.428	0.46	NA	NT	0.01407
	ALSIFSAVY	4	A1 (282), A3 (68.46), B58 (101.2), B62 (170.70)	4.529	0.61	NA	NT	0.09563
	YPSRSMTVY	5	A1 (246.49), A26 (273.6), B7 (529.53), B8, B62 (66.87)	4.734	0.92	NA	NT	0.27114
	LLITIVLQY	5	A1 (126.26), A3 (580.50), A26 (111.73), B58 (70.81), B62 (567.85)	5.823	0.87	NA	NT	0.17368
N	STPAQNAGY	5	A1 (122.99), A3 (874.13), A26 (187.76), B58 (811.5), B62 (98.00)	7.243	0.44	NA	NT	0.0303
	SAFMGMSQF	5	A26 (189.00), B7 (351.4), B8 (669.1), B58 (408.43), B62 (27.24)	3.793	0.44	NA	NT	0.4783
S	WSYTGSSFY	5	A1 (207), A3 (405.32), A26 (128.17), B58 (415.24), B62 (86.49)	6.884	1.06	NA	NT	0.15609
	YSTNITHLL	5	A1 (96.5), A2 (97.43), B39 (700.486), B58 (351.8), B62 (479.036)	4.884	0.62	NA	NT	0.19305
	ISYAGAYSY	4	A1 (766.56), A3 (404.38), B58 (40.114), B62 (59.841)	5.841	0.80	NA	NT	0.00822
	SVTIADPGY	4	A1 (112.5), A26 (436.77), B58 (385.76), B62 (112.78)	3.538	1.07	NA	NT	0.01912
	ALQEVVKAL	4	A2 (195.934), B7 (299.5), B8 (307.99), B62 (401.678)	3.634	0.52	NA	NT	0.01693
	TMKKIYPAL	5	A2 (110.358), A24 (401.48), B8 (119), B39 (103.48), B62 (615.176)	4.943	0.52	NA	NT	0.14738
	MVYVITVKY	4	A3 (143.81), A26 (636.223), B58 (311.386), B62 (402.71)	4.385	0.91	NA	NT	0.11532
	FLFATVPIY	5	A3 (509.09), A26 (488.03), B39 (266.701), B58 (266.701), B62 (46.45)	6.127	0.56	NA	NT	0.22243
ORF4b	HSPGKNLRY	5	A1 (491.53), A3 (430.28), A26 (220.95), B58 (274.79), B62 (410.42)	5.421	0.43	NA	NT	0.16471
	SVVTQPTHY	4	A1 (532.9), A3 (344.9), A26 (126), B62 (128.60)	5.572	0.53	NA	NT	0.01912
Coronavirus endopeptidase C30	ASFSVLACY	5	A1 (114.33), A3 (91.58), A26 (339.2), B58 (300.82), B62 (127.94)	5.841	0.62	NA	NT	0.09719

AT, Antigenicity score (HLA supertype with binding affinity scores ≥ 0.75 were taken); NA, Non-allergen; NT, Non-toxic.

due to off-target reactions with the host proteins (76). Overall, the physicochemical examination of the polypeptide-based constructs reveals that these are appropriate for immunological applications.

3.10 Human homology and novelty analysis

The selected B-cell, CTL, and HTL epitopes and the final vaccine constructs, were subjected to BLASTp homology search against human proteome (taxid: 9606). The query coverage of shortlisted epitopes and vaccine constructs did not significantly resemble human protein sequences (Neo-1 = 7%, Neo-2 = 21%, Neo-3 = 21%, Neo-4 = 14%). Thus, the predicted vaccines would not elicit autoimmune reactions in the host. Besides, all prioritized B-cell, MHC-I, and MHC-II epitopes were indicated as novel prediction upon checking the IEDB database.

3.11 Prediction of secondary structure

Using the NetSurfP - 3.0 server, we predicted the secondary structure of our final vaccines. According to the projected secondary structure, the Neo-1 vaccine have 15.59% α-helix,

10.3389/fimmu.2022.956776

TABLE 4 List of prioritized MHC-II (HTL) binding epitopes by NetMHCIIpan.

Protein	MHCII Epitopes	No. of HLA alleles	HLA alleles/IC50 value (nM)	AT	Allerg.	Tox.
M	LPNEITVAKPNVLIA	3	DRB1*0101/(10.96), DRB1*0701/(27.53), DRB1*1501/(111.34)	0.41	NA	NT
	LIALKMVKRQSYGTN	3	DRB1*0301/(180.49), DRB1*1101/(46.15), DRB1*1301/(17.3)	1.00	NA	NT
N	TKSFNMVQAFGLRGA	3	DRB1*0101/(3.56), DRB1*0701/(7.18), DRB1*1501/(31.54)	0.82	NA	NT
	PKVITKKDAAAAKNK	6	DRB1*0101/(213.03), DRB1*0401/(116.37), DRB1*0801/(105.17), DRB1*1101/(77.54), DRB1*1301/(212.59), DRB1*1501/(115.52)	0.75	NA	NT
	SGAIKLDPKNPNYNK	4	DRB1*0301/(96.66), DRB1*0801/(177.66), DRB1*1101/(69.44), DRB1*1301/(79.35)	1.11	NA	NT
	PRWYFYYTGTGPEAA	4	DRB1*1501/(356.18), DRB1*0401/(155.66), DRB1*0701/(90.79), DRB1*0801/(233.88)	0.76	NA	NT
S	STSYYSAKPVGAYYE	5	DRB1*0101/(6.26), DRB1*0401/(52.57), DRB1*0701/(88.17), DRB1*0801/(274.17), DRB1*1101/(96.90)	0.43	NA	NT
	PEPITTLNTRYVAPQ	5	DRB1*0101/(36.29), DRB1*0401/(514.57), DRB1*0701/(422.39), DRB1*1301/(126.66), DRB1*1501/(119.58)	0.82	NA	NT
	ISYDIYGITGTGVFQ	3	DRB1*0101/(19.77), DRB1*0701/(56.03), DRB1*1501/(155.33)	0.67	NA	NT
	YVAGYKVLPPLMDVN	3	DRB1*0101/(18.05), DRB1*0401/(81.41), DRB1*1101/(88.74)	0.48	NA	NT
	GTQYVYSASNHKSTA	4	DRB1*0101/(34.89), DRB1*0401/(96.98), DRB1*0701/(153.93), DRB1*1101/(156.61)	0.45	NA	NT
	TQYVYSASNHKSTAN	3	DRB1*0101/(34.76), DRB1*0401/(96.43), DRB1*1501/(225.61)	0.42	NA	NT
	IIGFHSDDGNYYCVA	3	DRB1*0301/(112.99), DRB1*0401/(258.69), DRB1*1101/(493.8)	0.46	NA	NT
ORF4a	TAKYTPAPGTSLHPV	3	DRB1*0101/(17.07), DRB1*0401/(109.35), DRB1*0701/(448.9)	0.71	NA	NT
ORF4b	ARDISPIAVFLRNVR	3	DRB1*0301/(28.77), DRB1*1301/(368.61), DRB1*1501/(91.19)	1.04	NA	NT
ORF5b	STVFVPATRDSVPLH	3	DRB1*0301/(436.22), DRB1*0801/(334.17), DRB1*1101/(41.04)	0.59	NA	NT
PLPro	SPDFVAFNVFHGMET	6	DRB1*0101/(42.9), DRB1*0401/(84.87), DRB1*0701/(137.13), DRB1*0801/(194.89), DRB1*1101/(173.16), DRB1*1501/(52.95)	0.71	NA	NT
	FRTVVLNNKNSYRSQ	3	DRB1*0301/(65.99), DRB1*1101/(180.05), DRB1*1501/(21.08)	0.83	NA	NT
NSP10	KGKFVQIPSQCTRDP	5	DRB1*0101/(7.91), DRB1*0401/(34.54), DRB1*070I/(61.37), DRB1*0801/(191.62), DRB1*1101/(88.24)	0.57	NA	NT
	GTGIAISVKPESTAD	4	DRB1*0301/(102.9), DRB1*0801/(75.29), DRB1*1101/(350.6), DRB1*1301/(350.6),	1.21	NA	NT
ADP	SKCYRAMNAYPLVVT	3	DRB1*0101/(2.95), DRB1*0701/(12.61), DRB1*1501/(30.76)	0.70	NA	NT
Binding module	AKNILHVVGPDARAK	3	DRB1*0101/(12.5), DRB1*0301/(897.1), DRB1*1501/(65.47)	0.66	NA	NT
CoV endopeptidase C30	QQLYTGFQGKQILGS	3	DRB1*0101/(51.5), DRB1*0701/(118.1), DRB1*1501/(220.03)	0.41	NA	NT
	TGTFTVIMRPNYTIK	3	DRB1*0801/(72.77), DRB1*1101/(22.62), DRB1*1301/(25.63)	0.77	NA	NT

Predicted IC_{50} value by the IEDB tool (SMM method) is shown in parenthesis for each epitope with the respective HLA allele. AT, Antigenicity score; Allerg, Allergenicity; Tox, Toxicity (HLA allele with binding affinity scores≥ 0.75 were taken); NA, Non-antigenic; NT, Non-toxic. * is a part of nomenclatures indicating the method is molecular typing.

13.4% β-strand, and 70.75% coil. The Neo-2 and Neo-3 vaccine showed 9.13% and 10.26% α-helix, 22.6% and 18.97% β-strand, 68.27% and 70.75% secondary coil structure, respectively. The secondary structure of the Neo-4 comprises of 15.42% α-helix, 35% β-strand, and 62.65% coil. In addition, solvent accessibility (ACC) and disorder areas (DISO) of Neo-1-4 were also estimated. Neo-1 has 358 residues, 94% of which are projected to be exposed and only 6% are buried, while 60 residues (16.71%) are expected to be in disordered regions. Neo-2 vaccine construct is predicted to have 83.10% of residues exposed and 16.89% hidden, whereas 88 residues (20%) of Neo-2 are expected to lie in disordered regions. Similarly, in Neo-3, 89.5% of residues are projected to be exposed and 10.5% to be buried, and 47 residues (10.49%) are anticipated to be in disordered regions. Likewise, 83.17% residues of Neo-4 are exposed and 16.89% hidden, while 80 residues (17.70%) are expected to locate in disordered regions. The predicted secondary structure features, relative surface accessibility, and disorder regions of Neo-1–4 is depicted in Figure 3 and Figure S1.

3.12 Tertiary structure modeling and refinement of constructed vaccine

The RoseTTAFold tool and GalaxyRefine server were employed to predict and improve the modeled structures of the final vaccine constructs. We picked model 1 for Neo-1, 2,

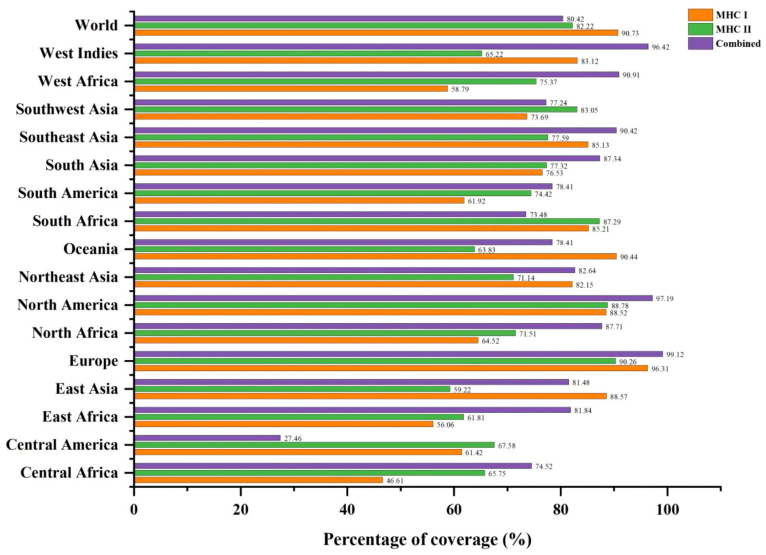


FIGURE 2
Global population coverage of selected T-cell epitopes (MHC-I, MHC-II, and combined MHC) based on their corresponding HLA binding alleles.

and 4 and model 4 for Neo-3 based on the model quality of the five refined models projected by GalaxyRefine (Table S7). The ERRAT quality score, ProSA Z-score, and Ramachandran plot analysis were used to verify the overall model quality of the refined vaccine structures. ERRAT gave an overall quality score of 81.26, 95.87, 97.18, and 90.41 for Neo-1 to 4, respectively. The ProSA predicted Z-score of Neo-1 to 4 were -6.08, -6.58, -8.32, and -5.72, respectively, which is within the scoring range of similar-sized native proteins, suggesting high overall model quality. Using ProSA, we further checked the quality of the local model, and the residue scores are presented in Figure 4. The negative values indicate that the model structure is free of errors. We further used the PROCHECK server (77) for Ramachandran plot analysis. The refined model structure of Neo-1contains 81.6%, 13%, and 1.4% of residues in most favored, additionally allowed, and generously allowed regions, respectively, whereas only 4% of residues lie in disallowed regions. Neo-2 refined model has collectively 99.4% residues in allowed regions, while only 0.6% of residues lie in the disallowed region. Similarly, Neo-3 and Neo-4 refined modelled structures has 98.5% and 99.2% of residues in allowed regions and 1.5% and 0.8% in disallowed regions, respectively (Figure 4 and Table S8). The final 3D-model of each vaccine construct is illustrated in Figure S2.

3.13 Prediction of conformational B-cell epitopes

The new protein's structure and folding can lead to new conformational B-cell epitopes, prompting further predictions. Such epitopes in the refined modeled structure were predicted through the ElliPro server (34). Tables S9, S10 enlists predicted discontinuous B-cell epitopes for all four vaccine constructs. Briefly, seven conformational B-cell epitopes were predicted for Neo-1 involving a total of 183 residues with a score ranging from 0.562 to 0.796. For Neo-2, eight conformational B-cell epitopes were predicted, which involved 212 residues with score in range of 0.526 to 0.981. Similarly, predicted conformational B-cell epitopes for Neo-3 (score ranged from 0.575 to 0.817) and Neo-4 (score ranged from 0.512 to 0.769) involved 219 and 211 residues, respectively.

3.14 Molecular docking analysis

The interactions between the antigenic molecule and immune receptor molecule at the molecular level are critical to successfully activating the transport of antigenic molecule and immune response (78, 79). To study possible interactions and binding energy, docking was performed between several immune receptors (TLR2, TLR3, TLR4, MHC-I and II) and developed vaccines (Neo-1–4). TLRs can elicit an immunological response upon viral recognition. We selected the best-docked complex among 30 docked poses based on the lowest energy score, indicating that the constructed vaccines fit in receptor efficiently with strong binding affinity. Amongst all TLRs, TLR3 showed the best binding with Neo-4 and Neo-1 with binding energies of -114.47 kcal/mol and -101.08 kcal/mol, respectively, followed by Neo-2 (-91.88 kcal/mol) and Neo-3 (-81.01 kcal/mol). For all these TLRs, Neo-1 and Neo-4

10.3389/fimmu.2022.956776

TABLE 5 The predicted vaccine constructs for NeoCoV.

Vaccine Name	Adjuvant	Epitope count	Length	Vaccine Constructs
NeoCoV vaccine- 1 (Neo-1)	β-defensin	HTL (6) CTL (5) B-cell (5)	358	GIINTLQKYYCRVRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAK**AKFVAAWTLKAAA**GGGSVFTVVCA ITLAAYGTNSGVAIYAAYSTPAQNAGYAAYWSYTGSSFYAAYHSPGKNLRYGGGSLPNEITVAKPNVLIAGP GPGTKSFNMVQAFGLRGAGPGPGSTSYYSAKPVGAYYEGPGPGTAKYTPAPGTSLHPVGGPGSTVFVPATR DSVPLHGPGPGSPDFVAFNVFHGMETGGG*STGRSVYVKFQESKPPLPPEE*KK*MSNMTQLSEQQIIAIIKDWN*K K*GNSSRGASPGPSGVGAPGGD*KKN*NLPPPLLSNSTGTDFKDEL*KKLGIGGDRTERLTQEMELSNW*HHHHHH
NeoCoV vaccine- 2 (Neo-2)	Ribosomal protein	HTL (6) CTL (5) B-cell (5)	438	MAKLSTDELLKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKI GVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVKEAAAK**AKFVAAWTLK AAA**GGGSASFSVLACYAAYALSIFSAVY**AAY**SAFMGMSQFAAYSAFMGMSQFAAYYSTNITHLLGGGSKGKF VQIPSQCTRDPGPGPGSKCYRAMNAYPLVVTQQLYTGFQGKQILGSGPGPGLIALKMVKRQSYGTNGPGPG PKVITKKDAAAAKNKGPGPGPEPITTLNTRYVAPQGGGS*KPVQLVPVSPVDHGGESNDSKK*K*FDSGTLSKASD WKCKVTDV*KK*GAVQQESDEYILTRGPLQVGKKL*VSMTNHSFSVQKHVGAPAN*KK*TNQPRGRGRNPKPRAAPN TT*HHHHHH
NeoCoV vaccine- 3 (Neo-3)	Heparin-binding hemagglutinin	HTL (6) CTL (3) B-cell (5)	448	MAENPNIDDLPLAALGAADLALATVNDLIANLRERAEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFT TEELRKAAEGYLEAATNRYNELVERGEAALQRLRSQTAFEDASARAEGYVDQAVELTQEALGTVASQTRAV GERAAKLVGIELEAAAK**AKFVAAWTLKAAA**GGGSLLITIVLQYAAYISYAGAYSYAAYSVVTQPTHYGGGS ARDISPIAVFLRNVRGPGPGFRTVVLNNKNSYRSQGPGPGGTGIAISVKPESTADGPGPGAKNILHVVGPDA RAKGPGPGTGTFTVIMRPNYTIKGPGPGSGAIKLDPKNPNYNKGGGS*PKKEKKQKAPKEESNDQEMA*KK*AN AKIVTLPGNDATGYCPSV*KK*FFNGADISDTIPDEKQHGCS*KK*EEGATDAPSTFGTRNPNND*SKK*NSPTTGQLWA YNFGGVPYRV*HHHHHH
NeoCoV vaccine- 4 (Neo-4)	Ribosomal protein	HTL (6) CTL (5) B-cell (4)	415	MSDINKLAENLKIVEVNDLAKILKEKYGLDPSANLAIPSLPKAEILDKSKEKTSFDLILKGAGSAKLTVVKRIKD LIGLGLKESKDLVDNVPKHLKKGLSKEEAESLKKQLEEVGAEVELKEAAAK**AKFVAAWTLKAAA**GGGSSVTI ADPGYAAYALQEVVKALAAYTMKKIYPALAAYMVYVITVKYAAYFLFATVPIYGGGSPRWYFYYTGTGPEA AGPGPGISYDIYGITGTGVFQGPGPGYVAGYKVLPPLMDVNGPGPGGTQYVYSASNHKSTAGPGPGTQYVY SASNHKSTANGPGPGIIGFHSDDGNYYCVAGGGS*DLGTQYVYSASNHKSTANDA*KK*NYGATNKDDVVKPGG RASQQ*KK*LLQPRTESKCPGNSNYVSYF*KK*ESVKTPQTVPLTTGFGGEFN*HHHHHH

The sequences highlighted in bold are the pan HLA DR-binding epitope (PADRE), whereas the italics sequences are the B-cell epitopes.

exhibited lower binding energies than Neo-2 and Neo-3. Furthermore, Neo-1–4 displayed good interaction energy (ranging from-76.52 to-85.43 kcal/mol) with both classes of MHC molecules. According to the docking study, the proposed vaccines and selected immune receptors have significant interactions and successful binding. The docking calculations are given in Table 7.

Interaction analysis revealed strong molecular contacts between the designed vaccines (Neo-1 and Neo-4) and TLR3. A total of 31 residues of the Neo-1 construct interacted through hydrogen bonds with TLR3 residues, which are crucial to maintaining complex stability. Neo-1 residues that mediated hydrogen bond interaction with TLR3 with a bond distance ≤ 2.75Å are: Thr32, Arg33, Lys118 (2×), Lys174, Pro185, Thr187, and Lys189. Besides, the constructed vaccine residues that formed a hydrogen bond with TLR3 residues with a bond distance in the range of 2.76–3Å include Arg12, Ala16, Arg33, Pro92, Ala93, Ser103, Ser107, His114 (2×), and Ser115, Pro143, Gly146 (2×), Pro163, Lys174, Gly182, Pro183 (2×), Lys189, and Gly195. With a bond distance >3Å, Neo-1 residues, such as Arg12, Lys189, and Tyr180, also showed hydrogen bonding with TLR3 residues. In addition, four residues of Neo-1, including Arg33, Lys118, Lys174, and Lys189(2×), formed salt bridges with TLR3 residues (bond distance in the range of 2.64Å to 2.76Å). Detailed interactions of Neo-1 construct with TLR3 receptor are displayed in Figure 5 and Table S11.

Similarly, Neo-4 construct mediated 37 hydrogen bonds with TLR3 residues. Neo-1 residues, including, Lys71 (3×), Asp75, Lys82, Thr185, Lys276 (2×), Ala279, Lys295, Asn299, Asp324, and His415, showed hydrogen bond with TLR3 residues with a bond distance ≤ 2.75Å. Moreover, Lys74 (3×), Lys82, Ser204, Asn274 (2×), Lys276, Thr278, Lys295, Pro281, Gly282, Pro283, Asp324, Gln396, Thr401, Asn409 (2×), His412, His413, and His415 residues of Neo-1 formed a hydrogen bond with TLR3 residues with a bond distance 2.76–3Å. Neo-1 residues that showed hydrogen bonding with TLR3 with a bond distance of>3Å are Lys71, Asp324, and Gln396.

Furthermore, the Neo-4 construct showed a total of 10 salt bridges interactions with TLR3 residues through Lys71, Asp75, Lys82(2×), Lys276(2×), Lys295(2×), Asp324, and His412 residues (bond distance in the range of 2.66Å to 2.79Å). Detailed interactions of the Neo-4 construct with TLR3 receptor are displayed in Figure 5 and Table S12.

3.15 Vaccine optimization and *in-silico* cloning

An efficient vaccine expression in the *E. coli* system is required for computational cloning (44). The prioritized B-cell and T-cell predicted epitopes and suitable adjuvants and linkers were used to create four vaccines (Neo-1–4). These vaccine

10.3389/fimmu.2022.956776

TABLE 6 Allergenicity, antigenicity, toxicity, surface accessibility, and physicochemical characteristics of proposed polypeptide-based vaccine constructs (Neo-1-4).

Properties	Neo-1	Neo-2	Neo-3	Neo-4
Sol-Pro	0.946675	0.993184	0.869426	0.757577
Protein Sol	0.517 (soluble)	0.641	0.517	0.515
Molecular weight	37678.63 Da (avg.)	45770.41 Da	47868.55 Da	44042.97 Da
Formula	$C_{1676}H_{2596}N_{472}O_{496}S_{12}$	$C_{2038}H_{3250}N_{562}O_{608}S_{13}$	$C_{2099}H_{3329}N_{615}O_{658}S_5$	$C_{1991}H_{3095}N_{521}O_{595}S_6$
Theoretical pI	9.56(basic)	9.47 (basic)	8.97 (basic)	9.15 (basic)
Ext. coefficient	47,705 M^{-1} cm^{-1}	30,620 M^{-1} cm^{-1}	34, 965 M^{-1} cm^{-1}	52,845 M^{-1} cm^{-1}
Instability index	35.94 (stable)	29.00 (stable)	33.82 (stable)	25.29 (stable)
Aliphatic index	60.67 (thermostable)	73.63 (thermostable)	69.62 (thermostable)	75.98(thermostable)
Grand average of hydropathicity (GRAVY)	-0.421(hydrophilic)	-0.245 (hydrophilic)	-0.586 (hydrophilic)	-0.347 (hydrophilic)
Half-Life (satisfactory)	30 hrs (Mammalian reticulocytes, *in vitro*). >20 hrs (yeast, *in vivo* >10 hrs (*Escherichia coli*, *in vivo*).	30 hrs (Mammalian reticulocytes, *in vitro*). >20 hrs (yeast, *in vivo*). >10 hrs (*Escherichia coli*, *in vivo*).	30 hrs (Mammalian reticulocytes, *in vitro*). >20 hrs (yeast, *in vivo*). >10 hrs (*Escherichia coli*, *in vivo*).	30 hrs (Mammalian reticulocytes, *in vitro*). >20 hrs (yeast, *in vivo*). >10 hrs (*Escherichia coli*, *in vivo*).
Allergenicity	AllerTop (Non-allergen), Allergen FP (Non-allergen)	AllerTop (Non-allergen), Allergen FP (Non-allergen)	AllerTop (Non-allergen), Allergen FP (Non-allergen)	AllerTop (Non-allergen), Allergen FP (Non-allergen)
Antigenicity	ANTIGENpro (0.97), VexiJen (0.56)	ANTIGENpro (0.95), VexiJen (0.52)	ANTIGENpro (0.93), VexiJen (0.53)	ANTIGENpro (0.92), VexiJen (0.44)
Toxicity	Non-toxic	Non-toxic	Non-toxic	Non-toxic
Surface accessibility	Position: residues Inside: 1-49 TMhelix: 50-72 Outside: 73-358	Position: residues Inside: 1-150 TMhelix: 151-173 Outside: 174-478	Outside	Position: residues Inside: 1-177 TMhelix: 178-200 Outside: 201-415

sequences were submitted separately in the Backtranseq tool to modify the codon used to most sequenced prokaryotic species. The cDNA sequence of Neo-1–4 was 1089, 1326, 1428, and 1257 nucleotides, respectively, according to this analysis. The obtained CAI (Neo-1 = 0.94; Neo-2 = 0.99; Neo-3 = 0.97; Neo-4 = 0.95) showed that the modified sequences are comprised of codons capable of using the target organism's cellular machinery. Also, the modified sequences had a GC content ranging from 50.91% to 53.76%. This sequence information indicates that the proposed vaccines will be expressed efficiently and consistently in *E. coli*. The cleavage sites for *Xho*I and *Nde*I enzymes were added at the N and C termini of Neo-1–4 cDNA (Figure S3), and the final optimized sequences were then inserted into the pET28a (+) vector using the SnapGene software to clone the proposed vaccines. For Neo-1–4, the cloned plasmids had final length of 6373, 6610, 6640, and 6541 bp, respectively (Figure 6).

3.16 Immunostimulatory response of the vaccine constructs

To investigate the formation of adaptive immunity as well as immunological interactions, an immune simulation study was conducted for Neo-1 and Neo-4 vaccine complexes. The results of immunological simulation after three *in silico* injections of the vaccine constructs are shown in Figure 7, S4 and S5. A higher level of IgM was identified as a primary reaction in the simulation outcomes. Moreover, increased antibody activity levels were found in secondary and tertiary reactions, comprising IgG1+IgG2, IgM, and IgM+IgG antibodies. Subsequently, there was faster antigen clearance (reduced antigen load). As evidenced by higher quantities of simulated B-cells and memory B-cells growth, the vaccine elicited a substantial long-lasting immune response.

The growing rate of T-cells (including helper, cytotoxic, and regulatory T-cells) were seen, that are associated with relative immunity development. In addition, dendritic and macrophage cells showed higher concentrations, indicating good antigen presentation. Furthermore, the population of natural killer cells remained steady during immune simulation. Finally, different cytokines, notably higher quantities of the cytokine IFN-γ (over 400,000 ng/ml) were also detected. Those molecules evoke immune properties to ensure that the vaccine can succeed in human body.

3.17 Molecular dynamics simulation

The root means square deviation (RMSD) of the Cαatoms was estimated for two vaccine complexes. The average RMSD of

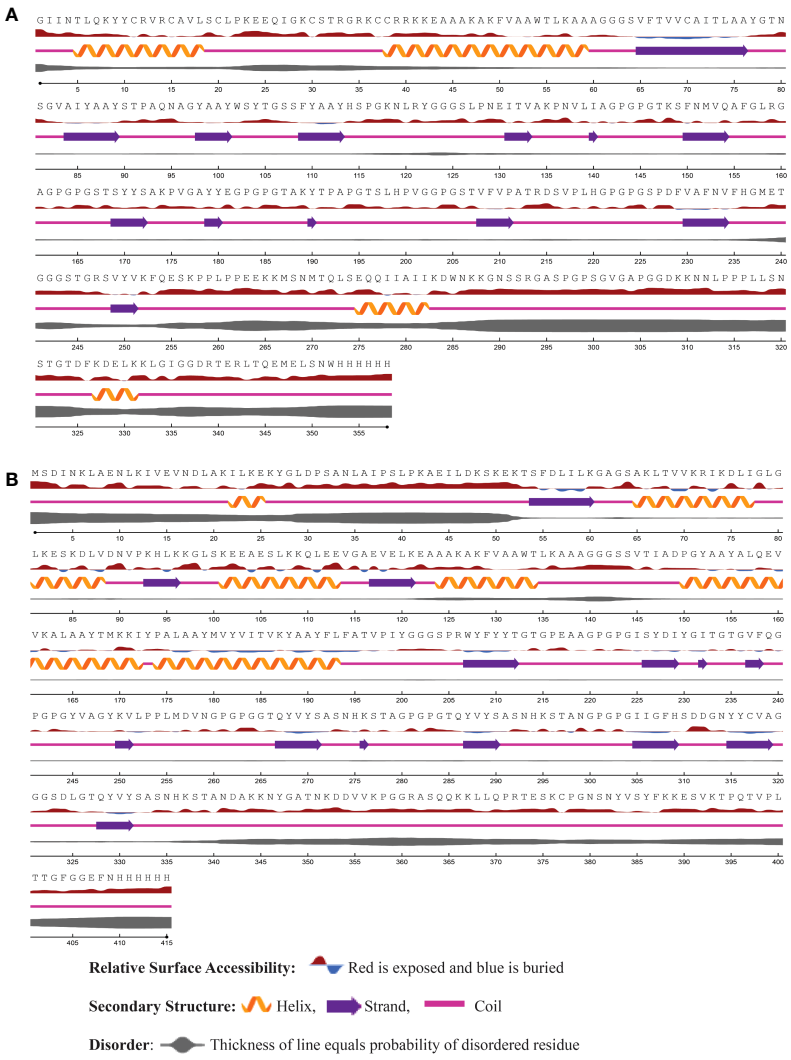


FIGURE 3
The predicted secondary structure features, relative surface accessibility, and disorder regions of **(A)** Neo-1 **(B)** Neo-4 using NetsrurfP-3.0 server.

the Neo-1 complex was 4.5501Å. This system experienced an increase in RMSD descriptors until the 30ns time; after that upward trend was ceased, and the system remained stable. Neo-4 complex experienced a continuous rise in RMSD until 50ns, then underwent smaller fluctuations and remained increased again toward the end of simulation. The average RMSD for this complex was 8.3117Å. Overall, there were no significant variations in both docked complexes during 110ns simulation, indicating the significance of our simulation analysis.

In order to understand changes in volume, SASA for both complexes was analyzed based on simulation trajectories. The average SASA value of Neo-1 and TLR-3 complex was 46952.58 Å, which is similar to the whole SASA profile after the 30ns. In contrast, the average SASA value for Neo-4 and TLR3 complex was 52342.26 Å. The SASA values in both simulated complexes increased initially, indicating that the volume of protein was expanded during the early phase.

The radius of gyration (Rg) of simulation trajectory gives information on the protein's compact nature, with a greater Rg profile indicating less rigidity in the biological system. The average Rg value of the Neo-1 and Neo-4 complex was 32.9086Å and 40.4795 Å, respectively. There was an initial rise in the Rg profile of Neo-1 and TLR-3 complex. Besides, this complex Rg descriptor sustained between 32.1696 to 33.8119 Å between 0 to 110 ns, with several oscillations. This might be the cause of the system's loose packing. The Rg value of the Neo-4 and TLR-3 complex also increased till 35ns and remained similar till the end, though there were few fluctuations.

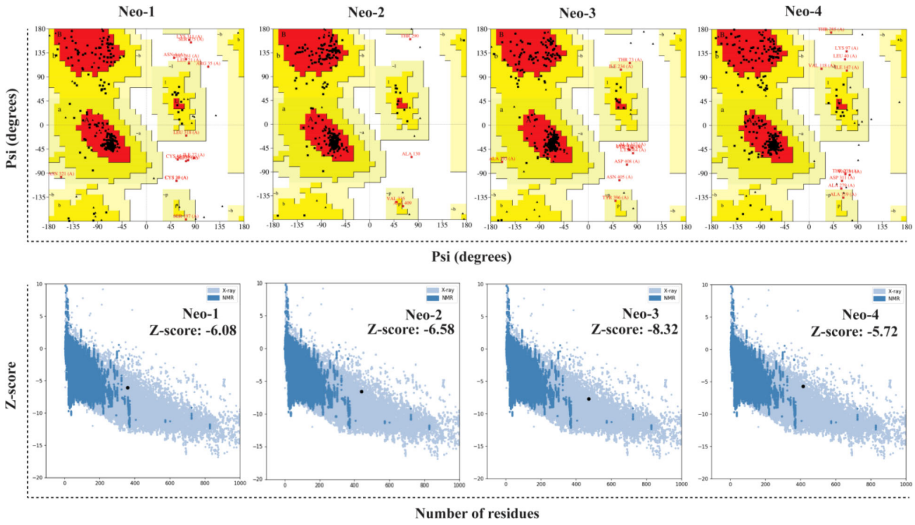


FIGURE 4
Validation of the predicted 3D structure models of NeoCoV vaccines with PROCHECK and ProSA-webserver.

Additionally, the root means square fluctuation (RMSF) profile was used to analyze protein flexibility across amino acid residues. The average RMSF value for Neo-1 and Neo-4 complex was 1.9823 Å and 4.1212 Å. Fluctuations of most residues in Neo-1 and TLR-3 complex were below 2.5 Å, with few residues, such as 690-720, 840-900, and 970-990, indicating larger flexibility. RMSF profile of Neo-4 and TLR3-complex revealed that most residues in this complex had RMSF profiles

TABLE 7 Molecular docking results of Neo-1-4 with selected TLRs.

Vaccine construct	Receptor	Name of target	Score (kcal/mol)	E-Conf	E-Place
Neo-1	2Z7X	TLR2	-86.52	-7275.00	-19.84
	2A0Z	TLR3	**-101.08**	**-7271.54**	**-25.89**
	3FXI	TLR4	-82.04	-6219.70	-20.79
	1AKJ	MHC-I	-85.43	-7157.54	-20.21
	3L6F	MHC-II	-83.97	-6521.34	-19.54
Neo-2	3FXI	TLR4	-70.64	-10124.58	-32.51
	2A0Z	TLR3	-91.88	-10116.98	-26.57
	2Z7X	TLR2	-78.13	-8714.57	-31.01
	1AKJ	MHC-I	-81.21	-9324.76	-23.56
	3L6F	MHC-II	-78.54	-8731.90	-28.54
Neo-3	3FXI	TLR4	-68.93	-14795.16	-36.42
	2A0Z	TLR3	-81.01	-14808.04	-29.46
	2Z7X	TLR2	-86.32	-13277.29	-32.06
	1AKJ	MHC-I	-76.52	-13789.54	-32.65
	3L6F	MHC-II	-83.90	-14567.65	-34.67
Neo-4	3FXI	TLR4	-81.73	-8482.36	-30.68
	2A0Z	TLR3	**-114.47**	**-8515.66**	**-35.05**
	2Z7X	TLR2	-79.56	-7223.11	-33.44
	1AKJ	MHC-I	-84.76	-8675.12	-32.67
	3L6F	MHC-II	-81.87	-8234.65	-31.89

E-Conf, Energy of Confirmation; E-Place, Energy of Placement.

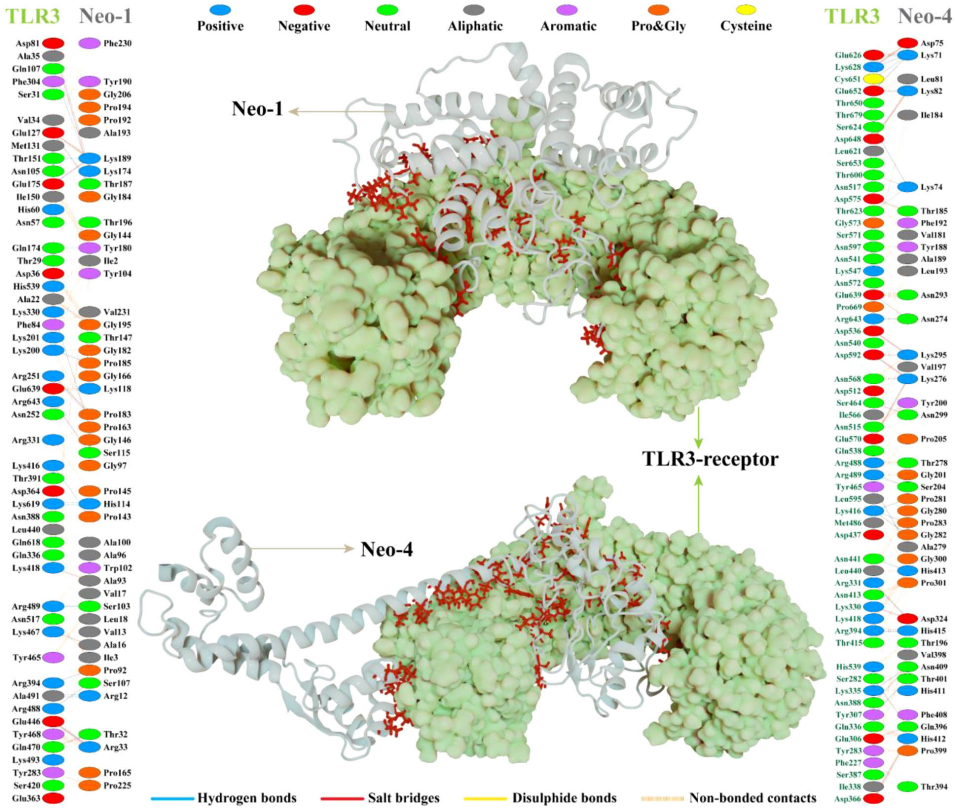


FIGURE 5
NeoCoV vaccine constructs (Neo-1 and Neo-4) docked with human Toll-like receptor-3 (TLR3). The best-docked vaccine-TLR3complex is depicted in the middle. Neo-1 and Neo-4 are represented with a silver cartoon model, and TLR3 is shown in a light green surface model. The interacting residues of Neo-1 and Neo-4 with TLR3 receptor are shown on either side. Hydrogen bonds are indicated in blue lines. The color of interacting residues reflects the properties of amino acid.

below 4.5 Å, with residues (especially in the region 1-250) showed larger fluctuations. These findings define the stability and stiffness of the constructed vaccines' complex.

The transition phases of each system were analyzed by studying free energy landscape (FEL). To investigate the evolution from initial positions to metastable states, first two eigenvectors of the Neo-1- TLR-3 and Neo-4-TLR-3 complexes were employed to generate the trajectories' FEL. To better understand the structural changes in the TLR3 upon multi-epitope vaccine binding, the low energy states in each complex were displayed. FEL plot shows the high energy, intermediate energy, and stable energy states in red, yellow, and blue, respectively. Both complexes revealed more translational conformations during the simulation. Multiple metastable states were observed in both complexes during structural changes, which is shown by high and low energy distribution. RMSD, SASA, Rg, RMSF, and FEL analysis of Neo-1/TLR3 and Neo-4/TLR3 complex is graphically presented in Figures 8, 9.

During simulation, several hydrogen bonds were noted between vaccine and the TLR3 binding interface. For Neo-1 and TLR3 complex, hydrogen bonds between residues $Arg12_{Neo-1}$ and $Ala470_{TLR3}$ and $Pro163_{Neo-1}$ and $Lys309_{TLR3}$ had more than 20% occupancy during the simulation. Also, $Lys189_{Neo-1}$ mediated multiple hydrogen bonds with Glu106 and Glu154 of TLR3 with 20% occupancy, indicating that this residue is crucial for specificity. Besides. Hydrogen bonds between residues $Arg33_{Neo-1}$ and Glu425(2×), $Thr187_{Neo-1}$ and Glu106, $Lys174_{Neo-1}$ and Asp60, and $Gly182_{Neo-1}$ and Lys179 had over 10% occupancy during the simulation. All hydrogen bond occupancy details between the Neo-1 construct and TLR3 complex is provided in Table S13. Similarly, for Neo-4 and TLR3 complex, higher hydrogen bonds occupancy was monitored for $Gly282_{Neo-4}$ and Lys395 (65%) and $Lys295_{Neo-4}$ and Asp515 (63%). In addition, the hydrogen bond between residues $Lys276_{Neo-4}$ and Glu549, $Asn299_{Neo-4}$ and Ser443, $Lys295_{Neo-4}$ and Asp515 had over 45% occupancy throughout the simulation. All hydrogen bond occupancy details between Neo-4 construct and TLR3 receptor is provided in Table S14.

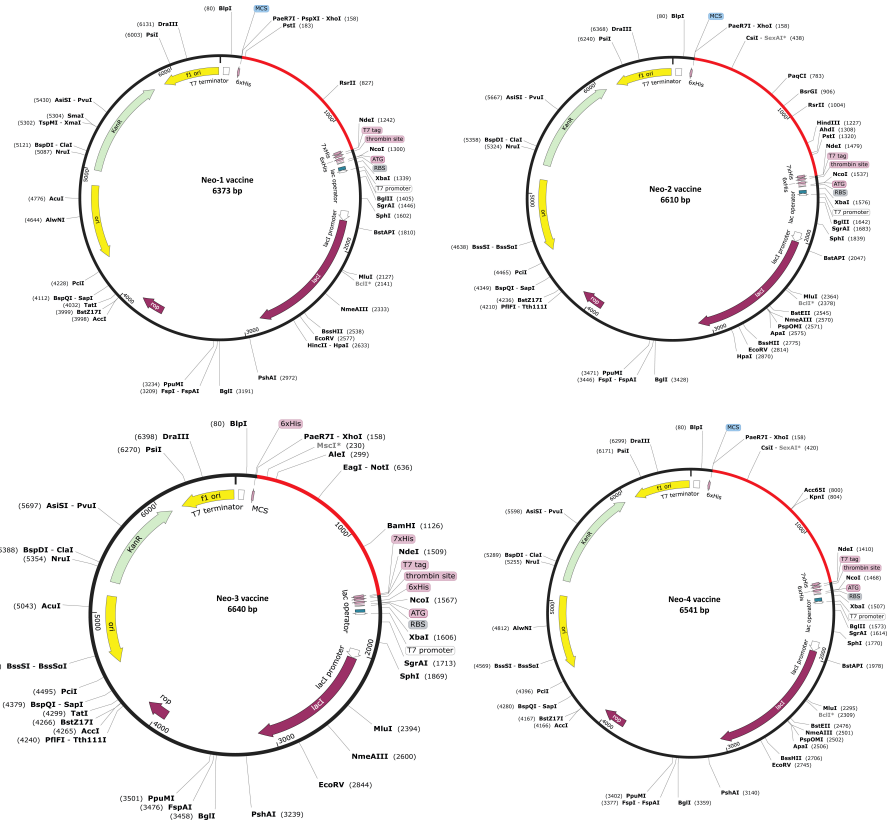


FIGURE 6
In silico restriction cloning of Neo-1–4 vaccine constructs. Using the SnapGene program, the final vaccine constructs codon-optimized sequence (red color) was cloned between the XhoI (158) and NdeI (936) restriction enzyme loci in the pET-28a (+) expression vector (black color). Efficient expression of the designed constructs can be carried out in E. coli strain K12 for effective vaccine production.

3.18 MM/GBSA binding free energy estimation

The binding affinity of the complex, also known as the Gibbs free energy (ΔG) in thermodynamics, is critical for predicting whether an interaction will occur in the cell under specified conditions. Thus, MM/GBSA calculation was conducted to evaluate the binding affinity of vaccine-receptor complex. The total binding energies (ΔTOTAL) of Neo-1–TLR3 and Neo-4–TLR3 complexes are -29.4841 kcal/mol and -36.7910 kcal/mol, respectively (Table 8). Interaction energies of both systems indicated energetically favorable and stable binding, as shown by negative Gibbs free energy values. Both Vander Waals (VDWAALS) and electrostatic energies (EEL) dominate the intermolecular interactions.

4 Discussion

The use of immunoinformatics in the development of promising vaccines against diverse microbes, particularly viruses, are becoming more widely accepted as the first line of vaccine development (80). Epitope-based vaccines have gained much interest due to their potential benefits over traditional immunization (13). Multi-epitope vaccines have distinctive design approaches with the following characteristics, distinguishing them from conventional vaccines and single-epitope vaccines: (1) they are comprised of several MHC-restricted epitopes that T-cell receptors can identify from different T-cell subsets (2) they include CTL, HTL, and B-cell epitopes that may concurrently elicit potent cellular and humoral immune responses (3) they are made up of several epitopes from various viral antigens that can broaden the spectrum of targeted viruses (4) they introduce a few substances that can act as adjuvants, enhancing immunogenicity and producing persistent immune responses (5) they lessen undesirable substances that would otherwise cause adverse effects or abnormal immunological reactions (81). With such benefits, well-designed multi-epitope vaccines should develop into potent preventative and therapeutic treatments for viral infections. Multi-epitope vaccine for SARS-CoV-2 (13, 82) and MERS-CoV (83) have been designed using immunoinformatics-guided methods in recent years. *In silico* predicted multi-epitope vaccine was validated experimentally by *in-vitro* research against *Mycobacterium tuberculosis* (84). Recently,

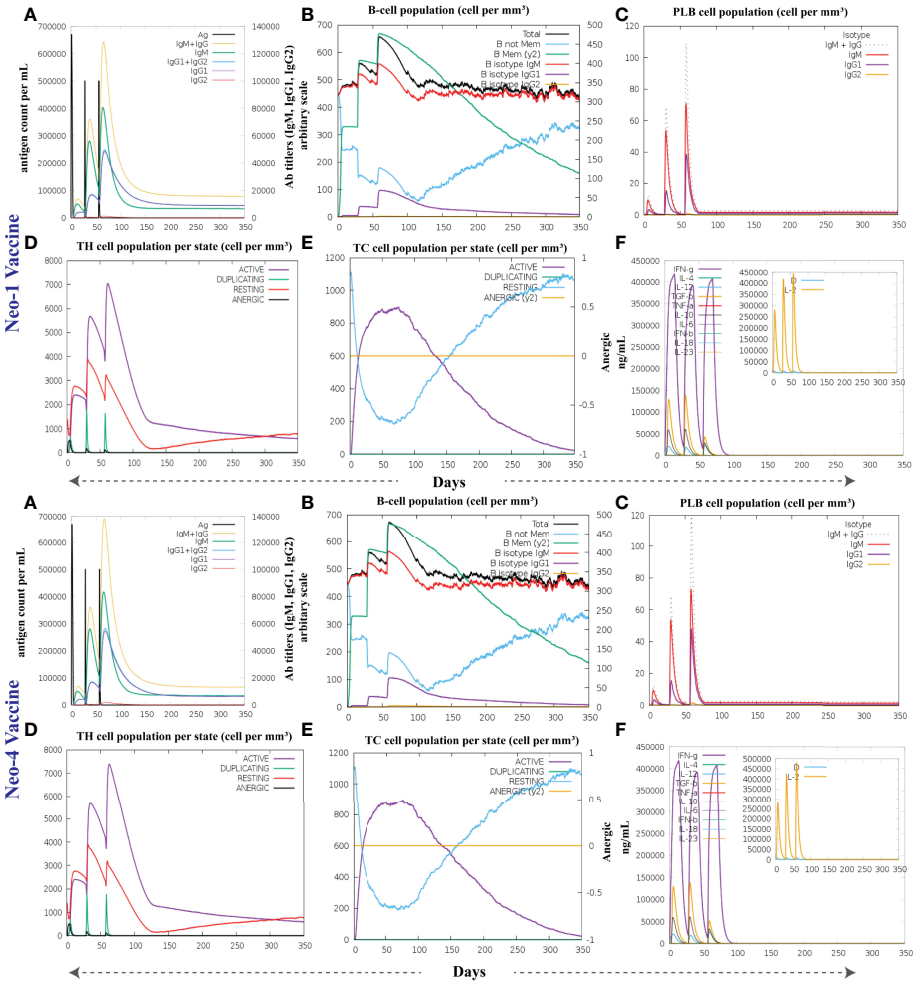


FIGURE 7
The immune response of Neo-1 and Neo-4 evaluated using C-Immsim server **(A)** Response of antibodies and antibodies complex to antigen **(B)** Total count per entity state of B-lymphocytes **(C)** Count of plasma B-lymphocytes **(D)** Total count per entity state of CD4 T helper lymphocyte **(E)** Total count per entity state of CD8 T-cytotoxic lymphocytes **(F)** cytokine concentrations and interleukin in various states shown in a smaller graph with the Simpson index shown in dotted line. In three successive immunological reactions, all units are expressed in cells/mm^3.

Yu et al. employed an immunoinformatics approach to design a multivalent vaccine against SARS-CoV-2 and its variants and confirmed vaccine immunogenicity *in vitro* and *in vivo* (85). In addition, designing a vaccine using a similar method has been shown to develop prophylactic efficiency *in vivo*, and several of these vaccines have reached the clinical trial stage (83). Considering the potential future threat associated with NeoCoV high infectivity and high transmittance rate in human population, prophylactic vaccine could be desirable. Thus, the present study aimed to exploit computational techniques to formulate multi-peptide NeoCoV vaccine candidates capable of eliciting immunological responses in possible infection of NeoCoV in humans.

Herein, NeoCoV protein components were collected to formulate a multi-epitope-based vaccine. Before antigenicity assessment of the sequences, their physical and chemical characteristics were predicted. Using the VaxiJen v2.0 virus model, the viral components were categorized as antigens and non-antigens using a cut-off of 0.4. Non-antigens were categorized as scores below this cut-off and were discarded. For appropriate component selection, the localization and transmembrane helices of the protein sequences were also assessed. This procedure determines if a prospective vaccine is suitable for experimental validation during the vaccine formulation process (86, 87). The results of all structural proteins and ORF4a, ORF4b, ORF8b, ORF5b, viral protease, NSP10, ADP Binding module, and Coronavirus endopeptidase C30 from NSPs were analyzed for epitope-based vaccine design.

In addition, these viral antigenic components were used to predict B and T-cell epitopes. A promising vaccine should be able to produce antigen-induced immunity, which is long-

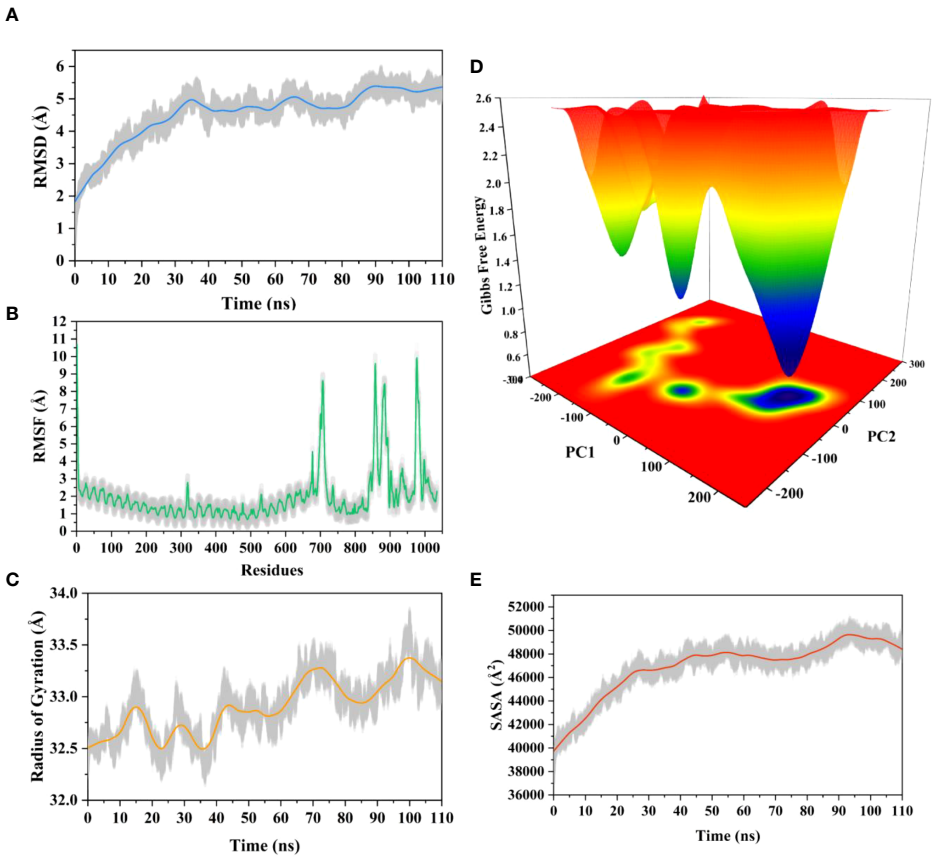


FIGURE 8
Molecular dynamics simulation of Neo-1 and TLR-3 complex at 110ns **(A)** Root means square deviation (RMSD) plot of the complex, representing mild fluctuations **(B)** The Root means square fluctuation (RMSF) plot of the docked complex **(C)** The Radius of Gyration plot of the docked complex **(D)** Free energy landscapes (FELs) of the docked complex. High, intermediate, and low/stable energy states are shown in red, yellow/green, and light-to-dark blue color in the graph, respectively **(E)** Alterations in Solvent-Accessible Surface Area (SASA) profile of the docked complex during the simulation.

lasting and adaptive. Both MHC classes were expected to have a total of 44 epitopes. In contrast, the B-cells were predicted to have 19 epitopes that were immunogenic, non-toxic, and safe from causing allergic reactions. The CTL epitopes help build long-lasting immunity against viruses and infected cells (88). At the same time, HTL epitopes are linked with both humoral immunity and cell-mediated immunity. These epitopes trigger a helper T-cell response (CD4+ T cell), leading to the formation of protective memory CD8+ T cells and B-cell activation (89).

In our study, multi-epitope vaccines were constructed by fusing prioritized epitopes with appropriate adjuvants and linkers. Although safer, subunit vaccinations are often less immunogenic and effective, necessitating the use of an adjuvant. In order to enhance and direct the adaptive immune response to vaccine antigens, adjuvants are thus essential (90). Different adjuvants— β-defensin, 50S ribosomal protein L7/L12, and HBHA—were used to construct NeoCoV vaccines. By activating innate immunity cells and attracting naïve T cells through the chemokine receptor CCR-6, β-defensin peptides stimulate innate immunity cells (91). The maturation of dendritic cells, HTLs, CTLs, and IFN-γ-producing cells is induced by the 50S ribosomal protein L7/L12 when naive T-cells are stimulated (61). HBHA is a novel TLR4 agonist with minimal systemic toxicity, in addition to possessing a strong immunostimulatory potential and the potential to induce dendritic cells maturation in a TLR4-dependent manner (92). Linkers are necessary for extended conformation (flexibility), protein folding, and functional domain separation, all of which help produce a more stable protein structure (93). EAAAK linker (added after the adjuvant) was utilized to provide stiffness, minimizing any interference from other protein regions in the interaction between the adjuvant and its receptor. Following this linker, we added PADRE sequence to boost the vaccine immunogenicity. Alternatively, GGGS linker (added to separate epitope class) provides flexibility to the structure. (94). GPGPG linker (used for linking HTL epitopes) prevents junctional immunogenicity and can induce T helper cell

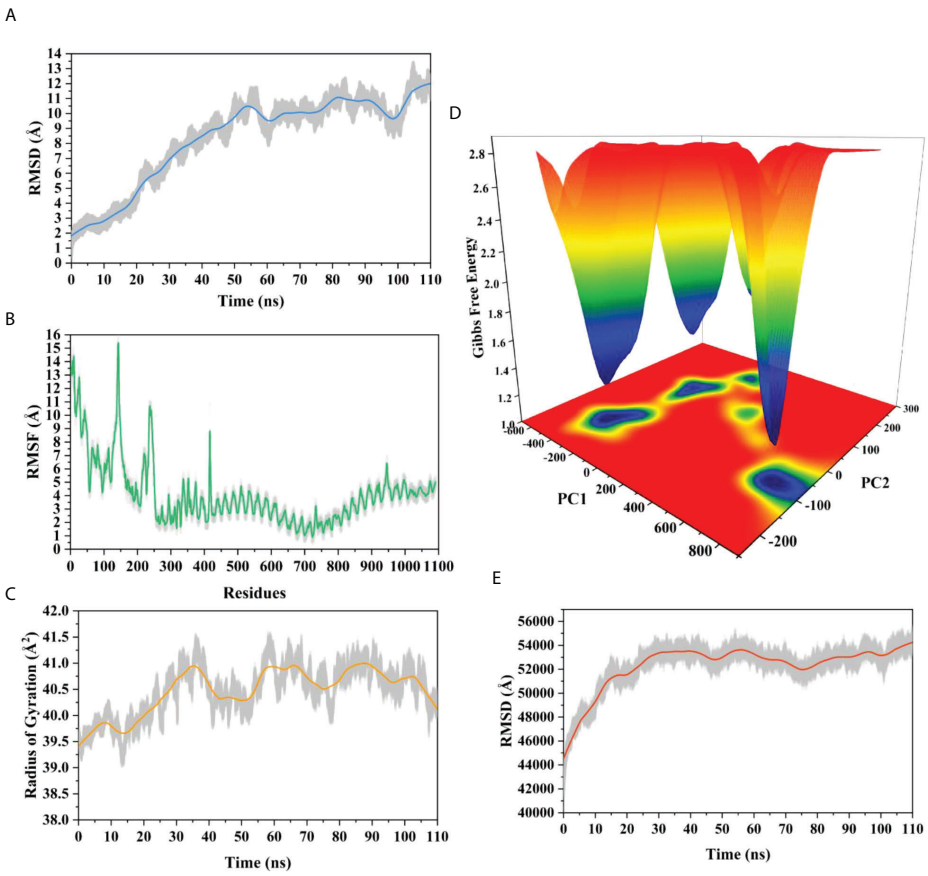


FIGURE 9
Molecular dynamics simulation of Neo-4 and TLR-3 complex at 110ns **(A)** Root means square deviation (RMSD) plot of the complex, representing mild fluctuations **(B)** The Root means square fluctuation (RMSF) plot of the docked complex **(C)** The Radius of Gyration plot of the docked complex **(D)** Free energy landscapes (FELs) of the docked complex. High, intermediate, and low/stable energy states are shown in red, yellow/green, and light-to dark blue color in the graph, respectively **(E)** Alterations in Solvent-Accessible Surface Area (SASA) profile of the docked complex during the simulation.

immunological response. AAY liner (used for connecting CTL epitopes) promotes epitope presentation by assisting epitopes in forming suitable sites for binding to the TAP transporter (13). In mammals, the AAY linker is also the proteasomal cleavage site. B cell epitopes were connected by a KK linker. During the processing and presentation of epitopes through MHC-II molecules for antibody induction, the KK linker sequence is a target for the lysosomal protease enzyme (94).

ERRAT, PROCHECK, and ProSA-web servers were used to verify the predicted 3D structures of multi-epitope vaccines

TABLE 8 MM/GBSA free energy estimation and individual free energy components of vaccines-TLR3 complexes.

Energy Component	Neo-1		Neo-4	
	Average	Std. Err. Mean	Average	Std. Err. Mean
VDWAALS	-134.5904	0.2434	-169.2386	0.3853
EEL	-719.0948	1.9689	-1076.0320	1.5889
EGB	840.8048	1.9405	1230.8375	1.6378
ESURF	-16.6036	0.0338	-22.3579	0.0517
ΔG gas	-853.6853	1.9922	-1245.2706	1.6934
ΔG solvation	824.2012	1.9315	1208.4796	1.6126
ΔTOTAL	-29.4841	0.1862	-36.7910	0.2565

which showed that the overall model quality reflects X-ray crystallographic properties. Overall, the validation results suggest that the predicted vaccine structures are of high quality. Consequently, the predicted 3D structures of vaccines are reliable and can be used in downstream evaluation.

Molecular docking is vital *in silico* method to investigate pattern of interaction and affinities of ligand-receptor binding using a lock-and-key method. Protein-protein docking is often used in immunoinformatics to analyze the best, stable, and effective vaccine by examining the binding modes of proteins, interacting atoms, and binding energies. Following the same method, optimal vaccine construct(s) were predicted against immune molecules (TLR2, TLR3, and TLR4). These receptors are important in the formation of an efficient immune response, according to several immunoinformatics studies (13, 44, 95–98).TLRs are involved in both innate and adaptive immunity, as well as immune activation. The role of TLR2 and TLR4 has also been investigated in the activation of structural proteins of virus that leads to the generation of inflammatory cytokines (99). Besides, TLR3 detects viral infection and triggers an innate immunity signaling pathway (96).

The proposed constructs (Neo-1–4) were docked well with receptors (TLRs, MHC class-I and II), revealing that the designed vaccines fit well into the receptors binding pockets; hence the vaccine constructs can elicit persistent immune responses. Due to the higher docking energies and excellent performance, Neo-1 and Neo-4 were further subjected to MD simulation.

MD simulation is a robust technique to capture the motion of atoms at the atomic level, which is very difficult using experimental methods (100). The conformational stability and compactness of the vaccine-TLR3 complex were confirmed using the RMSD, RMSF, and Rg descriptors. The SASA and FEL profile of complexes revealed structural alterations induced upon the binding of the vaccine construct (ligand) with the receptor. Moreover, hydrogen bond analysis showed that hydrogen bonds were stable during the simulation and would possibly play an essential role in complex stability (101). Estimating binding free energies with MM/GBSA indicated negative ΔG scores for vaccine-TLR3 complexes; hence the values were consistent with docking scores (102). To reduce codon bias (103), computational cloning was performed on a pET28a (+) vector following codon optimization with the JCAT web service. The optimized nucleotides' CAIs and GC content were within the acceptable limits of 0.8–1.0 and 30–70%, respectively. These results demonstrated that the proposed constructs are durable and can be expressed efficiently in *E. coli* (strain K12).

An immune response simulator, the C-ImmSim server, was deployed to assess the potency of the predicted vaccine to generate an immunological response (60). This method

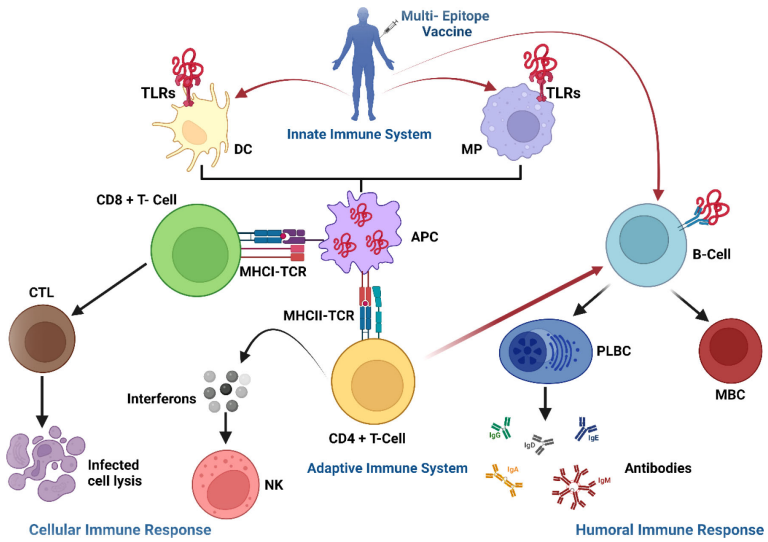


FIGURE 10
Proposed immune route in response to a multiepitope-based subunit vaccine in the host. The vaccine enters the cell through toll-like receptors (TLRs) and attaches to dendritic cells (DC) and macrophages (MP), triggering an innate immune response. The vaccine's epitopes are then digested by antigen-presenting cells (APC) and presented to T-cells, which triggered an adaptive immunological response. T-cells activate other immune cells or destroy infected cells directly (cellular immune response), whilst plasma B-cells (PLBC) create antibodies to neutralize viruses and memory B-cells (MBC) preserve all the information needed to mount a powerful immune response in the event of re-infection (humoral immune response). TCR, T-cell receptors; CTL, cytotoxic T lymphocytes; IFNs, interferons; NK, natural killer cells; CD, cluster of differentiation; MHC, major histocompatibility complex; TCR, T-cell receptors; CTL, cytotoxic T lymphocytes; IFNs, interferons; NK, natural killer cells; PLBC, Plasma B-cells.

models the essential components of a functional mammalian system (bone marrow, thymus, and lymph node). The immune cells' response to the developed vaccines, including CTL, HTL, antibodies, B-cell, dendritic cells, macrophages, and cytokines, was observed since an effective vaccine must form a lifelong adaptive immunity, hence mimic the antigen induced natural immunity. Primary antibodies (IgG and IgM), T-cell, B-cell and cytokines were triggered by immune simulation by Neo-1 and Neo-4. The proposed immune route in response to a multiepitope-based subunit vaccine (designed in this study) in the host is provided in the Figure 10. The developed vaccines, Neo-1 and Neo-4 may protect against NeoCoV, like previous multi-epitope vaccine design studies exploiting immunoinformatics tools (104–106). Further validation of safety and efficacy profile of the designed vaccines against NeoCoV is warranted at experimental level.

As for the NeoCoV-related viruses, various challenges for antiviral vaccines are documented. For instance, the high-affinity binding mode between MERS-CoV RBD and human DDP4 receptor required a high concentration of targeted antibodies (ribavirin) higher than what can be clinically acceptable in humans (107–109). The same is the case with SARS-CoV-2 spike protein and ACE2 receptor binding underscores the high-affinity binding requirements of the neutralizing antibodies induced by the COVID-19 vaccine (110). Another issue is the prerequisite of suitable animal models, which has hampered the vaccine safety and efficacy testing of MERS-CoV and SARS-CoV-2 disease (110, 111). Immunopathological concerns associated with subsets T cells (Th2) were raised by an earlier report regarding the MERS-CoV (110). Besides safety issues associated with the live attenuated vaccine, hypertensive lung pathology was reported for transgenic mice vaccinated with inactivated MERS-CoV vaccine (112, 113). In clinical trials, reactogenicity after vaccination and suppression of innate immune system activation is also reported for anti-SARS-CoV-2 vaccines (110). It is, however, yet to be determined whether anti-NeoCoV vaccines would demonstrate similar risks in animal models or clinical trials. On the other hand, novel vaccine development may encounter some typical challenges, including complications in the development of the production process, formulation, analytical assays, and impediments in the optimization of analytical assays. Moreover, vaccine construction can face various difficulties in low resource settings, such as technicalities, implementation of clinical trials, funding, introduction, and commercialization (114). Previously, Yan et al. demostrated that neutralizing antibodies that are elicited by the present COVID-19 vaccines against SARS-CoV-2 and MERS-CoV could not cross-neutralize the NeoCoV infection (2). Our study reports the first attempt of using comprehensive computational approach to design a multi-epitope vaccine for NeoCoV. We expect that our work could provide a starting point for the wet-lab therapeutic studies against NeoCoV infection.

Limitation

This work presents an alternate vaccination method based on the multi-epitope assembly of NeoCoV genome protein components to address antigenic complexity. Given the limitations of our method, we were very stringent and only chose top candidate epitopes confirmed by multiple tools. Although immunoinformatics techniques were used to propose the NeoCoV vaccines, and they were predicted immunogenic, it is uncertain how much protection against NeoCoV infection they provide. This is something that future experiments will have to evaluate upon the validation of epitopes. Immunoinformatics methods are beneficial for *in silico* research and may direct laboratory investigations, saving time and money. The next step is to conduct *in vitro* immunological experiments to verify the proposed vaccines, ascertain their immunogenicity, antigenicity, safety, efficacy and develop challenge-protection preclinical trials to validate these methods.

5 Conclusion

Multi-epitope vaccines have already gained popularity and showed protective efficacy *in vivo*, with several undergoing clinical studies. The current study was based on an immunoinformatics-driven strategy to identify possible antigenic epitopes for inclusion in a NeoCoV vaccine candidate. Four multi-epitope vaccines were developed using three antigen categories of NeoCoV proteins including CTL, HTL, and B-cell-epitopes. Computational analysis of physicochemical and antigenic characteristics of vaccines was performed. Molecular docking and molecular dynamics simulations were used to investigate the stability profile and molecular interactions between the proposed vaccines and immunological receptors. The computer simulation showed that vaccine constructs can generate an immunological response. Several immunoinformatics methodologies were successively used to create and assess a vaccine that may provide protective immunity against viral infection; nevertheless, experimental testing is necessary to determine the precise effectiveness. The vaccine may be synthesized before being tested *in vitro* and *in vivo* in the experimental assay. We also recommend more research into the synthesis and biological activities of the multi-epitope vaccines that has been developed.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SH, AK and AA-H conceived and designed the study. SA, MW and SH performed experiments. AA, AI and MI analyzed the data. SA, MW, SH and AK wrote the manuscript with inputs and comments from all co-authors. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.956776/full#supplementary-material

References

1. Ur Rehman MF, Fariha C, Anwar A, Shahzad N, Ahmad M, Mukhtar S, et al. Novel coronavirus disease (COVID-19) pandemic: A recent mini review. *Comput Struct Biotechnol J* (2021) 19:612–23. doi: 10.1016/j.csbj.2020.12.033

2. Yan H, Wang X, Xiong Q, Cao L, Ma C, Liu C, et al. Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. *bioRxiv[preprint]* (2022). doi: 10.21203/rs.3.rs-1292418/v1

3. Hok L, Rimac H, Mavri J, Vianello R. COVID-19 infection and neurodegeneration: Computational evidence for interactions between the SARS-CoV-2 spike protein and monoamine oxidase enzymes. *Comput Struct Biotechnol J* (2022) 20:1254–63. doi: 10.1016/j.csbj.2022.02.020

4. Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir. *Virol J* (2016) 13:1–7. doi: 10.1186/s12985-016-0544-0

5. Hassan MM, Hussain MA, Kambal S, Elshikh AA, Gendeel OR, Ahmed SA, et al. NeoCoV is closer to MERS-CoV than SARS-CoV. *Infect Diseases: Res Treat* (2020) 13:1178633720930711. doi: 10.1177/1178633720930711

6. Dhawan M, Thakur N, Priyanka MS, Choudhary OP. NeoCoV: A foresight of the next possible pandemic. *Int J Surg (London England)* (2022) 99:106255. doi: 10.1016/j.ijsu.2022.106255

7. Organization, W.H. *Weekly epidemiological update on COVID-19 - 22 march 2022* (2022). Available at: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—22-march-2022.

8. Raj VS, Mou H, Smits SL, Dekkers DH, Müller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* (2013) 495:251–4. doi: 10.1038/nature12005

9. Luo C-M, Wang N, Yang X-L, Liu H-Z, Zhang W, Li B, et al. Discovery of novel bat coronaviruses in south China that use the same receptor as middle East respiratory syndrome coronavirus. *J Virol* (2018) 92:e00116–00118. doi: 10.1128/JVI.00116-18

10. Trigueiro-Louro J, Correia V, Figueiredo-Nunes I, Gíria M, Rebelo-De-Andrade H. Unlocking COVID therapeutic targets: A structure-based rationale against SARS-CoV-2, SARS-CoV and MERS-CoV spike. *Comput Struct Biotechnol J* (2020) 18:2117–31. doi: 10.1016/j.csbj.2020.07.017

11. Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, et al. Close relative of human middle East respiratory syndrome coronavirus in bat, south Africa. *Emerg Infect Dis* (2013) 19:1697–9. doi: 10.3201/eid1910.130946B.

12. Choudhary OP, Ali RK, Maulud SQ, Dhawan M, Mohammed TA. Will the next spillover pandemic be deadlier than the COVID-19?: a wake-up call. *Int J Surg (London England)* (2022) 97:106208–8. doi: 10.1016/j.ijsu.2021.106208

13. Dong R, Chu Z, Yu F, Zha Y. Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front Immunol* (2020) 11:1784. doi: 10.3389/fimmu.2020.01784

14. Nabel GJ. HIV Vaccine strategies. *Vaccine* (2002) 20:1945–7. doi: 10.1016/S0264-410X(02)00074-9

15. Sweredoski MJ, Baldi P. COBEpro: a novel system for predicting continuous b-cell epitopes. *Protein Engineering Design Selection* (2009) 22:113–20. doi: 10.1093/protein/gzn075

16. Duffy L, O'reilly SC. Toll-like receptors in the pathogenesis of autoimmune diseases: recent and emerging translational developments. *Immuno Targets Ther* (2016) 5:69. doi: 10.2147/ITT.S89795

17. Liu W, Lu Y, Chen Y. Bioinformatics analysis of SARS-cov m protein provides information for vaccine development. *Prog Natural Sci* (2003) 13:844–7. doi: 10.1080/10020070312331344530

18. Sedeyn K, Schepens B, Saelens X. Respiratory syncytial virus nonstructural proteins 1 and 2: Exceptional disrupters of innate immune responses. *PloS Pathog* (2019) 15:e1007984. doi: 10.1371/journal.ppat.1007984

19. Sagara I, Ellis RD, Dicko A, Niambele MB, Kamate B, Guindo O, et al. A randomized and controlled phase 1 study of the safety and immunogenicity of the AMA1-C1/Alhydrogel®+ CPG 7909 vaccine for plasmodium falciparum malaria in semi-immune malian adults. *Vaccine* (2009) 27:7292–8. doi: 10.1016/j.vaccine.2009.10.087

20. Kumar Pandey R, Ojha R, Mishra A, Kumar Prajapati V. Designing b-and T-cell multi-epitope based subunit vaccine using immunoinformatics approach to control zika virus infection. *J Cell Biochem* (2018) 119:7631–42. doi: 10.1002/jcb.27110

21. Rost B, Sander C, Casadio R, Fariselli P. Transmembrane helices predicted at 95% accuracy. *Protein Sci* (1995) 4:521–33. doi: 10.1002/pro.5560040318

22. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* (2017) 33:3387–95. doi: 10.1093/bioinformatics/btx431

23. Westerhout J, Krone T, Snippe A, Babé L, Mcclain S, Ladics GS, et al. Allergenicity prediction of novel and modified proteins: Not a mission impossible! development of a random forest allergenicity prediction model. *Regul Toxicol Pharmacol* (2019) 107:104422. doi: 10.1016/j.yrtph.2019.104422

24. Dimitrov I, Flower DR, Doytchinova I. "AllerTOP-a server for in silico prediction of allergens". *BMC bioinformatics: BioMed central.* (2013) 14:S4. doi: 10.1186/1471-2105-14-S6-S4.

25. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* (2014) 30:846–51. doi: 10.1093/bioinformatics/btt619

26. Fadaka AO, Pretorius A, Klein A. MicroRNA assisted gene regulation in colorectal cancer. *Int J Mol Sci* (2019) 20:4899. doi: 10.3390/ijms20194899

27. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* (2007) 8:1–7. doi: 10.1186/1471-2105-8-4

28. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. "Peptide toxicity prediction,". In: *Computational peptidology*. Springer: Humana Press, New York, NY (2015). p. 143–57.

29. El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear b-cell epitopes using string kernels. *J Mol Recognition: Interdiscip J* (2008) 21:243–55. doi: 10.1002/jmr.893

30. Osatomi K, Sumiyoshi H. Complete nucleotide sequence of dengue type 3 virus genome RNA. *Virology* (1990) 176:643–7. doi: 10.1016/0042-6822(90)90037-R

31. Faria AR, Costa MM, Giusta MS, Grimaldi GJr., Penido ML, Gazzinelli RT, et al. High-throughput analysis of synthetic peptides for the immunodiagnosis of canine visceral leishmaniasis. *PloS Negl Trop Dis* (2011) 5:e1310. doi: 10.1371/journal.pntd.0001310

32. Saha S, Raghava GP. "Prediction methods for b-cell epitopes,". In: *Immunoinformatics*. Springer: Humana Press, New York, NY . (2007). p. 387–94.

33. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* (2017) 45:W24–9. doi: 10.1093/nar/gkx346

34. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf* (2008) 9:1–8. doi: 10.1186/1471-2105-9-514

35. 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) 8:1–12. doi: 10.1186/1471-2105-8-424

36. 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. doi: 10.1371/journal.pcbi.1003266

37. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol* (2013) 191:5831–9. doi: 10.4049/jimmunol.1302101

38. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* (2004) 22:3195–204. doi: 10.1016/j.vaccine.2004.02.005

39. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* (2018) 154:394–406. doi: 10.1111/imm.12889

40. Kruiswijk C, Richard G, Salverda ML, Hindocha P, Martin WD, De Groot AS, et al. In silico identification and modification of T cell epitopes in pertussis antigens associated with tolerance. *Hum Vaccines Immunother* (2020) 16:277–85. doi: 10.1080/21645515.2019.1703453

41. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: a server for quantitative prediction of peptide–MHC binding. *Nucleic Acids Res* (2003) 31:3621–4. doi: 10.1093/nar/gkg510

42. Adhikari UK, Tayebi M, Rahman MM. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *J Immunol Res* (2018) 2018:6718083. doi: 10.1155/2018/6718083

43. dBui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinf* (2006) 7:1–5. doi: 10.1186/1471-2105-7-153

44. Fadaka AO, Sibuyi NRS, Martin DR, Goboza M, Klein A, Madiehe AM, et al. Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. *Sci Rep* (2021) 11:1–22. doi: 10.1038/s41598-021-99227-7

45. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *Proteomics Protoc Handb* (2005), 571–607. doi: 10.1385/1-59259-890-0:571

46. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* (2009) 25:2200–7. doi: 10.1093/bioinformatics/btp386

47. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein–sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* (2017) 33:3098–100. doi: 10.1093/bioinformatics/btx345

48. Abedi Karjiban R, Abdul Rahman MB, Basri M, Salleh AB, Jacobs D, Abdul Wahab H. Molecular dynamics study of the structure, flexibility and dynamics of thermostable L1 lipase at high temperatures. *Protein J* (2009) 28:14–23. doi: 10.1007/s10930-008-9159-7

49. Fadaka AO, Sibuyi NRS, Madiehe AM, Meyer M. Computational insight of dexamethasone against potential targets of SARS-CoV-2. *J Biomol Structure Dynamics* (2022) 40:875–85. doi: 10.1080/07391102.2020.1819880

50. Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y. Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS coronavirus-2 (SARS-CoV-2). *Immunobiology* (2020) 225:151955. doi: 10.1016/j.imbio.2020.151955

51. Baek M, Dimaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (2021) 373:871–6. doi: 10.1126/science.abj8754

52. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* (2013) 41:W384–8. doi: 10.1093/nar/gkt458

53. Tripathi NK, Shrivastava A. Recent developments in recombinant protein–based dengue vaccines. *Front Immunol* (2018) 9:1919 doi: 10.3389/fimmu.2018.01919

54. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* (1993) 2:1511–9. doi: 10.1002/pro.5560020916

55. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* (2007) 35:W407–10. doi: 10.1093/nar/gkm290

56. Ulc, C.C.G. *"Molecular operating environment. 2020:1010 sherbrooke st. West, suite #910, Montreal, QC, Canada, H3A 2R7."*. focal press Waltham, Massachusetts (2020).

57. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci* (2018) 27:129–34. doi: 10.1002/pro.3289

58. Hess R. *Blender foundations: The essential guide to learning blender 2.5*. Routledge: focal press Waltham, Massachusetts (2013).

59. 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. doi: 10.1093/nar/gki376

60. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PloS One* (2010) 5:e9862. doi: 10.1371/journal.pone.0009862

61. Sami SA, Marma KKS, Mahmud S, Khan MAN, Albogami S, El-Shehawi AM, et al. Designing of a multi-epitope vaccine against the structural proteins of marburg virus exploiting the immunoinformatics approach. *ACS Omega* (2021) 6:32043–71. doi: 10.1021/acsomega.1c04817

62. Case DA, Belfon K, Ben-Shalom I, Brozell SR, Cerutti D, Cheatham T, et al. (2020). Amber 2020.

63. Tian C, Kasavajhala K, Belfon KA, Raguette L, Huang H, Migues AN, et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* (2019) 16:528–52. doi: 10.1021/acs.jctc.9b00591

64. Sengupta A, Li Z, Song LF, Li P, Merz JKM. Parameterization of monovalent ions for the Opc3, opc, Tip3p-fb, and Tip4p-fb water models. *J Chem Inf Model* (2021) 61:869–80. doi: 10.1021/acs.jcim.0c01390

65. Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* (2001) 22:501–8. doi: 10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V

66. Darden T, York D, Pedersen L. Particle mesh ewald: An n· log (N) method for ewald sums in large systems. *J Chem Phys* (1993) 98:10089–92. doi: 10.1063/1.464397

67. Roe DR, Cheatham Iii TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* (2013) 9:3084–95. doi: 10.1021/ct400341p

68. Ziegel E. *"Numerical recipes: The art of scientific computing"*. Taylor & Francis: Taylor & Francis Group England (1987).

69. Yang T, Wu JC, Yan C, Wang Y, Luo R, Gonzales MB, et al. Virtual screening using molecular simulations. *Proteins: Structure Function Bioinf* (2011) 79:1940–51. doi: 10.1002/prot.23018

70. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* (2011) 51:69–82. doi: 10.1021/ci100275a

71. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure Function Bioinf* (2004) 55:383–94. doi: 10.1002/prot.20033

72. Ur Rahman M, Liu H, Wadood A, Chen H-F. Allosteric mechanism of cyclopropylindolobenzazepine inhibitors for HCV NS5B RdRp *via* dynamic correlation network analysis. *Mol Biosyst* (2016) 12:3280–93. doi: 10.1039/C6MB00521G

73. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T-and b-cell epitope prediction. *J Immunol Res* (2017) 2017:2680160. doi: 10.1155/2017/2680160

74. Baseer S, Ahmad S, Ranaghan KE, Azam SS. Towards a peptide-based vaccine against shigella sonnei: A subtractive reverse vaccinology based approach. *Biologicals* (2017) 50:87–99. doi: 10.1016/j.biologicals.2017.08.004

75. Yang MM, Wang J, Dong L, Kong DJ, Teng Y, Liu P, et al. Lack of association of C3 gene with uveitis: additional insights into the genetic profile of uveitis regarding complement pathway genes. *Sci Rep* (2017) 7:1–8. doi: 10.1038/s41598-017-00833-1

76. Ueda H, Howson JM, Esposito L, Heward J, Chamberlain G, Rainbow DB, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* (2003) 423:506–11. doi: 10.1038/nature01621

77. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* (1993) 26:283–91. doi: 10.1107/S0021889892009944

78. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol* (2010) 125:S3–S23. doi: 10.1016/j.jaci.2009.12.980

79. Fadaka AO, Aruleba RT, Sibuyi NRS, Klein A, Madiehe AM, Meyer M. Inhibitory potential of repurposed drugs against the SARS-CoV-2 main protease: a computational-aided approach. *J Biomol Structure Dynamics* (2022) 40:3416–27. doi: 10.1080/07391102.2020.1847197.

80. Dar HA, Zaheer T, Shehroz M, Ullah N, Naz K, Muhammad SA, et al. Immunoinformatics-aided design and evaluation of a potential multi-epitope vaccine against klebsiella pneumoniae. *Vaccines* (2019) 7:88. doi: 10.3390/vaccines7030088

81. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* (2018) 15:182–4. doi: 10.1038/cmi.2017.92

82. Jyotisha, Singh S, Qureshi IA. Multi-epitope vaccine against SARS-CoV-2 applying immunoinformatics and molecular dynamics simulation approaches. *J Biomol Structure Dynamics* (2020) 40:2917–33. doi: 10.1080/07391102.2020.1844060.

83. Mahmud S, Rafi M, Paul GK, Promi MM, Shimu M, Sultana S, et al. Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Sci Rep* (2021) 11:1–20. doi: 10.1038/s41598-021-92176-1

84. Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, et al. In silico predicted mycobacterial epitope elicits *in vitro* T-cell responses. *Mol Immunol* (2014) 61:16–22. doi: 10.1016/j.molimm.2014.04.009

85. Yu M, Zhu Y, Li Y, Chen Z, Li Z, Wang J, et al. Design of a recombinant multivalent epitope vaccine based on SARS-CoV-2 and its variants in immunoinformatics approach. *Front Immunol* (2022) 2116. doi: 10.3389/fimmu.2022.884433

86. Kumar Jaiswal A, Tiwari S, Jamal SB, Barh D, Azevedo V, Soares SC. An in silico identification of common putative vaccine candidates against treponema pallidum: A reverse vaccinology and subtractive genomics based approach. *Int J Mol Sci* (2017) 18:402. doi: 10.3390/ijms18020402

87. Naz K, Naz A, Ashraf ST, Rizwan M, Ahmad J, Baumbach J, et al. PanRV: Pangenome-reverse vaccinology approach for identifications of potential vaccine candidates in microbial pangenome. *BMC Bioinf* (2019) 20:1–10. doi: 10.1186/s12859-019-2713-9

88. Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, et al. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep* (2020) 10:1–24. doi: 10.1038/s41598-020-67749-1

89. Sauer K, Harris T. An effective COVID-19 vaccine needs to engage T cells. *Front Immunol* (2020) 2371. doi: 10.3389/fimmu.2020.581807

90. Apostolico JDS, Lunardelli VAS, Coirada FC, Boscardin SB, Rosa DS. Adjuvants: classification, modus operandi, and licensing. *J Immunol Res* (2016) 2016:1459394. doi: 10.1155/2016/1459394

91. Mohan T, Sharma C, Bhat AA, Rao D. Modulation of HIV peptide antigen specific cellular immune response by synthetic α-and β-defensin peptides. *Vaccine* (2013) 31:1707–16. doi: 10.1016/j.vaccine.2013.01.041

92. Lei Y, Shao J, Ma F, Lei C, Chang H, Zhang Y. Enhanced efficacy of a multi-epitope vaccine for type a and O foot–and-mouth disease virus by fusing multiple epitopes with mycobacterium tuberculosis heparin-binding hemagglutinin (HBHA), a novel TLR4 agonist. *Mol Immunol* (2020) 121:118–26. doi: 10.1016/j.molimm.2020.02.018

93. Nezafat N, Ghasemi Y, Javadi G, Khoshnoud MJ, Omidinia E. A novel multi-epitope peptide vaccine against cancer: an in silico approach. *J Theor Biol* (2014) 349:121–34. doi: 10.1016/j.jtbi.2014.01.018

94. 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) 19:1–14. doi: 10.1186/s43141-021-00160-z

95. Kerepesi LA, Leon O, Lustigman S, Abraham D. Protective immunity to the larval stages of onchocerca volvulus is dependent on toll-like receptor 4. *Infect Immun* (2005) 73:8291–7. doi: 10.1128/IAI.73.12.8291-8297.2005

96. Vercammen E, Staal J, Beyaert R. Sensing of viral infection and activation of innate immunity by toll-like receptor 3. *Clin Microbiol Rev* (2008) 21:13–25. doi: 10.1128/CMR.00022-07

97. Vijay K. Toll-like receptors in immunity and inflammatory diseases: Past, present, and future. *Int Immunopharmacol* (2018) 59:391–412. doi: 10.1016/j.intimp.2018.03.002

98. Kumar N, Sood D, Chandra R. Design and optimization of a subunit vaccine targeting COVID-19 molecular shreds using an immunoinformatics framework. *Rsc Adv* (2020) 10:35856–72. doi: 10.1039/D0RA06849G

99. Compton T, Kurt-Jones EA, Boehme KW, Belko J, Latz E, Golenbock DT, et al. Human cytomegalovirus activates inflammatory cytokine responses *via* CD14 and toll-like receptor 2. *J Virol* (2003) 77:4588–96. doi: 10.1128/JVI.77.8.4588-4596.2003

100. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron* (2018) 99:1129–43. doi: 10.1016/j.neuron.2018.08.011

101. Lippert T, Rarey M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J Cheminformatics* (2009) 1:1–12. doi: 10.1186/1758-2946-1-13

102. Du X, Li Y, Xia Y-L, Ai S-M, Liang J, Sang P, et al. Insights into protein–ligand interactions: mechanisms, models, and methods. *Int J Mol Sci* (2016) 17:144. doi: 10.3390/ijms17020144

103. Marín A, Gallardo M, Kato Y, Shirahige K, Gutiérrez G, Ohta K, et al. Relationship between g+ c content, ORF-length and mRNA concentration in saccharomyces cerevisiae. *Yeast* (2003) 20:703–11. doi: 10.1002/yea.992

104. Ismail S, Ahmad S, Azam SS. Vaccinomics to design a novel single chimeric subunit vaccine for broad-spectrum immunological applications targeting nosocomial enterobacteriaceae pathogens. *Eur J Pharm Sci* (2020) 146:105258. doi: 10.1016/j.ejps.2020.105258

105. Tahir Ul Qamar M, Shokat Z, Muneer I, Ashfaq UA, Javed H, Anwar F, et al. Multiepitope-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach. *Vaccines* (2020) 8:288. doi: 10.3390/vaccines8020288

106. Devi A, Chaitanya NS. In silico designing of multi-epitope vaccine construct against human coronavirus infections. *J Biomol Structure Dynamics* (2021) 39:6903–17. doi: 10.1080/07391102.2020.1804460

107. Al-Tawfiq JA, Momattin H, Dib J, Memish ZA. Ribavirin and interferon therapy in patients infected with the middle East respiratory syndrome coronavirus: an observational study. *Int J Infect Dis* (2014) 20:42–6. doi: 10.1016/j.ijid.2013.12.003

108. Du L, Yang Y, Zhou Y, Lu L, Li F, Jiang S. MERS-CoV spike protein: a key target for antivirals. *Expert Opin Ther Targets* (2017) 21:131–43. doi: 10.1080/14728222.2017.1271415

109. Widagdo W, Okba NM, Raj VS, Haagmans BL. MERS-coronavirus: From discovery to intervention. *One Health* (2017) 3:11–6. doi: 10.1016/j.onehlt.2016.12.001

110. Li T, Zhang T, Gu Y, Li S, Xia N. Current progress and challenges in the design and development of a successful COVID-19 vaccine. *Fundam Res* (2021) 1:139–50. doi: 10.1016/j.fmre.2021.01.011

111. Schindewolf C, Menachery VD. Middle East respiratory syndrome vaccine candidates: cautious optimism. *Viruses* (2019) 11:74. doi: 10.3390/v11010074

112. Agrawal AS, Tao X, Algaissi A, Garron T, Narayanan K, Peng B-H, et al. Immunization with inactivated middle East respiratory syndrome coronavirus vaccine leads to lung immunopathology on challenge with live virus. *Hum Vaccines Immunother* (2016) 12:2351–6. doi: 10.1080/21645515.2016.1177688

113. Menachery VD, Gralinski LE, Mitchell HD, Dinnon Iii KH, Leist SR, Yount BLJr., et al. Combination attenuation offers strategy for live attenuated coronavirus vaccines. *J Virol* (2018) 92:e00710–00718. doi: 10.1128/JVI.00710-18

114. Heaton PM. Challenges of developing novel vaccines with particular global health importance. *Front Immunol* (2020) 11:517290. doi: 10.3389/fimmu.2020.517290

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 August 2022
DOI 10.3389/fimmu.2022.941865

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Syed Aun Muhammad,
Bahauddin Zakariya University, Pakistan
Irina Isakova-Sivak,
Institute of Experimental Medicine (RAS), Russia

*CORRESPONDENCE
Jarrod J. Mousa
jarrod.mousa@uga.edu

SPECIALTY SECTION
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

RECEIVED 20 May 2022
ACCEPTED 14 July 2022
PUBLISHED 08 August 2022

CITATION
Huang J, Miller RJ and Mousa JJ (2022) A Pan-Pneumovirus vaccine based on immunodominant epitopes of the fusion protein.
Front. Immunol. 13:941865.
doi: 10.3389/fimmu.2022.941865

A Pan-Pneumovirus vaccine based on immunodominant epitopes of the fusion protein

Jiachen Huang[1,2], Rose J. Miller[1,2] and Jarrod J. Mousa[1,2,3]*

[1]Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, United States, [2]Center for Vaccines and Immunology, College of Veterinary Medicine, University of Georgia, Athens, GA, United States, [3]Department of Biochemistry and Molecular Biology, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, United States

Respiratory syncytial virus (RSV) and human metapneumovirus (hMPV) are two leading causes of severe respiratory infections in children, the elderly, and immunocompromised patients. The fusion (F) protein is the major target of neutralizing antibodies. Recent developments in stabilizing the pre-fusion conformation of the F proteins, and identifying immunodominant epitopes that elicit potent neutralizing antibodies have led to the testing of numerous pre-fusion RSV F-based vaccines in clinical trials. We designed and tested the immunogenicity and protective efficacy of a chimeric fusion protein that contains immunodominant epitopes of RSV F and hMPV F (RHMS-1). RHMS-1 has several advantages over vaccination with pre-fusion RSV F or hMPV F, including a focus on recalling B cells to the most important protective epitopes and the ability to induce protection against two viruses with a single antigen. RHMS-1 was generated as a trimeric recombinant protein, and analysis by negative-stain electron microscopy demonstrated the protein resembles the pre-fusion conformation. Probing of RHMS-1 antigenicity using a panel of RSV and hMPV F-specific monoclonal antibodies (mAbs) revealed the protein retains features of both viruses, including the pre-fusion site Ø epitope of RSV F. Mice immunized with RHMS-1 generated neutralizing antibodies to both viruses and were completely protected from RSV or hMPV challenge. Overall, this study demonstrates protection against two viruses with a single antigen and supports testing of RHMS-1 in additional pre-clinical animal models.

KEYWORDS

RSV (respiratory syncytial virus), human metapneumovirus (hMPV), vaccine, structure-based vaccine design, neutralizing antibodies, fusion protein, epitope

Introduction

Respiratory syncytial virus (RSV) and human metapneumovirus (hMPV) are significant causes of acute lower respiratory tract infections (ALRI) in infants and young children (1–4). RSV was first identified in 1956, and was subsequently recognized as a common cause of respiratory illness in early life (5). The majority of children experience at least one RSV infection before 2 years of age, and infants under 6 months old have a higher risk of severe disease requiring hospitalization (6). hMPV was identified in 2001, and it is the second most common cause of viral lower respiratory infection in children (7). In contrast to RSV, the peak age for infant hospitalizations caused by hMPV infections is 6-12 months old, with a nearly 100% exposure rate by the age of 5 (3). Reinfections of both RSV and hMPV are common throughout life, which usually cause mild symptoms in healthy adults. However, for certain populations including immunocompromised patients, individuals over 65 years of age, and people with underlying conditions such as asthma or chronic obstructive pulmonary disease (COPD), infection with RSV or hMPV may lead to severe bronchiolitis and pneumonia (8–11).

Both hMPV and RSV are enveloped negative-sense single-stranded RNA viruses that belong to the Pneumoviridae family (12). The 15.2 kb genome of RSV consists of 10 genes (encoding 11 proteins), while the 13.3 kb hMPV genome has 8 open reading frames (lacking NS1 and NS2). On the membrane of the Pneumovirus virion, the attachment (G) protein is responsible for viral adherence, the fusion (F) protein mediates membrane fusion and viral entry to host cells (13), while the small hydrophobic (SH) protein is a putative viroporin that forms ion channels and is involved in the prevention of apoptosis of infected cells (14, 15). RSV F and G as well as hMPV F proteins are the main targets of neutralizing antibodies. However, SH is poorly immunogenic and cannot induce neutralizing antibodies, likely due to its small size and low abundance on the virions (16). The M2 gene of RSV and hMPV encodes two overlapping open reading frames for M2-1 and M2-2 proteins. The viral capsid is made from a layer of the matrix (M) and M2-1 proteins (17) where M2-1 connects M with the internal ribonucleoprotein complexes (RNPs) of viral genomic RNA, nucleoprotein (N), large RNA polymerase (L), and the phosphoprotein (P). M2-2 is served as a regulator that is involved in the balance between RNA synthesis and transcription in both RSV and hMPV (18, 19). The RSV non-structural proteins NS1 and NS2 are involved in the inhibition of alpha/beta interferons (IFN-α/β) *in vitro* (20, 21), and RSV infection was attenuated without NS1 and NS2 (22, 23).

The F glycoproteins of RSV and hMPV are highly similar in structure and share ~30% amino acid sequence identity. Both F proteins belong to the class I viral fusion protein family and play indispensable roles in viral attachment as well as membrane fusion. To become fusion competent, the F0 precursor must be cleaved into F1 and F2 subunits that are linked by two disulfide bonds to generate a mature meta-stable homotrimer (24). RSV F is cleaved at two furin cleavage sites with the p27 fragment in between F1 and F2 removed, whereas hMPV F has only one cleavage site that can be cleaved by the host membrane protease TMPRSS2 (25). To initiate the fusion process, the hydrophobic fusion peptide on the N terminus of the F2 subunit is exposed and inserted into the host cell membrane, which triggers the conformational rearrangements that turn the F protein into the stable post-fusion state, and brings the viral and host cell membranes together for lipid mixing.

Multiple antigenic sites have been identified on both RSV F and hMPV F proteins. Among the six known antigenic sites of RSV F, pre-fusion-specific sites Ø and V are targeted by over 60% of neutralizing antibodies in humans (26, 27), indicating these sites are vital for immune recognition and antibody neutralization (28). hMPV F shares three antigenic sites (III, IV, V) with RSV F, as several antibodies have been found to cross-react with RSV and hMPV F at these epitopes (29–32). In addition, the area between sites III and IV was found to be a distinct hMPV site that is recognized by a mAb called DS7 (33, 34). Studies have shown that, unlike RSV F-specific antibodies, the majority of hMPV F-specific antibodies target epitopes present in both pre-fusion and post-fusion conformations, likely due to glycosylation present near pre-fusion-specific sites on the head of hMPV F (35, 36).

Currently, there are no vaccines available for either RSV or hMPV. Previous attempts with formalin-inactivated RSV and hMPV vaccines revealed that low affinity, non-neutralizing F-specific antibodies induced by denatured post-fusion proteins cannot provide protection, and lead to vaccine-enhanced disease (37–41). By stabilizing the RSV F protein in the pre-fusion conformation, several studies have demonstrated improvement in neutralizing antibody titers (42, 43). The bivalent Pfizer RSV vaccine that contains pre-fusion F from RSV subgroups A and B (44) and the pre-fusion RSV F vaccine from GlaxoSmithKline (GSK) (45) are two leading candidates that under phase III clinical trials at present. For hMPV, pre-fusion and post-fusion F proteins induced comparable neutralizing antibodies in mice (36) and immunization with post-fusion F completely protected mice from hMPV challenge (46). Several epitope-focused vaccine designs have been tested for RSV and hMPV F. A head-only RSV F protein boosted titers of neutralizing Abs targeting antigenic sites Ø and II (47). In a different study, RSV F was modified by glycan-masking that blocked poorly neutralizing epitopes on a nanoparticle, which induced a more potent neutralizing Ab response than a pre-fusion F trimer (48). Based on computational protein design strategies, RSV F site II was presented on a scaffold fused with RSV N-based nanoparticles, which boosted subdominant neutralizing antibody responses targeting antigenic site II in mice (49, 50). In addition, RSV F neutralizing antigenic sites (Ø, II, IV) were tested on *de novo* protein scaffolds respectively, and a mixture of these epitope-based immunogens induced focused immune responses toward the target antigenic sites (51). All of the

studies above demonstrate the concept that engineered RSV F epitope-based immunogens can induce and boost neutralizing RSV F antibodies.

The idea of universal vaccine development provides the possibility of preventing multiple viruses/viral variants by a single immunogen. Due to the similarities between RSV and hMPV F, researchers have tried to generate universal RSV/MPV vaccines by grafting the helix-turn-helix motif (site II) from RSV F onto hMPV F, however, this chimera induced neutralizing antibody responses only to hMPV, but not RSV (52). A similar study that grafted RSV F and hMPV F epitopes on pre-fusion and post-fusion F proteins showed that chimeric proteins swapping either site II or site IV can induce cross-neutralizing antibodies in mice, but a challenge with pre-fusion candidates was lacking (53). For the influenza hemagglutinin (HA) protein, chimeric immunogens generated by swapping the HA head with zoonotic subtypes while retaining the conserved HA stem successfully induced antibodies targeting the subdominant HA stem (54, 55).

Based on these findings and the knowledge about structures and the immunodominant epitopes of RSV and hMPV F proteins, we designed a novel chimeric immunogen that contains the head of RSV F and the stem of hMPV F. The **R**SV **h**ead h**M**PV **s**tem construct **1** (RHMS-1) protein was stably expressed as a pre-fusion trimer that preserved the structural features on key antigenic sites for both RSV and hMPV F proteins. RHMS-1 retains immunodominant epitopes of both F proteins, including antigenic sites Ø, V, and II of RSV F, and sites IV, DS7, and III of hMPV F. Immunization of mice with RHMS-1 induced potent neutralizing antibodies that protected mice from both RSV and hMPV challenge. Overall, our data demonstrate that RHMS-1 can be a promising universal vaccine against both Pneumoviruses.

Materials and methods

Expression and purification of proteins

Plasmids encoding cDNAs of Pneumovirus fusion proteins were synthesized (GenScript) and cloned into the pcDNA3.1$^+$ vector. The stable cell line that expresses the hMPV B2 F protein was utilized as previously described (46, 56). The rest of the F proteins and monoclonal antibodies (mAbs) were transiently expressed in Expi293F cells or from hybridomas (for mAbs) as previously described (56). The proteins were harvested from the supernatant of cell cultures and purified by HisTrap Excel (for his-tagged proteins) or Protein G (for antibodies) columns (GE Healthcare Life Sciences). RHMS-1 (trimer), RSV A2 F DsCav1 (trimer) (42), and trypsin-treated hMPV B2 F (monomer) were further purified by size exclusion chromatography on a Superdex S200, 16/600 column (GE Healthcare Life Sciences). The influenza A/California/07/2009 hemagglutinin (HA) protein was expressed and purified as previously described (57).

Negative-stain electron microscopy analysis

Purified RHMS-1 (trimer) was applied on carbon-coated copper grids (5 µL of 10 µg/mL protein solution) for 3 min. The grid was washed in water twice and then stained with Nano-W (Nanoprobes) for 1 min. Negative-stain electron micrographs were acquired using a JEOL JEM1011 transmission electron microscope equipped with a high-contrast 2K-by-2K AMT midmount digital camera.

ELISA of RHMS-1 with mAbs or human/mouse serum

384-well plates (Greiner Bio-One) were coated with 2 µg/mL of antigen in PBS overnight at 4°C. The plates were then washed once with water before blocking for 1 hour with the blocking buffer. Primary mAbs (starting at 20 µg/mL and followed by 3-fold dilutions) or serial dilutions of human/mouse serum (starting with 1:50 and followed by 3-fold dilutions) were applied to wells for 1 hour after three washes with water. Plates were washed with water three times before applying 25 µL of secondary antibody (goat anti-human IgG Fc-AP Southern Biotech, 2048-04; goat anti-mouse IgG Fc-AP Southern Biotech, 1033-04) at a dilution of 1:4,000 in blocking buffer. After incubation for 1 hour, the plates were washed five times with 0.05% PBS-Tween-20, and 25 µL of a PNPP (p-nitrophenyl phosphate) substrate solution (1 mg/mL PNPP in 1 M Tris base) was added to each well. The plates were incubated at room temperature for 1 hour before reading the optical density at 405 nm (OD_{405}) on a BioTek plate reader. Data were analyzed in GraphPad Prism using a nonlinear regression curve fit and the log(agonist)-versus-response function to calculate the binding EC_{50} values. Mouse serum IgG endpoint titers were calculated from the highest dilution of a serum sample that produced OD_{405} readings of >0.3 above the background readings and were shown in a $\log_{10}$ scale as previously described (46).

Binding kinetics of fusion proteins with human mAbs by biolayer interferometry

SEC Purified hMPV F 130BV, RSV F DsCav1, and RHMS-1 were loaded onto HIS1K biosensors at 100 µg/mL for 60 s, then the binding kinetics of mAbs (10 µg/mL) were analyzed by association for 300 s and dissociation for 300 s. Octet data analysis software

was used to analyze the data. Binding curves were independently graphed in GraphPad Prism for data visualization.

ELISA screening of human PBMCs

As previously described (58), peripheral blood mononuclear cells (PBMCs) and plasma were isolated from human subject blood samples using CPT tubes (BD, 362753), and PBMCs were frozen in the liquid nitrogen vapor phase until further use. For serology screening, the plasma samples of 41 subjects were used for ELISA as described above. The IgG binding was quantified by the area under the curve (AUC) values using GraphPad Prism. For PBMC screening, 4 of the 41 subjects were selected, and for each subject, ten million PBMCs were mixed with 8 million previously frozen and gamma-irradiated NIH 3T3 cells modified to express human CD40L, human interleukin-21 (IL-21), and human B-cell activating factor (BAFF) in 80 mL StemCell medium A (StemCell Technologies) containing 1 μg/mL of cyclosporine A (Millipore-Sigma). The mixture of cells was plated in four 96-well plates at 200 μL per well in StemCell medium A. After 6 days, undiluted culture supernatants were screened by ELISA for IgG binding to the RHMS-1 (trimer), RSV A2 DsCav1 F (trimer), and trypsin-treated hMPV B2 F (monomer). Each well is represented by a dot with the $OD_{405\ nm}$ against RSV/hMPV F as the x coordinate, and the $OD_{405\ nm}$ against RHMS-1 as the y coordinate.

Animal immunization and hMPV/RSV challenge

BALB/c mice (6 to 8 weeks old; The Jackson Laboratory) were immunized in a prime-boost regimen with 25 μL purified RHMS-1 (trimer), RSV A2 DsCav1 F (trimer), or trypsin-treated hMPV B2 F (monomer) (20 μg protein/mouse) + 25 μL AddaS03 adjuvant *via* the subcutaneous route into the loose skin over the neck, while mice in control groups were immunized with PBS + AddaS03 adjuvant (8 animals per group). Three weeks after prime, the mice were boosted with the same amount of the antigens + adjuvant. Three weeks after the boost, mice were bled and then intranasally challenged with RSV A2 ($2.8x10^6$ PFU per mouse) or hMPV TN/93-32 ($3x10^5$ PFU per mouse). Mice were sacrificed 5 days post-challenge, and lungs were collected and homogenized for virus titration as previously described (46). Briefly, RSV-challenged lung homogenates were plated on HEp-2 cells (EMEM+2% FBS) while hMPV-challenged lung homogenates were plated on LLC-MK2 cells (EMEM + 5 μg/mL trypsin-EDTA and 100 μg/mL $CaCl_2$) in 24 well plates. After 4-5 days, the cells were fixed with 10% neutral buffered formalin and the plaques of both viruses were immunostained with human mAbs MPV364 (for hMPV) (58) or 101F (for RSV) (59). Plaques were counted under a stereomicroscope.

Virus neutralization assays with immunized mice sera

For serum neutralization assays, the serum of 4/8 mice were randomly picked from each group. Heat-inactivated mouse serum was serially diluted (starting at 1:25 and followed by 3-fold dilutions) and incubated 1:1 with a suspension of hMPV (CAN/97-83 and TN/93-32) or RSV (A2 and B) for 1 hour at room temperature. PBS or serially diluted serum samples from naïve mice were mixed with viruses as the negative control. LLC-MK2 (for hMPV) or HEp-2 cells (for RSV) in 24-well plates were then inoculated with the serum-virus mixture (50 μL/well) for 1 hour and rocked at room temperature before adding the overlay (0.75% Methylcellulose in EMEM+ 2% FBS for HEp-2 cells; 0.75% Methylcellulose in EMEM + 5 μg/mL trypsin-EDTA and 100 μg/mL $CaCl_2$ for LLC-MK2 cells). After 4-5 days, the plaques were stained as described above. The percent neutralization was calculated by (PFU in control wells – PFU in serum wells)/PFU in control wells × 100%.

Competition binding of human mAbs with immunized mice sera by biolayer interferometry

RSV A2 DsCav1 and hMPV F 130BV, as well as all the human reference mAbs were diluted at 100 μg/mL in Octet buffer (PBS + 0.02% Tween20, 0.1% BSA) and loaded onto HIS1K Biosensors. Vaccinated mice serum samples from each vaccination group were pooled and serial diluted (1/10, 1/100, and 1/1000) in ChonBlock blocking buffer (25% ChonBlock buffer diluted in Octet buffer) (Chondrex Inc.), while pooled serum from PBS vaccinated mice was diluted at 1/10 in Chonblock blocking buffer. The loaded biosensors were immersed into wells containing diluted serum samples for 300 s, the non-specific binding signal was then removed by immersing the sensors in Octet buffer. Following this, biosensors were immersed into wells containing 100 μg/mL of a human reference mAb that binds to different epitopes for 300 s. Percent inhibition of pooled mice serum was calculated by (1 – max signal of the vaccinated serum/max signal of the naive serum) × 100%.

Analysis of mouse IgG isotypes in mouse serum

IgG isotype responses in mouse serum were measured by ELISA. High-binding 384 well plates were coated with antigens corresponding to immunization groups (RHMS-1, RSV A2 DsCav1, hMPV B2 F monomer) as described above. After blocking and washing the plate, serially diluted serum samples

(pooled together for each mouse group) in blocking buffer (1:100, followed by 3-fold dilutions) were added to the plate. After incubation for 1 hr, the plates were washed three times with water. Each IgG isotype secondary antibody (goat anti-mouse IgG_1 Fc-AP Southern Biotech, 1070-04; goat anti-mouse IgG_{2a} Fc-AP Southern Biotech, 1080-04; goat anti-mouse IgG_{2b} Fc-AP Southern Biotech, 1090-04; goat anti-mouse IgG_{2c} Fc-AP Southern Biotech, 1079-04; goat anti-mouse IgG_3 Fc-AP Southern Biotech, 1100-04) was diluted 1:4,000 in blocking buffer. Following incubation, the plate was developed with the PNPP substrate solution, the OD_{405} values were measured, and the IgG endpoint titers were determined and graphed on GraphPad Prism as described above.

Th1/Th2 cytokine analysis of mouse lungs

Lung homogenates were thawed on ice and centrifuged at 10,000 x g for 10 minutes. The supernatant (pooled together for each mouse group) was used for measuring IFN-γ, IL-2, IL-4, and IL-10 levels by using a mouse Th1/Th2 ELISA kit (ThermoFisher, 88–7711-44) per the manufacturer's instructions. Undiluted samples were normalized to the baseline OD_{405} values and the concentrations of each cytokine were calculated based on each respective standard curve.

Results

RHMS-1 was expressed as a pre-fusion trimer

We designed the RHMS-1 protein based on the pre-fusion structures of RSV F (5UDE) and hMPV F (5WB0) using ChimeraX (60), and a model of RHMS-1 based on these structures is shown in Figure 1A. The design of RHMS-1 maintains the signal peptide, two cleavage sites, DsCav1 mutations (S115C, S190F, V207L, S290C) (42), and the fusion peptide of RSV F. Part of the F2 N-terminus (residues 26-54) and the F1 C-terminus (residues 315-531) was replaced by the homologous hMPV F sequences, with two junctions located on β2 and β7 strands. Two glycosylation sites (RSV F-N70, hMPV F-N353) are retained in RHMS-1. A GCN4 trimerization domain and a hexa-histidine tag were appended to the F1 C-terminus (Figure 1B). Intact RSV F sites II, V, and Ø and hMPV F sites IV, site III, and DS7 site were adopted from the original sequence (Figure 1C).

RHMS-1 was expressed in HEK293F cells and size exclusion chromatography (SEC) showed RHMS-1 was mainly expressed as a trimeric protein, but the size was slightly bigger than RSV F trimers (Figure 2A). For hMPV F, trimers and monomers were observed after trypsin treatment as previously described (46), but the size of trimeric hMPV F is smaller than RHMS-1 and RSV F, likely due to trypsinization (Figure 2A). Like RSV F, RHMS-1 was deduced to be cleaved after expression, as the F2 domain (~15 kDa) was observed on the gel under reducing conditions (Figure 2B), and the sizes of RHMS-1 and RSV F bands are consistent with the peaks shown in Figure 2A. Negative-stain EM analysis demonstrated that the majority of RHMS-1 particles were in the pre-fusion conformation based on "ball-like" structures resembling pre-fusion RSV and hMPV F (36, 42) (Figure 2C), indicating the DsCav1 mutations work well in stabilizing the structure of RHMS-1.

RHMS-1 shares immunological features of both RSV and hMPV F

To determine if RHMS-1 retains the correct conformation of each antigenic site, mAbs specifically targeting these sites were tested for binding by ELISA (Figure 3A). For both RSV F and RHMS-1, mAb D25 binds to site Ø (61), mAb hRSV90 binds to site V (62), and motavizumab binds to site II (63, 64) at similar EC_{50} values (Figure 3B). For both hMPV F and RHMS-1, mAbs DS7 and MPV196 bind to the DS7 site (33, 58). mAb 101F binds to all three antigens on site IV (29) while mAb MPE8 binds to site III on RSV F and RHMS-1 (65), but not monomeric hMPV F, likely due to the cross-protomer epitope that is only partially displayed on hMPV F monomer. The binding site of mAb MPV364 partially overlaps with hMPV site III, but it was also predicted to interact with the head of hMPV F (58), and mAb MPV458 binds to the 66-87 peptide on the head of hMPV F (56), therefore, both mAbs MPV364 and MPV458 do not bind to RHMS-1 as expected. In addition, we determined the binding kinetics of trimeric hMPV F, RSV F, and RHMS-1 against reference mAbs through BLI (Figure 3C). In consistence with the ELISA data, mAbs MPE8 and 101F bind to all three constructs. mAbs D25 and motavizumab showed similar binding kinetics with RSV F and RHMS-1, while mAbs DS7 and a hMPV F-specific site IV mAb, MPV481 (66), showed similar binding kinetics with hMPV F and RHMS-1. Overall, our data suggest that RHMS-1 maintains the conformational structures on the included antigenic epitopes from the RSV F and hMPV F proteins.

RHMS-1 can be recognized by B cells pre-exposed to RSV F or hMPV F

To verify if epitopes on RHMS-1 can be recognized by the human immune system in a similar manner compared to RSV F or hMPV F, we screened plasma IgG responses from 41 human subjects against these proteins by ELISA. Overall, a simple linear regression fit showed positive correlations of serum IgG bindings for both RSV F vs. RHMS-1 (Figure 4A) and hMPV F vs. RHMS-1 (Figure 4B), but no correlation was observed for an irrelevant antigen, influenza A/California/07/2009 HA protein vs. RHMS-1

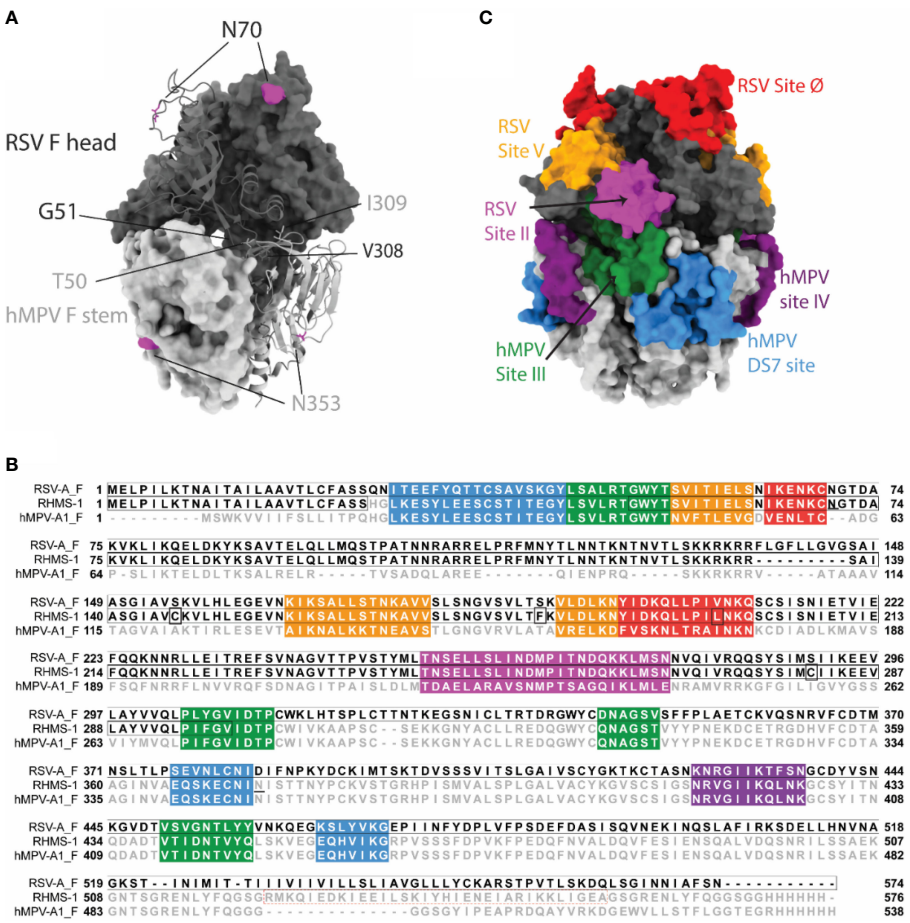


FIGURE 1
RHMS-1 protein design. **(A)** The diagram generated with the head (gray) of pre-fusion RSV F (5UDE) and the stem (silver) of pre-fusion hMPV F (5WB0) shows one protomer in cartoon and the rest of two protomers in surface. The glycosylation sites are shown in magenta. **(B)** Sequence alignment of RSV-A F, RHMS-1, and hMPV-A1 F generated by Jalview. The sequences of known antigenic sites are highlighted: RSV site Ø – red, RSV site V – orange, RSV site II – magenta, hMPV site III – forest, hMPV site IV – purple, hMPV DS7 site – blue. In RHMS-1 sequence, four Ds-Cav1 mutations are circled in boxes, two N-linked glycosylation sites are underlined and the GCN4 trimerization domain is circled in red dashed box. **(C)** The antigenic sites colored in accordance with the sequences highlighted in Figure 1B are displayed on the diagram. Both figure 1A and C were made by ChimeraX.

(Figure 4C). The serological screening data suggest epitopes are conserved between the native F proteins and the epitopes included on RHMS-1. The antibody responses at the cellular level were also tested by measuring the binding of supernatant from stimulated B cells in four subjects. The B cells in PBMCs were activated through coincubation with NIH 3T3 cells expressed human CD40L, human interleukin-21 (IL-21), and human B-cell activating factor (BAFF) to stimulate growth and IgG secretion to the culture supernatant as previously described (58). For all of the subjects we tested, the majority of RSV F-positive B cells are also positive for RHMS-1 (Figure 5A). This finding is expected as the majority of human B cells target the head of the RSV F protein, which is retained in RHMS-1. Such correlations are still present for hMPV F (Figure 5B), although the frequencies of hMPV F-positive B cells are generally lower than RSV F-positive B cells, therefore, populations of hMPV F negative, RHMS-1 positive B cells close to the Y axes were seen.

Vaccination with RHMS-1 elicits potent neutralizing antibodies and cross-protection in mice

To evaluate the immunological properties of RHMS-1, it was tested as a vaccine in the mouse model. BALB/c mice were subcutaneously primed and boosted with 20 μg of RHMS-1, RSV F DsCav1, hMPV monomeric B2 F, or PBS in an emulsion formulated with AddaS03 adjuvant and then challenged with RSV or hMPV (Figure 6A). All RHMS-1 vaccinated mice showed serum IgG binding titers against both RSV F DsCav1

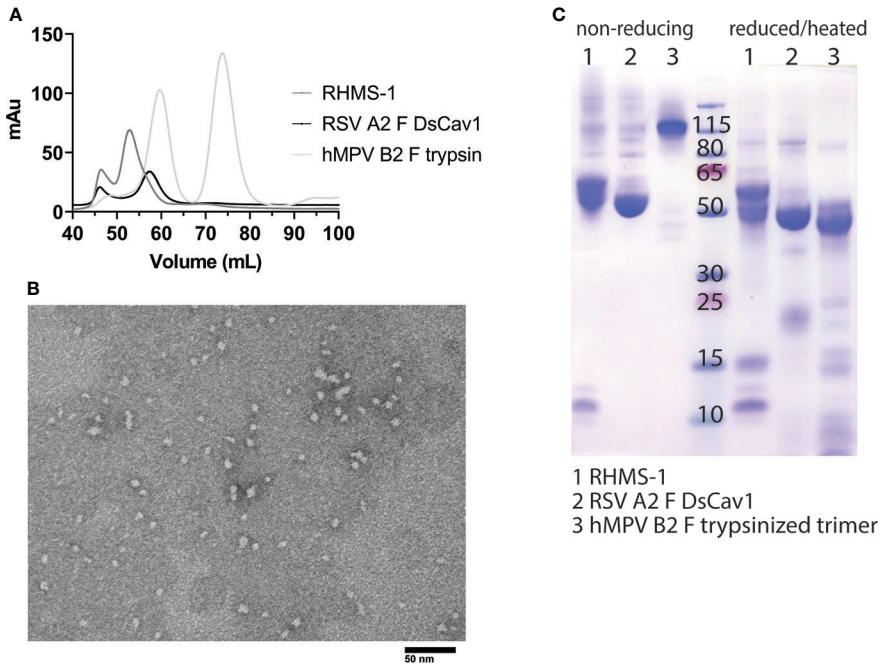


FIGURE 2
Purification and Negative-stain EM of RHMS-1. **(A)** Size exclusion chromatography curves of RHMS-1 (gray), RSV A2 F DsCav1 (black), and trypsinized hMPV B2 F (silver). mAu: milli absorbance units. **(B)** SDS-PAGE of F proteins in non-reducing and reduced/heated conditions. Numbers shown at the marker lane represent kDa values. **(C)** Representative negative-stain electron micrograph of RHMS-1 obtained from fractions 50-60 mL from the size exclusion chromatogram shown in **(A)**.

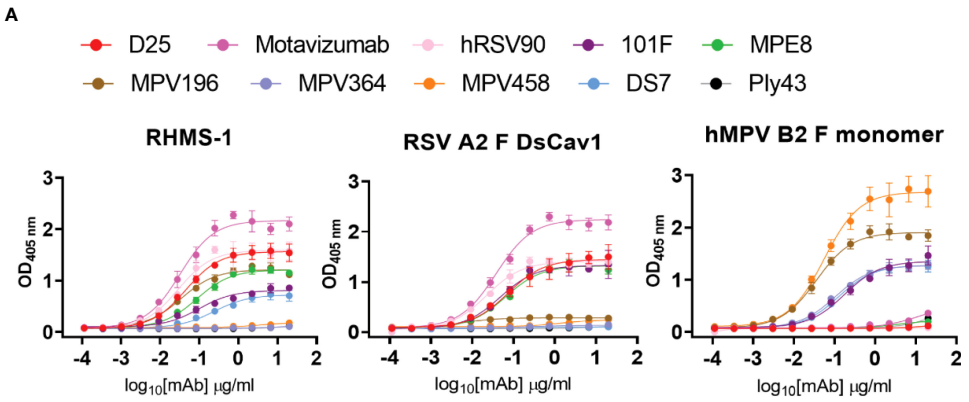
and hMPV monomeric B2 F proteins (Figure 6B, C). Representative viruses from each RSV subgroup and each hMPV genotype were neutralized by RHMS-1 immunized mouse serum 3 weeks after the boost (Figure 6D, E). RSV F DsCav1 immunization failed to induce IgG that cross-recognize hMPV monomeric B2 F, while hMPV monomeric B2 F immunized mice showed moderate binding, but non-neutralizing IgG against RSV. Three weeks after the boost, mice were intranasally challenged with RSV A2 or hMPV TN/93-32. The virus titers in the lung homogenate were determined 5 days post challenge. Vaccination with RHMS-1 completely protected the mice from the challenges of both viruses, while RSV A2 F DsCav1 and hMPV B2 F monomer vaccinated groups protected mice only against the autologous virus (Figure 6F, G).

To determine if RHMS-1 induced epitope-specific antibodies against RSV F head and hMPV F stem, pooled Day 42 serum samples were used for competition binding assay by BLI (Figure 6H). For hMPV F 130BV, RHMS-1 vaccinated serum showed potent competition with mAbs 101F, MPE8, and DS7, but not with the hMPV F head-specific antibody mAb MPV467 (66). For RSV F DsCav1, RHMS-1 vaccinated serum (1/10 diluted) inhibited the binding of mAbs MPE8, 101F, D25, and motavizumab by ~50% compared to the mock vaccinated serum. Interestingly, RHMS-1 induced site III and site IV-competing antibodies for both RSV F and hMPV F, which may be due to the similar sequences at these 2 sites (Figure 1B). Another potential explanation is that site III and site IV are located at the interface of RHMS-1, and certain binding angles of antibodies on the RSV F head may interfere with the antibodies that bind to the hMPV F stem or vice versa.

No sign of vaccine-enhanced disease was observed based on IgG isotype and cytokine profiles

To further assess the immune response elicited by vaccination and challenge, we determined the IgG isotypes elicited by immunization in mice (Figure 7A). Serum for each vaccine group was measured for binding against the vaccinating antigen, and we observed RHMS-1 vaccinated mice had a predominantly IgG_1 immune response, suggestive of a Th2-skewed immune response (67). Similar results were obtained for hMPV B2 F and RSV A2 F DsCav1 immunized mice with additional IgG_{2a} antibody detected for these groups. BALB/c mice have a Th2-skewed immune response (68), and these data indicate the AddaS03 adjuvant cannot overcome this bias for these antigens. We further assessed the in mouse lungs cytokine response after RSV and hMPV challenge (Figure 7B). IFN-γ,

10.3389/fimmu.2022.941865



B

EC50 (ng/mL)

300

10

EC50 (ng/mL)	Antigenic site	RHMS-1	RSV A2 F DsCav1	MPV B2 F monomer
D25	RSV Ø	45	72	
hRSV90	RSV V	28	22	
Motavizumab	RSV II	32	38	
101F	RSV/hMPV IV	99	48	177
MPE8	RSV/hMPV III	86	64	
MPV196	hMPV III/DS7	34		36
MPV364	hMPV III			105
MPV458	hMPV 66-87			55
DS7	hMPV DS7	238		127
Ply43	Neg ctrl			

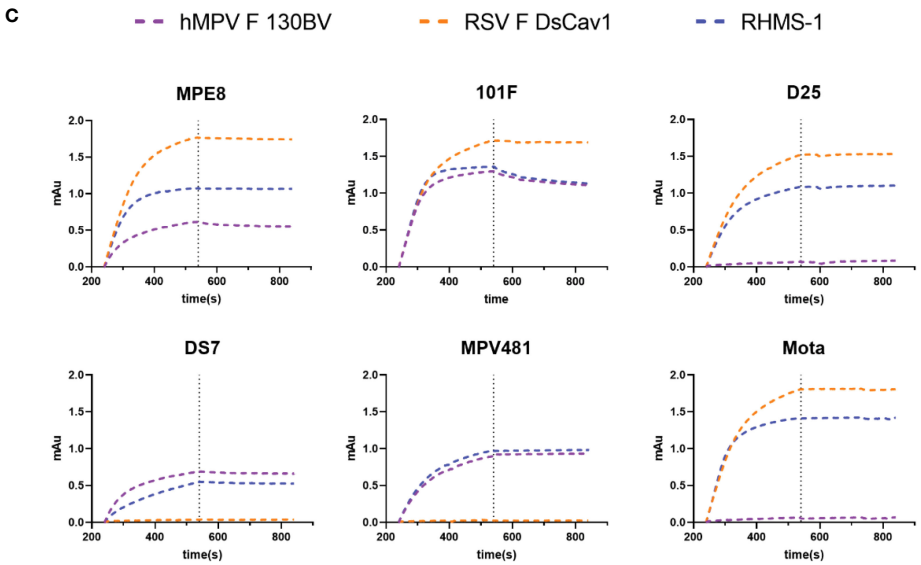


FIGURE 3
Antigenic site-specific mAbs binding to F proteins. **(A)** ELISA binding curves of mAbs targeting different RSV/hMPV F antigenic sites against RHMS-1, RSV A2 F DsCav1, and trypsinized hMPV B2 F monomer. **(B)** EC_{50} values of the binding curves in **(A)**. The binding curves and the EC_{50} values were generated by GraphPad Prism. **(C)** BLI binding kinetics of RHMS-1, RSV A2 F DsCav1, and hMPV F 130BV with reference mAbs.

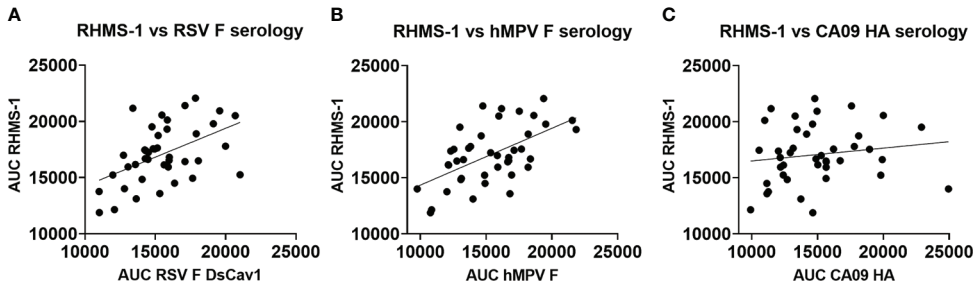


FIGURE 4
Serology of human plasma against F proteins. Area under the curve analysis of plasma IgG binding to RHMS-1 vs. RSV A2 F DsCav1 **(A)**, RHMS-1 vs. trypsinized hMPV B2 F monomer **(B)**, and RHMS-1 vs. influenza HA (A/California/07/2009). Each dot represents one subject, and the lines indicate the linear regression fit of the data sets (R^2 = 0.174, p=0.0067 for **(A)**; R^2 = 0.324, p=0.0001 for **(B)**; R^2 = 0.023, p=0.347 for **(C)**]. Figures and data analysis was generated by GraphPad Prism.

IL-2, IL-4, and IL-10 responses were measured and compared to control vehicle immunized and challenged mice, as well as mice that were not immunized or challenged. We observed increases in IFN-γ responses for all challenged groups. IL-4 responses, indicating a Th2 immune response, were highest in the RSV A2 F and hMPV B2 F immunized mice compared to RHMS-1, suggesting RHMS-1 can limit the Th2 immune responses associated with potential vaccine enhanced disease.

Discussion

Multiple protein engineering strategies have been investigated to generate cross-protective or epitope-based antigens against RSV F or hMPV F. Previous attempts at grafting a single antigenic site on RSV F or hMPV F from one to another induced cross-neutralizing antibodies but very limited protection against the heterologous virus challenge, likely due to the epitopes on the backbone still dominate the immune responses (53). RHMS-1 contains multiple immunodominant epitopes of both RSV F and hMPV F in relatively equal proportions, including at least three RSV F-specific and three hMPV F-specific antigenic sites. Therefore, more balanced immune responses against both RSV F and hMPV F would be expected, and it is less likely to drive escape mutations focused on a single epitope. After prime and boost, RHMS-1 induced comparable levels of hMPV F/RSV F-specific serum IgG titers. Although the serum neutralization against RSV is not as potent as that against hMPV, RHMS-1 immunization completely protected the mice from both RSV and hMPV challenges, suggesting RHMS-1 is a promising antigen that can be used as a vaccine to induce cross-neutralizing and cross-

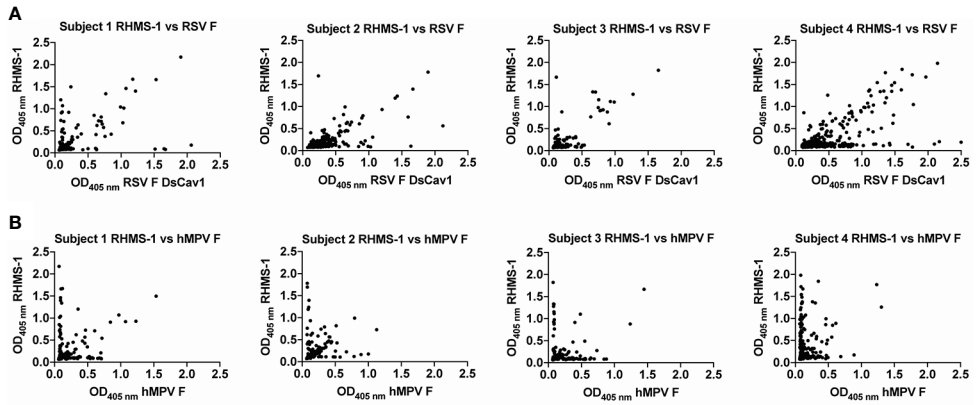


FIGURE 5
Human PBMCs binding to F proteins. ELISA $OD_{405\ nm}$ values of B cell culture supernatants binding to RHMS-1 vs. RSV A2 F DsCav1 **(A)** and RHMS-1 vs. trypsinized hMPV B2 F monomer **(B)**. Each dot represents the B cell supernatant in a single well of a 384 well plate initially containing 20,000 PBMCs. Figures were generated by GraphPad Prism.

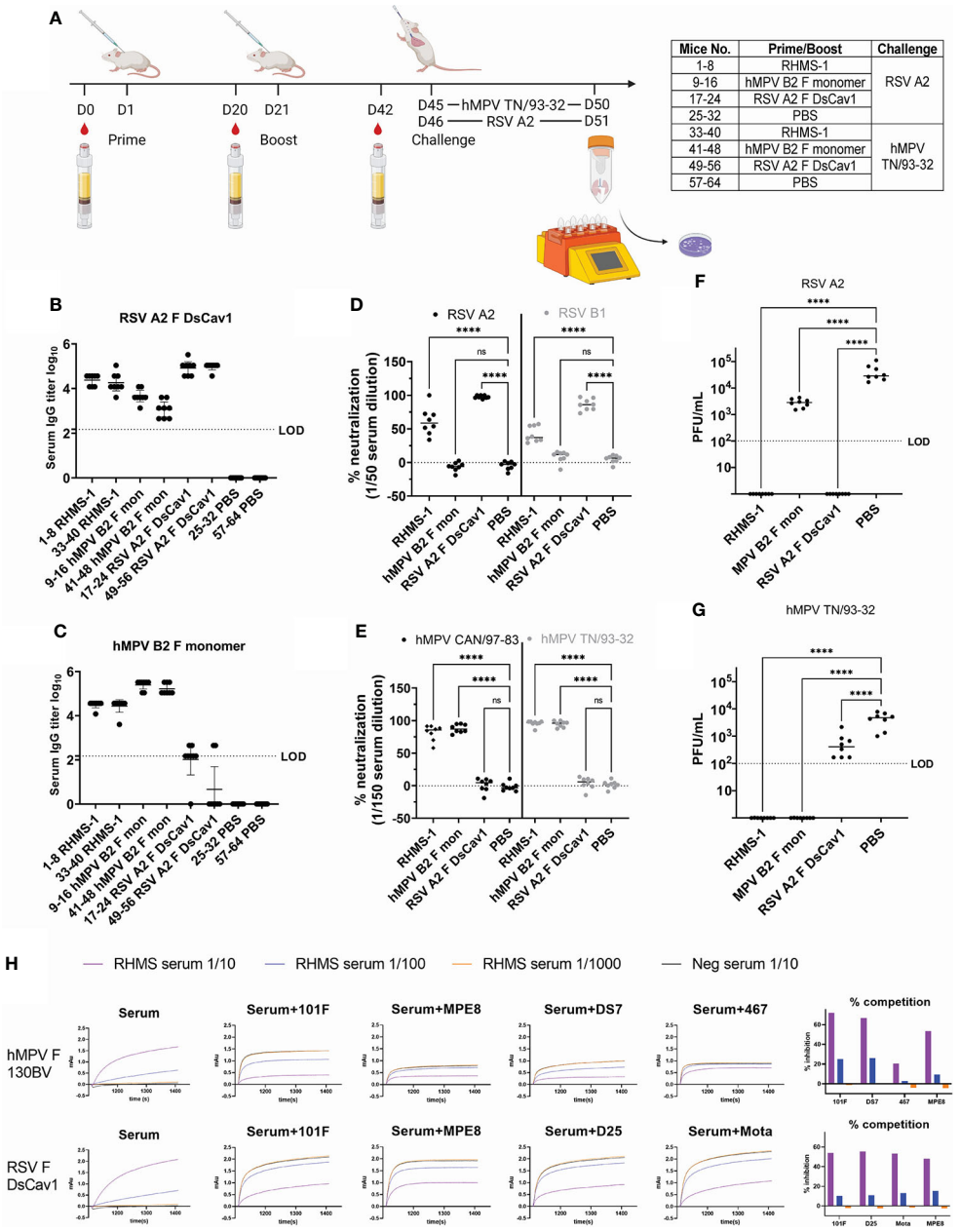


FIGURE 6
Mouse immunization and challenge studies. **(A)** Study regimen. Day 42 serum IgG titers against RSV A2 F DsCav1 **(B)** and trypsinized hMPV B2 F monomer **(C)** before challenge. **(D)** Serum neutralization against RSV A2 and B1 strains at 1/50 dilution and against **(E)** hMPV CAN/97-83 and TN/93-32 at a 1/150 dilution. Viral titers for RSV A2 **(F)** and hMPV TN/93-32 **(G)** in the lung homogenates of mice 5 days post-challenge. LOD, limit of detection. **(H)**. Competition binding of RHMS-1 (serial diluted at 1/10, 1/100, and 1/1000) or mock (1/10 diluted) vaccinated Day 42 serum with reference mAbs. ns, not significant. ****P < 0.0001.

protecting antibodies against RSV and hMPV. These findings also suggest additional optimization of antigen and adjuvants is needed to optimize the elicited neutralizing antibody responses.

Our ELISA screening data showed that the human subjects tested had pre-existing immunity against both RSV F and hMPV F. Interestingly, subjects had an overall higher frequency of RSV F-specific B cells than hMPV F-specific B cells, which is likely due to the higher prevalence of RSV than hMPV. Since initial exposures to an antigen can influence subsequent immune responses against similar antigens, termed original antigenic sin (69), pre-existing immunity to RSV or hMPV may affect the efficacy of RHMS-1 or other RSV and

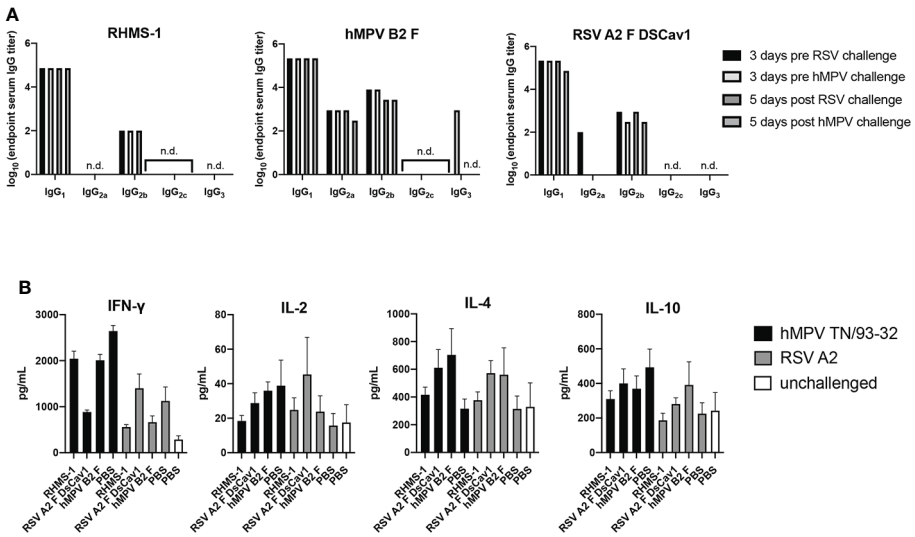


FIGURE 7
TH1/TH2 antibody and cytokine response after immunization and challenge. **(A)** Mouse serum IgG isotype ELISA against each vaccinating antigen (RHMS-1, hMPV B2 F, RSV A2 DsCav1) three days prior to viral challenge and five days post-challenge (hMPV TN/93-32, RSV A2). n.d., no detection above OD405 = 0.3. **(B)** Cytokine ELISA in the lung homogenates of mice five days post-challenge (hMPV TN/93-32, RSV A2). Error bars are the standard deviation.

hMPV F-based vaccine candidates. To address this problem, future studies will be needed to test RHMS-1 in animal models that are pre-immune to RSV and hMPV. Furthermore, despite the encouraging results obtained from the RHMS-1 immunization/challenge in this study, it is still necessary to compare the efficacy of RHMS-1 versus a mixture of RSV F and hMPV F in animal models.

Previous studies have shown that mice immunized with either pre-fusion RSV F or post-fusion hMPV F did not induce significant cross-neutralizing antibodies (29). Similar results were observed in this study with pre-fusion RSV F and monomeric hMPV F. The serum of 9 out of 16 mice immunized with RSV F + AddaS03 showed little binding to monomeric hMPV F just above the detection limit, while all of the monomeric hMPV F + AddaS03 immunized mice serum had moderate binding to RSV F. However, the serum of both groups failed to cross-neutralize the viruses *in vitro* (Figures 6B–E). Interestingly, the lung virus counts in RSV F immunized/hMPV challenged and hMPV F immunized/RSV challenged groups are reduced ~10 fold compared to mock immunized/hMPV challenged and mock immunized/RSV challenged groups, respectively (Figure 6F), indicating poorly and non-neutralizing, cross-reactive antibodies could play a role in limiting virus replication in the lungs of mice, possibly through Fc-mediated effector functions that need to be further characterized. Based on our current data, vaccine-induced immunopathology is not expected in animals treated with RHMS-1, because 1) dominant pre-fusion conformation particles of RHMS-1 were observed by negative-stain EM; 2) there is no significant change of Th2-associated cytokine levels in the post-challenge lungs from RHMS-1 vaccinated groups vs. the other control groups. Nevertheless, future studies in more permissive animal models such as cotton rats and African Green Monkeys will be needed to further evaluate the efficacy and safety of RHMS-1, as well as studies using a formalin-inactivated virus comparators to assess pathology. While not directly assessed in this report, it is not expected that RHMS-1 would cause liver toxicity due to a record of safety for RSV F and hMPV F in animal models and humans (70–73).

Although the RHMS-1 construct appears to be a pre-fusion trimer by negative-stain electron microscopy, the stability of this protein requires further evaluation, and could likely be optimized to improve both stability and antigenicity using computational protein design tools. For example, by using interprotomer disulfides (IP-DSs) that link protomers of the hMPV F trimer, both stabilized pre-fusion and post-fusion F proteins elicited significantly higher neutralizing responses than the hMPV F proteins without IP-DSs (74). Such a strategy could also be applied to stabilize the RHMS-1 construct. In addition, it would be interesting to further explore other vaccine platforms for RHMS-1. For example, mRNA vaccines can elicit robust antibody titers as well as T cell responses, which were not examined in this study. Live-attenuated vaccine would be another potential platform as it can induce mucosal immunity to prevent viral transmission, however, the fusogenicity of RHMS-1 will need to be assessed to determine if incorporation into live-attenuated vaccine platforms is warranted.

In summary, we generated and evaluated a chimeric RSV F and hMPV F protein, RHMS-1. To our knowledge, this is the first immunogen that elicits a protective immune response against both RSV and hMPV. Moreover, RHMS-1 can be readily applied to both traditional and novel vaccine delivery platforms like viral vectors, VLPs, nanoparticles, and mRNA. Further optimization of RHMS-1 could lead to a safe and effective universal Pneumovirus vaccine.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by University of Georgia IRB, and subjects were recruited to the University of Georgia Clinical and Translational Research Unit. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by University of Georgia IACUC committee.

Author contributions

JH and RM conducted the experiments. JH, RM, and JJM. analyzed the data. JH and JJM wrote and revised the manuscript. RM revised the manuscript. JJM procured funding. All authors contributed to the article and approved the submitted version.

Funding

These studies were supported by National Institutes of Health grants 1R01AI143865 (JJM) and 1K01OD026569 (JJM). The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Conflict of interest

JH and JJM are listed as inventors on a provisional patent application describing the vaccine candidate.

The remaining author declarers that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: A systematic review and meta-analysis. *Lancet* (2010) 375:1545–55. doi: 10.1016/S0140-6736(10)60206-1

2. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: A systematic review and modelling study. *Lancet* (2017) 390:946–58. doi: 10.1016/S0140-6736(17)30938-8

3. Edwards KM, Zhu Y, Griffin MR, Weinberg GA, Hall CB, Szilagyi PG, et al. Burden of human metapneumovirus infection in young children. *N Engl J Med* (2013) 368:633–43. doi: 10.1056/NEJMoa1204630

4. Schuster JE, Williams JV. Human metapneumovirus. *Pediatr Rev* (2013) 34:558. doi: 10.1542/pir.34.12.558

5. Chanock R, Finberg L. Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA). II. Epidemiologie aspects of infection in infants and young children. *Am J Hyg* (1957) 66:291–300. doi: 10.1093/oxfordjournals.aje.a119902

6. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, et al. The burden of respiratory syncytial virus infection in young children. *N Engl J Med* (2009) 360:588–98. doi: 10.1056/NEJMoa0804877

7. den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RAM, et al. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med* (2001) 7:719–24. doi: 10.1038/89098

8. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adults. *N Engl J Med* (2005) 352:1749–59. doi: 10.1056/NEJMoa043951

9. Falsey AR, McElhaney JE, Beran J, Van Essen GA, Duval X, Esen M, et al. Respiratory syncytial virus and other respiratory viral infections in older adults with moderate to severe influenza-like illness. *J Infect Dis* (2014) 209:1873–81. doi: 10.1093/infdis/jit839

10. Dowell SF, Anderson LJ, Gary HEJr., Erdman DD, Plouffe JF, File TMJr., et al. Respiratory syncytial virus is an important cause of community-acquired lower respiratory infection among hospitalized adults. *J Infect Dis* (1996) 174:456–62. doi: 10.1093/infdis/174.3.456

11. Panda S, Mohakud NK, Pena L, Kumar S. Human metapneumovirus: Review of an important respiratory pathogen. *Int J Infect Dis* (2014) 25:45–52. doi: 10.1016/j.ijid.2014.03.1394

12. Afonso CL, Amarasinghe GK, Bányai K, Bào Y, Basler CF, Bavari S, et al. Taxonomy of the order mononegavirales: update 2016. *Arch Virol* (2016) 161:2351–60. doi: 10.1007/s00705-016-2880-1

13. Battles MB, McLellan JS. Respiratory syncytial virus entry and how to block it. *Nat Rev Microbiol* (2019) 17:233–45. doi: 10.1038/s41579-019-0149-x

14. Wilson RL, Fuentes SM, Wang P, Taddeo EC, Klatt A, Henderson AJ, et al. Function of small hydrophobic proteins of paramyxovirus. *J Virol* (2006) 80:1700–9. doi: 10.1128/JVI.80.4.1700-1709.2006

15. Bao X, Kolli D, Liu T, Shan Y, Garofalo RP, Casola A. Human metapneumovirus small hydrophobic protein inhibits NF-$κ$B transcriptional activity. *J Virol* (2008) 82:8224–9. doi: 10.1128/JVI.02584-07

16. Schepens B, Sedeyn K, Vande Ginste L, De Baets S, Schotsaert M, Roose K, et al. Protection and mechanism of action of a novel human respiratory syncytial virus vaccine candidate based on the extracellular domain of small hydrophobic protein. *EMBO Mol Med* (2014) 6:1436–54. doi: 10.15252/emmm.201404005

17. Marty A, Meanger J, Mills J, Shields B, Ghildyal R. Association of matrix protein of respiratory syncytial virus with the host cell membrane of infected cells. *Arch Virol* (2003) 149:199–210. doi: 10.1007/s00705-003-0183-9

18. Bermingham A, Collins PL. The M2–2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proc Natl Acad Sci* (1999) 96:11259–64. doi: 10.1073/pnas.96.20.11259

19. Buchholz UJ, Biacchesi S, Pham QN, Tran KC, Yang L, Luongo CL, et al. Deletion of M2 gene open reading frames 1 and 2 of human metapneumovirus: Effects on RNA synthesis, attenuation, and immunogenicity. *J Virol* (2005) 79:6588–97. doi: 10.1128/JVI.79.11.6588-6597.2005

20. Spann KM, Tran KC, Collins PL. Effects of nonstructural proteins NS1 and NS2 of human respiratory syncytial virus on interferon regulatory factor 3, NF-$κ$B, and proinflammatory cytokines. *J Virol* (2005) 79:5353–62. doi: 10.1128/JVI.79.9.5353-5362.2005

21. Lo MS, Brazas RM, Holtzman MJ. Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness. *J Virol* (2005) 79:9315–9. doi: 10.1128/JVI.79.14.9315-9319.2005

22. Jin H, Zhou H, Cheng X, Tang R, Munoz M, Nguyen N. Recombinant respiratory syncytial viruses with deletions in the NS1, NS2, SH, and M2-2 genes are attenuated *in vitro* and *in vivo*. *Virology* (2000) 273:210–8. doi: 10.1006/viro.2000.0393

23. Zhang W, Yang H, Kong X, Mohapatra S, Juan-Vergara HS, Hellermann G, et al. Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. *Nat Med* (2005) 11:56–62. doi: 10.1038/nm1174

24. Jardetzky TS, Lamb RA. Activation of paramyxovirus membrane fusion and virus entry. *Curr Opin Virol* (2014) 5:24–33. doi: 10.1016/j.coviro.2014.01.005

25. Shirogane Y, Takeda M, Iwasaki M, Ishiguro N, Takeuchi H, Nakatsu Y, et al. Efficient multiplication of human metapneumovirus in vero cells expressing the transmembrane serine protease TMPRSS2. *J Virol* (2008) 82:8942–6. doi: 10.1128/JVI.00676-08

26. Ngwuta JO, Chen M, Modjarrad K, Joyce MG, Kanekiyo M, Kumar A, et al. Prefusion f–specific antibodies determine the magnitude of RSV neutralizing activity in human sera. *Sci Transl Med* (2015) 7:309ra162–309ra162. doi: 10.1126/scitranslmed.aac4241

27. Gilman MSA, Castellanos CA, Chen M, Ngwuta JO, Goodwin E, Moin SM, et al. Rapid profiling of RSV antibody repertoires from the memory b cells of naturally infected adult donors. *Sci Immunol* (2016) 1:1–12. doi: 10.1126/sciimmunol.aaj1879

28. Andreano E, Paciello I, Bardelli M, Tavarini S, Sammicheli C, Frigimelica E, et al. The respiratory syncytial virus (RSV) prefusion f-protein functional antibody repertoire in adult healthy donors. *EMBO Mol Med* (2021) 13(6):e14035. doi: 10.15252/emmm.202114035

29. Más V, Rodriguez L, Olmedillas E, Cano O, Palomo C, Terrón MC, et al. Engineering, structure and immunogenicity of the human metapneumovirus f protein in the postfusion conformation. *PLoS Pathog* (2016) 12:e1005859. doi: 10.1371/journal.ppat.1005859

30. Wen X, Mousa JJ, Bates JT, Lamb RA, Crowe JE, Jardetzky TS. Structural basis for antibody cross-neutralization of respiratory syncytial virus and human metapneumovirus. *Nat Microbiol* (2017) 2:16272. doi: 10.1038/nmicrobiol.2016.272

31. Mousa JJ, Binshtein E, Human S, Fong RH, Alvarado G, Doranz BJ, et al. Human antibody recognition of antigenic site IV on pneumovirus fusion proteins. *PloS Pathog* (2018) 14:e1006837. doi: 10.1371/journal.ppat.1006837

32. Xiao X, Tang A, Cox KS, Wen Z, Callahan C, Sullivan NL, et al. Characterization of potent RSV neutralizing antibodies isolated from human memory b cells and identification of diverse RSV/hMPV cross-neutralizing epitopes. *MABS* (2019) 11(8):1415–27. doi: 10.1080/19420862.2019.1654304

33. Williams JV, Chen Z, Cseke G, Wright DW, Keefer CJ, Tollefson SJ, et al. A recombinant human monoclonal antibody to human metapneumovirus fusion protein that neutralizes virus in vitro and is effective therapeutically *in vivo*. *J Virol* (2007) 81:8315–24. doi: 10.1128/JVI.00106-07

34. Wen X, Krause JC, Leser GP, Cox RG, Lamb RA, Williams JV, et al. Structure of the human metapneumovirus fusion protein with neutralizing antibody identifies a pneumovirus antigenic site. *Nat Struc Mol Biol* (2012) 19:461–3. doi: 10.1038/nsmb.2250

35. Pilaev M, Shen Y, Carbonneau J, Venable M-C, Rhéaume C, Lavigne S, et al. Evaluation of pre- and post-fusion human metapneumovirus f proteins as subunit vaccine candidates in mice. *Vaccine* (2020) 38:2122–7. doi: 10.1016/j.vaccine.2020.01.047

36. Battles MB, Más V, Olmedillas E, Cano O, Vázquez M, Rodr\'\iguez L, et al. Structure and immunogenicity of pre-fusion-stabilized human metapneumovirus f glycoprotein. *Nat Commun* (2017) 8:1–11. doi: 10.1038/s41467-017-01708-9

37. Killikelly AM, Kanekiyo M, Graham BS. Pre-fusion f is absent on the surface of formalin-inactivated respiratory syncytial virus. *Sci Rep* (2016) 6:34108. doi: 10.1038/srep34108

38. Murphy BR, Prince GA, Walsh EE, Kim HW, Parrott RH, Hemming VG, et al. Dissociation between serum neutralizing and glycoprotein antibody responses of infants and children who received inactivated respiratory syncytial virus vaccine. *J Clin Microbiol* (1986) 24:197–202. doi: 10.1128/jcm.24.2.197-202.1986

39. Murphyl BR, Walsh EE. Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity. *J Clin Microbiol* (1988) 26(8):1595–7. doi: 10.1128/jcm.26.8.1595-1597.1988

40. de Swart RL, van den Hoogen BG, Kuiken T, Herfst S, van Amerongen G, Yüksel S, et al. Immunization of macaques with formalin-inactivated human metapneumovirus induces hypersensitivity to hMPV infection. *Vaccine* (2007) 25:8518–28. doi: 10.1016/j.vaccine.2007.10.022

41. Yim KC, Cragin RP, Boukhvalova MS, Blanco JCG, Hamlin M-È, Boivin G, et al. Human metapneumovirus: enhanced pulmonary disease in cotton rats immunized with formalin-inactivated virus vaccine and challenged. *Vaccine* (2007) 25:5034–40. doi: 10.1016/j.vaccine.2007.04.075

42. McLellan JS, Chen M, Joyce MG, Sastry M, Stewart-Jones GBE, Yang Y, et al. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* (2013) 342:592–8. doi: 10.1126/science.1243283

43. Joyce MG, Bao A, Chen M, Georgiev IS, Ou L, Bylund T, et al. Crystal structure and immunogenicity of the DS-Cav1-stabilized fusion glycoprotein from respiratory syncytial virus subtype b. *Pathog Immun* (2019) 4:294. doi: 10.20411/pai.v4i2.338

44. Walsh EE, Falsey AR, Scott DA, Gurtman A, Zareba AM, Jansen KU, et al. A randomized phase 1/2 study of a respiratory syncytial virus prefusion f vaccine. *J Infect Dis* (2022) 225:1357–66. doi: 10.1093/infdis/jiab612

45. Langley JM, Aggarwal N, Toma A, Halperin SA, McNeil SA, Fissette L, et al. A randomized, controlled, observer-blinded phase 1 study of the safety and immunogenicity of a respiratory syncytial virus vaccine with or without alum adjuvant. *J Infect Dis* (2017) 215:24–33. doi: 10.1093/infdis/jiw453

46. Jiachen H, Pradeep C, Lin L, Tamas N, Jackelyn M, Tripp RA, et al. Structure, immunogenicity, and conformation-dependent receptor binding of the postfusion human metapneumovirus f protein. *J Virol* (2021) 95:e00593–21. doi: 10.1128/JVI.00593-21

47. Boyington JC, Joyce MG, Sastry M, Stewart-Jones GBE, Chen M, Kong W-P, et al. Structure-based design of head-only fusion glycoprotein immunogens for respiratory syncytial virus. *PLoS One* (2016) 11:e0159709. doi: 10.1371/journal.pone.0159709

48. Swanson KA, Rainho-Tomko JN, Williams ZP, Lanza L, Peredelchuk M, Kishko M, et al. A respiratory syncytial virus (RSV) f protein nanoparticle vaccine focuses antibody responses to a conserved neutralization domain. *Sci Immunol* (2020) 5:eaba6466. doi: 10.1126/sciimmunol.aba6466

49. Sesterhenn F, Galloux M, Vollers SS, Csepregi L, Yang C, Descamps D, et al. Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PloS Biol* (2019) 17:e3000164. doi: 10.1371/journal.pbio.3000164

50. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, et al. Proof of principle for epitope-focused vaccine design. *Nature* (2014) 507:201–6. doi: 10.1038/nature12966

51. Sesterhenn F, Yang C, Bonet J, Cramer JT, Wen X, Wang Y, et al. De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science* (2020) 368(6492):eaay5051. doi: 10.1126/science.aay5051

52. Wen X, Pickens J, Mousa JJ, Leser GP, Lamb RA, Crowe JE, et al. A chimeric pneumovirus fusion protein carrying neutralizing epitopes of both MPV and RSV. *PLoS One* (2016) 11:e0155917. doi: 10.1371/journal.pone.0155917

53. Olmedillas E, Cano O, Martínez I, Luque D, Terrón MC, McLellan JS, et al. Chimeric pneumoviridae fusion proteins as immunogens to induce cross-neutralizing antibody responses. *EMBO Mol Med* (2018) 10:175–87. doi: 10.15252/emmm.201708078

54. Nachbagauer R, Kinzler D, Choi A, Hirsh A, Beaulieu E, Lecrenier N, et al. A chimeric haemagglutinin-based influenza split virion vaccine adjuvanted with AS03 induces protective stalk-reactive antibodies in mice. *NPJ Vaccines* (2016) 1:1–10. doi: 10.1038/npjvaccines.2016.15

55. Nachbagauer R, Feser J, Naficy A, Bernstein DI, Guptill J, Walter EB, et al. A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. *Nat Med* (2021) 27:106–14. doi: 10.1038/s41591-020-1118-7

56. Huang J, Diaz D, Mousa JJ. Antibody recognition of the pneumovirus fusion protein trimer interface. *PloS Pathog* (2020) 16:e1008942. doi: 10.1371/journal.ppat.1008942

57. Ecker JW, Kirchenbaum GA, Pierce SR, Skarlupka AL, Abreu RB, Cooper RE, et al. High-yield expression and purification of recombinant influenza virus proteins from stably-transfected mammalian cell lines. *Vaccines* (2020) 8:462. doi: 10.3390/vaccines8030462

58. Bar-Peled Y, Diaz D, Pena-Briseno A, Murray J, Huang J, Tripp RA, et al. A potent neutralizing site III-specific human antibody neutralizes human metapneumovirus *in vivo*. *J Virol* (2019) 93:e00342–19. doi: 10.1128/JVI.00342-19

59. McLellan JS, Chen M, Chang J-S, Yang Y, Kim A, Graham BS, et al. Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J Virol* (2010) 84:12236–44. doi: 10.1128/JVI.01579-10

60. 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. doi: 10.1002/pro.3943

61. McLellan JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, et al. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* (2013) 340:1113–7. doi: 10.1126/science.1234914

62. Mousa JJ, Kose N, Matta P, Gilchuk P, Crowe JE. A novel pre-fusion conformation-specific neutralizing epitope on the respiratory syncytial virus fusion protein. *Nat Microbiol* (2017) 2:16271. doi: 10.1038/nmicrobiol.2016.271

63. Wu H, Pfarr DS, Johnson S, Brewah YA, Woods RM, Patel NK, et al. Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *J Mol Biol* (2007) 368:652–65. doi: 10.1016/j.jmb.2007.02.024

64. Gilman MSA, Furmanova-Hollenstein P, Pascual G, van't Wout AB, Langedijk JPM, McLellan JS. Transient opening of trimeric prefusion RSV f proteins. *Nat Commun* (2019) 10:1–13. doi: 10.1038/s41467-019-09807-5

65. Corti D, Bianchi S, Vanzetta F, Minola A, Perez L, Agatic G, et al. Cross-neutralization of four paramyxoviruses by a human monoclonal antibody. *Nature* (2013) 501:439–43. doi: 10.1038/nature12442

66. Banerjee A, Huang J, Rush SA, Murray J, Gingerich AD, Royer F, et al. Structural basis for ultrapotent neutralization of human metapneumovirus. *Pro Nat Acad Sci* (2022) 119(25):484292. doi: 10.1073/pnas.2203326119

67. Firacative C, Gressler AE, Schubert K, Schulze B, Müller U, Brombacher F, et al. Identification of T helper (Th)1- and Th2-associated antigens of cryptococcus neoformans in a murine model of pulmonary infection. *Sci Rep* (2018) 8:2681. doi: 10.1038/s41598-018-21039-z

68. Watanabe H, Numata K, Ito T, Takagi K, Matsukawa A. Innate immune response in Th1- and Th2-dominant mouse strains. *Shock* (2004) 22:460–6. doi: 10.1097/01.shk.0000142249.08135.e9

69. Francis T. On the doctrine of original antigenic sin. *Proc Am Philos Soc* (1960) 104:572–8.

70. August A, Shaw CA, Lee H, Knightly C, Kalidindia S, Chu L, et al. Safety and immunogenicity of an mRNA-based human metapneumovirus and parainfluenza virus type 3 combined vaccine in healthy adults. *Open Forum Infect Dis* (2022) 9: ofac206. doi: 10.1093/ofid/ofac206

71. Ruckwardt TJ, Morabito KM, Phung E, Crank MC, Costner PJ, Holman LA, et al. Safety, tolerability, and immunogenicity of the respiratory syncytial virus prefusion f subunit vaccine DS-Cav1: A phase 1, randomised, open-label, dose-escalatio clinical trial. *Lancet Respir Med* (2021) 9:1111–20. doi: 10.1016/S2213-2600(21)00098-9

72. Fries L, Shinde V, Stoddard JJ, Thomas DN, Kpamegan E, Lu H, et al. Immunogenicity and safety of a respiratory syncytial virus fusion protein (RSV f) nanoparticle vaccine in older adults. *Immun Ageing* (2017) 14:8. doi: 10.1186/s12979-017-0090-7

73. Stokes AH, Franklin K, Fisher DE, Posobiec LM, Binazon O, Tripathi N, et al. Repeated dose toxicity study and developmental and reproductive toxicology studies of a respiratory syncytial virus candidate vaccine in rabbits and rats. *Int J Toxicol* (2021) 40:125–42. doi: 10.1177/1091581820985782

74. Stewart-Jones GBE, Gorman J, Ou L, Zhang B, Joyce MG, Yang L, et al. Interprotomer disulfide-stabilized variants of the human metapneumovirus fusion glycoprotein induce high titer-neutralizing responses. *Proc Natl Acad Sci* (2021) 18 (39):e2106196118. doi: 10.1073/pnas.2106196118

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fimmu.2022.963023

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Abbas Alibakhshi,
Shahid Beheshti University of Medical Sciences, Iran
Kenneth Lundstrom,
PanTherapeutics, Switzerland

*CORRESPONDENCE
Keith J. Chappell
k.chappell@uq.edu.au
Daniel Watterson
d.watterson@uq.edu.au
Andrew Young
a.young4@uq.edu.au

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

RECEIVED 07 June 2022
ACCEPTED 25 July 2022
PUBLISHED 18 August 2022

CITATION
Young A, Isaacs A, Scott CAP, Modhiran N, McMillan CLD, Cheung STM, Barr J, Marsh G, Thakur N, Bailey D, Li KSM, Luk HKH, Kok K-H, Lau SKP, Woo PCY, Furuyama W, Marzi A, Young PR, Chappell KJ and Watterson D (2022) A platform technology for generating subunit vaccines against diverse viral pathogens.
Front. Immunol. 13:963023.
doi: 10.3389/fimmu.2022.963023

A platform technology for generating subunit vaccines against diverse viral pathogens

Andrew Young[1,2*†], Ariel Isaacs[1†], Connor A. P. Scott[1†], Naphak Modhiran[1,2], Christopher L. D. McMillan[1], Stacey T. M. Cheung[1], Jennifer Barr[3], Glenn Marsh[3], Nazia Thakur[4,5], Dalan Bailey[4], Kenneth S. M. Li[6], Hayes K. H. Luk[6], Kin-Hang Kok[6], Susanna K. P. Lau[6], Patrick C. Y. Woo[6], Wakako Furuyama[7], Andrea Marzi[7], Paul R. Young[1,2,8], Keith J. Chappell[1,2,8*] and Daniel Watterson[1,2,8*]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia, [2]The Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, QLD, Australia, [3]CSIRO, Health and Biosecurity, Australian Centre for Disease Preparedness, Geelong, VIC, Australia, [4]The Pirbright Institute, Woking, United Kingdom, [5]Oxford Vaccine Group, Department of Paediatrics, Medical Sciences Division, University of Oxford, Oxford, United Kingdom, [6]Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [7]Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States, [8]Australian Infectious Disease Research Centre, The University of Queensland, Brisbane, QLD, Australia

The COVID-19 pandemic response has shown how vaccine platform technologies can be used to rapidly and effectively counteract a novel emerging infectious disease. The speed of development for mRNA and vector-based vaccines outpaced those of subunit vaccines, however, subunit vaccines can offer advantages in terms of safety and stability. Here we describe a subunit vaccine platform technology, the molecular clamp, in application to four viruses from divergent taxonomic families: Middle Eastern respiratory syndrome coronavirus (MERS-CoV), Ebola virus (EBOV), Lassa virus (LASV) and Nipah virus (NiV). The clamp streamlines subunit antigen production by both stabilising the immunologically important prefusion epitopes of trimeric viral fusion proteins while enabling purification without target-specific reagents by acting as an affinity tag. Conformations for each viral antigen were confirmed by monoclonal antibody binding, size exclusion chromatography and electron microscopy. Notably, all four antigens tested remained stable over four weeks of incubation at 40°C. Of the four vaccines tested, a neutralising immune response was stimulated by clamp stabilised MERS-CoV spike, EBOV glycoprotein and NiV fusion protein. Only the clamp stabilised LASV glycoprotein precursor failed to elicit virus neutralising antibodies. MERS-CoV and EBOV vaccine candidates were both tested in animal models and found to provide protection against viral challenge.

KEYWORDS

subunit, platform, viral, fusion, clamp, vaccine

1 Introduction

Despite significant advances in medical research the frequency of outbreaks of emerging infectious diseases (EIDs) is increasing (1). Within the past two decades there have been viral epidemics from severe acute respiratory syndrome coronavirus (SARS-CoV), Zika virus, influenza viruses H5N1 and H1N1, Middle Eastern respiratory syndrome coronavirus (MERS-CoV), Ebola virus (EBOV), Lassa virus (LASV) and Nipah virus (NiV) (1, 2). Currently, SARS-CoV-2 is causing widespread mortality and global disruption and has exposed broad deficiencies in preparative and counteractive measures to control EIDs (3). Demographic and environmental factors suggest that the frequency of outbreaks will continue, with a growing global population (particularly aging demographics), increasing population densities, international movement of people and goods and continued encroachment on the habitats of animal reservoirs (2).

Vaccination is central to infectious disease control, but conventional developmental pipelines are ill-suited to react quickly to novel threats, requiring more than 10 years on average from discovery to licensure (4). The response to SARS-CoV-2 has challenged this paradigm with over 76 vaccine candidates announced within 3 months of the publication of its genome (5) and the first vaccines authorised for emergency use by the World Health Organisation (WHO) within 1 year (6). These efforts highlighted the utility of platform technologies that allow the exchange of specific viral antigens or genes into scaffold vectors to streamline the vaccine discovery phase. Such platforms can circumvent the need to physically transfer viral isolates from outbreak epicentres, allowing vaccine candidates and characterisation reagents to be synthesised remotely using genetic sequence information alone. Protein subunit, nucleic acid (DNA and RNA), replicating and non-replicating vectors, and virus-like particle vaccines all demonstrated pre-clinical efficacy for SARS-CoV-2 within months of genome publication, demonstrating the potential to rapidly generate vaccines across a variety of platforms (7–12).

These candidate vaccines largely focused on the 'spike' fusion protein as the target immunogen. For enveloped viruses, fusion proteins are often the primary targets for neutralising humoral immunity due to their surface presentation and critical functional role during host cell entry (13). Fusion proteins are dynamic molecular structures that assume distinct conformations before and after merging the viral and host membranes. The 'prefusion' conformation represents the metastable surface presentation on the infectious virion. Upon engagement, these surface antigens drive membrane fusion by irreversibly rearranging into a lower free-energy 'postfusion' conformation. Immunity elicited to prefusion-specific epitopes is often more protective than immunity to the postfusion form, and monoclonal antibodies (mAbs) with potent virus-neutralising activity that specifically bind prefusion epitopes have been well-described (14–20). Thus, stabilising the prefusion conformation is a common approach to enhance the efficacy of fusion protein vaccines (21).

Of the vaccine technologies mentioned above, subunit vaccines are particularly amenable to molecular characterisation and rational design (22). Recombinant viral fusion proteins can be modified to stabilise their conformation and then screened for the presentation of key protective and prefusion-specific epitopes. Cytotoxic and poorly immunogenic subdomains can be modified or omitted, and by only containing viral fragments without replicative components, the safety profiles of subunit vaccines are typically more predictable than other vaccine classes. Subunit vaccine production is scalable, and antigens can be designed and formulated for enhanced thermostability, reducing deployment costs to increase vaccine affordability and access (22). However, recombinant proteins can be poorly immunogenic and often require two or more doses with co-stimulatory adjuvants to elicit protective antibody and T-cell responses. Further, solubilising or otherwise modifying recombinant antigens can destabilise neutralising epitopes, requiring the introduction of stabilising mutations, multimeric scaffolds and/or heterologous motifs. Rational antigen design strategies have had success for clinically significant viruses such as RSV and HIV (14, 23) but these approaches are often dependent on detailed structural information, rendering them unsuitable for generic application and rapid mobilisation for novel EIDs.

Here we report a heterologous glycoprotein motif called the 'molecular clamp' which facilitates stabilisation and purification of viral fusion proteins. This technology has previously been described in a proof-of-concept production study using Achimota paramyxovirus and Wenzhou mammarenavirus (24), in application to a range of influenza hemagglutinin subtypes (25) and more recently, in a SARS-CoV-2 subunit vaccine that was effective in Phase 1 human clinical trials (9, 26). We build on those studies by presenting the streamlined generation of subunit vaccine candidates to four taxonomically diverse, WHO-listed priority pathogens: MERS-CoV (*Coronaviridae*), EBOV (*Filoviridae*), LASV (*Arenaviradae*) and NiV (*Paramyxoviridae*). We show that the clamped fusion antigens trimerise efficiently and present prefusion epitopes. We also show that vaccination with three of these antigens elicits neutralising immunity *in vitro* from BALB/c immunised sera. The vaccines also protect hamster and mouse models from viral challenge with mouse adapted-EBOV (MA-EBOV) and MA-MERS-CoV respectively. Importantly, it is shown that antigen of high purity can be obtained using the clamp as an immunoaffinity purification tag in a single step, enabling the generation of prefusion-stabilised subunit vaccine candidates without the use of target-specific reagents, demonstrating this technology's potential for broader application and to novel emerging viruses.

2 Materials and methods

2.1 DNA vectors and protein expression

Amino acid sequences for NiV Malaysia strain fusion protein F (amino acids 1-483, GenBank ID: NP_112026.1), LASV Josiah strain glycoprotein precursor GPC (amino acids 1-427, GenBank ID: AAA46286.1), EBOV Mayinga 1976 strain glycoprotein GPΔmucin-like domain (MLD) (amino acids 1-303; 471-639, GenBank ID: AF086833.2) and MERS-CoV KFU-HKU-13 strain spike protein S (amino acids 1-1297, GenBank ID: AHX00711.1) were retrieved. The clamp domain was produced by connecting partial sequences encoding the gp41 subdomain of the HIV Env protein (amino acids 547-582, 625-662, GenBank ID: AAB50262.1) to the GP ectodomain, connected by a flexible linker (EBOV GPΔMLD: G2SG2; NiV F: GSG, MERS-CoV S: GSG; LASV GPC: G3SG3). Genetic constructs were codon optimised for expression in CHO-S cells (*Cricetulus griseus*) before nucleic acids were synthesised and cloned into a mammalian expression vector (pNBF; National Biologics Facility, Brisbane, Queensland, Australia) by inFusion cloning (ClonTech).

Plasmid DNA was extracted from bacterial cultures using PureYield DNA MidiPrep kit (Promega) and transfected into expiCHO cells using expiCHO expression kit (ThermoFisher). Suspension cultures were harvested 7 days post-transfection, before the supernatants were clarified by centrifugation at 5000 × g and then sterilised by filtration through 0.22 µm filters. Clamped proteins were purified from supernatant by immunoaffinity chromatography on a HiTrap NHS-activated column (GE Healthcare) conjugated with a molecular clamp-specific monoclonal antibody (mAb) (HIV1281; PDB 3P30) (17) using 400 mM NaCl PBS wash buffer and diethylamine elution buffer (5 mM EDTA, 100 mM Tris, 400 mM NaCl, 20 mM diethylamine, pH 11.5). Column elution fractions were neutralised with a 1:1 (v/v) ratio of 1 M Tris pH 6.8, concentrated, and buffer exchanged to PBS. Protein concentration was quantified using Nanodrop spectrophotometry.

Expression yield from CHO-S transient suspension culture was determined from 30 ml cultures performed in triplicate. Proteins were quantitated from absorbance spectra at 280 nm using a NanoDrop spectrophotometer (ThermoFisher), calibrated with extinction coefficients calculated from the full construct amino acid sequences using the Expasy ProtParam online tool (accessible from: https://web.expasy.org/protparam/).

Unstabilised MERS-CoV S was purchased from SinoBiological (CAT: 40069-V08B). The S protein originates from EMC/2012 strain (GenBank ID: AFS88936.1), consists of amino acids 1-1297 and is His-tagged. Control antigens for NiV (NiV FΔFP) and LASV (GPCysR4) have been previously described (18, 27).

2.2 *In vitro* protein characterisation

The purity and molecular weight of the proteins was assessed by loading 4 µg of denatured protein on a 4-12% SDS-PAGE (Biorad) under reducing conditions (100 mM dithiothreitol). Gels were stained in Coomassie brilliant blue R-250 for 1 hour and destained in 35% methanol and 10% acetic acid.

The oligomeric state of purified proteins was determined by size exclusion chromatography (SEC) using 50 µg of protein in a 300 µL loop connected to a Superose 6 Increase 10/300 GL (GE Healthcare Life Sciences) gel filtration column calibrated with a series of standards. Fractions of 1 mL were collected on the basis of elution volumes with peak absorbance values for subsequent analyses.

2.3 Antigen-specific monoclonal antibody binding

Protein antigenicity was assessed by enzyme-linked immunosorbent assay (ELISA) with prefusion-specific mAbs. For MERS-CoV S, receptor binding domain (RBD) mAbs 4C2 (28), m336 (29), LCA60 (30), MERS27 (31), MCA1 (32), JC57-14, CDC2-C2 (33) and D12 were used as well as a S2-specific mAb G4 (34). For LASV GPC, 37.7H (27), 12.1F and 25.10C (35) mAbs were analysed. For EBOV GP, h15758, h15765, h15959, h15960, h16042 (36), KZ52 (37), 2G4, 4G7, 1H3 (38), mab100, mab114 (39), and 13C6 (40) were used. For NiV F, prefusion-specific mAbs 5B3 (18) and mAb66 (41) were used. Briefly, 2 µg/mL of antigen in PBS was coated on a Nunc MaxiSorp 96-well plate (ThermoFisher) and incubated overnight at 4°C. Antigen was removed and plates were blocked with 150 µL per well of PBS with 0.05% Tween-20 supplemented with 5% milk diluent (blocking buffer) (KPL SeraCare) for 30 minutes at room temperature. mAbs were added to the plates pre-diluted in blocking buffer to 10 µg/mL and titrated 5-fold prior to incubation at 37°C for 1 hour. Plates were washed by water immersion thrice before addition of horseradish peroxidase (HRP)-conjugated secondary antibody (goat anti-human IgG) (ThermoFisher) diluted to 1:2000 in blocking buffer. Plates were incubated as before and washing was repeated. For signal generation, tetramethylbenzidine (TMB) (ThermoFisher) was added for 5-10 minutes at room temperature. Reaction was stopped with 1 M H_2SO_4 and optical density (OD) was measured at 450 nm. Background binding of mAbs against PBS was subtracted from binding to the respective antigen.

2.4 Transmission electron microscopy

SEC-purified antigens were deposited onto carbon-coated, glow-discharged 400 mesh copper grids (ProSciTech) at approximately 5 – 10 µg per ml. Samples were blotted off the grids and washed twice with water before staining with 1% (w/v) uranyl acetate for 2 mins. Grids were imaged using a Hitachi HT7700 Transmission Electron Microscope at 120 KeV and images were acquired using AXT 2kx2k CMOS. Subsequent micrograph processing was conducted using Relion 3.1 software, and contrast transfer functions of the images were corrected using CTFFIND. Particles were selected manually followed by reference-free alignment and two-dimensional classification.

Cryogenic electron microscopy (cryo-EM) was performed on NiV Fclamp proteins in complex with Fab fragments from 5B3, similar to a previous report (42). NiV Fclamp was mixed with purified Fab fragments at a w/w ratio of 1:2 respectively, and then incubated for 1 hour at 4°C. Samples were then loaded on to the Superose 6 Increase 10/300 GL column and size-exclusion was conducted as previously described. Fractions containing complexed antigens were identified by relative peak shifts and isolated, before being concentrated in 0.5 ml 3K MWCO concentrators. NiV Fclamp and 5B3 Fab complexes were diluted to 0.05 mg/ml and 4 µl was adsorbed onto flow discharged quantifoil grids (Q2/1). Grids were plunged frozen using EMGP2 system (Leica) and imaged on a Cryo-ARM 300 (JEOL) Field Emission Cryo-Electron Microscope equipped with a K3 detector (Gatan) in a super-resolution CDS acquisition mode. Movies were acquired with a 5 second exposure in super resolution mode at 0.1 s/frame using JADAS software at a magnification of ×50,000. This yielded a pixel size of 0.48 and a dose rate of 7.66 e/pix/sec. Data was motion corrected using MotionCor2 (v1.1.10), and contrast transfer functions (CTF) of each micrograph were determined using CTFFIND software. Any micrographs with aberrations were removed as informed by thon-ring images, and only micrographs with a resolution equal to or less than 4 Å were selected for downstream analyses. Initial 2D references were formed by manually picking particles in Relion 3.1 to inform auto-picking software with a box size of 640 pixels. 3D refinement was conducted using C3 symmetry with a total of 40,833 particles.

2.5 Animal immunisations and viral challenge studies

For the immunogenicity studies, seven-week-old BALB/c mice were housed in HEPA-filtered cages at the Australian Institute for Bioengineering and Nanotechnology (AIBN) Animal Facility at The University of Queensland, Australia. Procedures were approved by The University of Queensland Animal Ethics Committee (AEC numbers: SCMB/354/14/AIBN/UNIQUEST; SCMB/558/17). Each mouse was immunised under anesthesia by ketamine and xylazine with 5 µg of antigen or PBS adjuvanted with either 3 µg of Quil-A intradermally or 50 µg of Alhydrogel per dose intramuscularly as stated. Immunisations were conducted at 21-day intervals.

2.5.1 EBOV challenge

All infectious work with MA-EBOV was performed in the high-containment laboratories at the Rocky Mountain Laboratories (RML), Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. RML is an institution accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). All procedures followed standard operating procedures (SOPs) approved by the RML Institutional Biosafety Committee (IBC). Animal work was performed in strict accordance with the recommendations described in the Guide for the Care and Use of Laboratory Animals of the National Institute of Health, the Office of Animal Welfare and the Animal Welfare Act, United States Department of Agriculture. The study was approved by the RML Animal Care and Use Committee (ACUC). Procedures were conducted in animals anesthetized (inhalational isoflurane) by trained personnel under the supervision of veterinary staff. All efforts were made to ameliorate animal welfare and minimize animal suffering; food and water were available *ad libitum*.

For the EBOV challenge study, 4 - 6-week-old Syrian golden hamsters (*Mesocricetus auratus*) were immunised with either 5 µg of subunit antigen or PBS adjuvanted with 3 µg of Quil-A twice by intramuscular (IM) injection, 21 days apart. Positive control hamsters were immunised once with 100,000 PFU of VSV-EBOV at the 2nd dose time point by the same route. Twenty-one days after the final immunisation, the hamsters were challenged with 1,000 LD_{50} (100 focus-forming units) of MA-EBOV by injection into the peritoneal cavity. Body weight changes were monitored daily until the hamsters recovered from the acute disease. Hamsters developing severe disease were euthanized following ACUC-approved endpoint criteria. Blood samples were collected by retro-orbital bleeding at the time of boost for subunit immunisations or the time of primary immunisation with VSV-EBOV (21 days pre-challenge) and 6 days prior to challenge. Terminal bleeds were collected *via* cardiac puncture at study endpoint (42 days post challenge (n = 6) or 4 days post challenge for tissue/blood sampling groups (n = 4)). At 4 days post MA-EBOV challenge, 4 hamsters per group were anesthetised and bled *via* cardiac puncture for virological analysis. Spleen and liver samples were collected and stored at -80°C. Virus loads were determined in hamster blood and tissue samples as previously described (43). Briefly, Vero E6 cells were seeded in 48-well plates the day before titration. Tissues were weighed, homogenised in 1 mL serum-free DMEM and tissue and blood samples were serially diluted 10-fold. Media was removed from cells and inoculated with

each dilution in triplicate. After one hour, DMEM supplemented with 2% FBS, penicillin/streptomycin and L-glutamine was added and incubated at 37°C. Cells were monitored for cytopathic effect (CPE) and 50% tissue culture infectious dose ($TCID_{50}$) was calculated for each sample (blood, per mL; tissue per mg) employing the Reed and Muench method (44).

2.5.2 MERS-CoV challenge

The MERS-CoV mouse challenge experiment was approved by CULATR, HKU (CULATR 5067-19) and the Department of Health, the Government of the HKSAR under the Animals (Control of Experiments) Ordinance, Chapter 340 (19-384/385 in DH/SHS/8/2/3 Pt.32). The mouse study was carried out in strict compliance with animal welfare regulations. The mice were anesthetised by ketamine/xylazine when procedures were conducted. Standard guidelines prescribed in pain and distress in laboratory rodents and lagomorphs, Laboratory Animals 28, 97-112 (1994) were strictly followed and the well-being of animals were monitored daily with a scoring sheet to ensure minimal pain and distress experienced by the mice.

A congenic C57BL/6 mouse with mouse *DPP4* exons 10-12 replaced with the human *DPP4* codons was generated by Taconic Biosciences and provided by Paul McCray, University of Iowa to HKU (45). Human *DPP4* knockin mice, 6-8 weeks-old, were vaccinated with 1 μg or 5 μg of MERS-CoV Sclamp vaccine with MF59 adjuvant (Seqirus) or with 5 μg of MERS-CoV Sclamp without adjuvant. A comparator group of mice received 1.25 μg of formalin inactivated MERS-CoV and another group received PBS as a placebo. A booster vaccination of the same material was given after 3 weeks. Three weeks after the booster vaccination, mice were challenged with 10^4 PFU of MA-MERS-CoV (clone 6.1.2) in 20 μl *via* intranasal route. Mice were weighed and monitored daily for 3 days. Five mice from each group were sacrificed on 3 days post infection with lungs harvested. Mice inoculated with MEM were included as mock infection group.

Lung tissues from infected mice were weighed and homogenised with TissueRuptor II (Qiagen, Hilden, Germany) in 1 ml PBS. Viral RNA was directly extracted from lung tissue homogenates of infected mice using RNeasy Mini Kit according to manufacturer's instructions (Qiagen, Hilden, Germany). The RNA was eluted in 50 μl of RNase-free water and quantified. 100 ng of total RNA were used as template for cDNA synthesis by reverse transcription using SuperScript III kit (Invitrogen, San Diego, CA, USA). RT-qPCR was performed to determine the viral load of the infected mice lung tissues using LightCycler 480 SYBR Green I Master (Roche Diagnostics). Briefly, 5 μl of cDNA was used in each reaction with specific primers and probes targeting the nucleocapsid (N) gene of MERS-CoV as previously described (46). The relative abundance between different experimental groups was calculated by the ΔΔCt method normalised to mouse GAPDH.

To examine the histopathology, the lung tissues were fixed in 4% paraformaldehyde, embedded in paraffin, and stained with hematoxylin and eosin (H&E). Histopathological changes were observed using a Nikon 80i microscope and imaging system.

2.6 Measuring antigen-specific IgG responses

Serum reactivity to cognate antigens was determined by ELISA as described in section 2.3. In brief, ELISA plates were coated with 2 μg/ml of antigen and incubated overnight at 4°C. The plates were then treated with blocking buffer (1× KPL blocking solution concentrate (SeraCare) in PBST). Sera were serially diluted in blocking buffer before being added to the coated plate and incubated for 1 hour at 37°C. Bound antibodies were detected using goat anti- mouse IgG or goat anti-hamster IgG HRP conjugates (ThermoFisher) as appropriate. Bound conjugates were detected with 3,3′,5,5′-Tetramethylbenzidine (TMB) (ThermoFisher) before the reactions were quenched with 1 M H_2SO_4. Absorbances were read at 450 nm. Endpoint titres (EPTs) were then determined by taking the mean absorbance of the mock (serum-free) plus 3 standard deviations and interpolating this value into a one-site specific binding regression model using GraphPad Prism.

2.7 Virus and pseudovirus propagation and neutralisation assays

EBOV (Mayinga 1976, GenBank ID: AF086833.2), MA-EBOV (passage 3) (47) and VSV-EBOV (expressing EBOV-Kikwit GP (48)) were propagated in Vero E6 cells. The supernatants were clarified by centrifugation at 1500 × g for 10 min, aliquoted and stored in liquid nitrogen (MA-EBOV) or at -80°C (VSV-EBOV). MA-EBOV titers were determined by immuno-plaque assay. VSV-EBOV was titred on Vero E6 cells using standard plaque assay (49). EBOV (Mayinga 1976) neutralisation was quantified using a standard plaque assay in Vero E6 cell culture using previously described methods (50).

Live virus neutralisation assays were conducted under strict bio-containment procedures in the BSL4 laboratory at ACDP. Serial two-fold dilutions of sera were prepared in 96-well tissue culture plates in 50 μl DMEM (Dulbecco's Modified Eagle's Medium, supplemented with 1% antibiotic/antimycotic and 1% hepes). An equal volume of either NiV, MERS-CoV or LASV containing 200 $TCID_{50}$ was added to each well and the virus-serum mix incubated for 45 min at 37°C in a humidified 5% CO_2 incubator. Vero E6 cell suspension was prepared in DMEM containing 10% fetal bovine serum (FBS), then 100 μl of cell suspension containing 2×10^5 cells/ml was added to each well and the plates incubated for three to five days. The wells were

observed for signs of viral CPE and the titre was determined as the serum concentration in which viral CPE was not observed.

Pseudovirus neutralisation assays were conducted using lentivirus-based pseudotypes as previously described (51–53). Briefly, HEK293T cells were transfected with p8.91 (encoding for HIV-1 gag-pol), CSFLW (lentivirus backbone expressing a firefly luciferase reporter gene) and viral glycoprotein (NiV F + G or LASV GPC) using PEI transfection reagent. Supernatants containing pseudotyped virus were harvested at 48- and 72-hours post-transfection, pooled and centrifuged at 1,300 x *g* for 10 minutes at 4°C to remove cellular debris. Pseudo-particles were then titrated on target HEK293T cells to obtain a luciferase virus titre. For the micro neutralisation tests (mVNTs), sera were diluted in serum-free media in triplicate from a starting dilution of 1:10 and titrated 3-fold. A fixed titred volume of pseudo-particles was added at a dilution equivalent to 10^5 - 10^6 signal luciferase units in 50 µL DMEM-10% and incubated with sera for 1 hour at 37°C, 5% CO_2. HEK293T cells were then added at a density of 2 $\times 10^4$ in 100 µL and incubated at 37°C, 5% CO_2 for 48 hours. Firefly luciferase activity was then measured with BrightGlo luciferase reagent and a Glomax-Multi$^+$ Detection System (Promega). Pseudotyped virus neutralisation titres were calculated by interpolating the point at which there was 50% reduction in luciferase activity, relative to untreated controls (neutralisation dose 50%, ND_{50}).

2.8 Statistical analysis

Statistical analyses were performed using Prism 9 (GraphPad, San Diego, CA, USA). Yield differences were analysed with an unpaired, two-tailed t test. ELISA, virus and pseudovirus titres were analysed as log-transformations with ordinary one-way ANOVA and Tukey's multiple comparisons test. MERS-CoV Ct values, eosinophil and histology scores were analysed with 2way ANOVA with Tukey's multiple comparisons test. EBOV survival curves were analysed with a Mantel-Cox test. EBOV viral titres were analysed with ordinary one-way ANOVA with Tukey's multiple comparisons test. Statistical significance is indicated as: $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

3 Results

3.1 Antigen design, expression and characterisation

The molecular clamp domain is derived from the HIV gp41 fusion core protein and consists of a pair of alpha-helices connected by a flexible linker which facilitates the formation of a hairpin tertiary structure (Figure 1A). These paired helix hairpins enhance trimerisation by coalescing with high affinity into thermostable, six-helix bundles *via* a series of intermolecular interdigitations (55). Of the 8 'known' WHO-listed priority viral pathogens in need of vaccines and therapeutic interventions (one is a hypothetical 'Disease X'), five contain trimeric fusion proteins that are potentially amenable to clamp-stabilisation (Figure 1B) (54).

Synthetic DNA fragments encoding NiV F, LASV GPC, MERS-CoV S or EBOV GPΔMLD ectodomains were cloned into mammalian expression vectors containing the molecular clamp sequence appended in place of their respective transmembrane (TM) domain coding regions. These produced constructs encoding NiV Fclamp, LASV GPCclamp, MERS-CoV Sclamp and EBOV GPΔMLDclamp. The recombinant antigens were transiently expressed in mammalian suspension cultures (Chinese hamster ovary, CHO-S) before purification using immunoaffinity with a clamp-specific monoclonal antibody (mAb) [HIV1281 (17)]. Yields ranged from 8 mg/L (LASV GPCclamp) to 31.2 mg/L (EBOV GPΔMLDclamp) for the clamped antigens (Figure 1C).

We attempted to generate unstabilised, ectodomain-only control antigens by similar methods. Purification using ectodomain-specific mAbs yielded EBOV GPΔMLD (purified with mAb KZ52 (37)), but control antigens were not recoverable for MERS-CoV S, NiV F or LASV GPC. Ultimately, to obtain unclamped control antigens for these viruses, best-in-class comparator proteins were used. These were LASV GPCysR4 (27), NiV FΔFP (18) and MERS-CoV S (MERS-CoV S1+S2 N-(AA1-1291)-His-C, SinoBiological, catalogue number: 40069-V08B). The yield of EBOV GPΔMLDclamp and GPΔMLD when purified with KZ52 showed incorporation of the clamp enhanced recoverable yield 2-fold for these antigens ($p = 0.0267$, Figure 1D). SDS-PAGE stained with Coomassie blue showed that clamp-mediated immunoaffinity chromatography recovered antigens of predicted molecular weight (MW) profiles and high purity (Figure 1E, inlayed).

Oligomerisation was assessed using size-exclusion fast protein liquid chromatography (SE-FPLC) with fractions assayed by ELISA using mAbs reactive to prefusion epitopes (Figure 1E). Sclamp (MERS-CoV) was predominantly trimeric (70% of total area under curve (AUC, mAU ml^{-1}), 11.5-15 ml retention volumes (RV)) with some aggregation (9% AUC, 6-9.5 ml RV) and degradation/monomeric dissociation observed (13% AUC, 15.4-28 ml RV). The unstabilised counterpart, S, was largely aggregated (48% AUC; 6-10 ml RV) and exhibited a long 'tail'-like profile (52% AUC; 10-30 ml RV), indicative of protein degradation. No local peak was observable corresponding to trimer within the S profile, and the receptor-binding domain (RBD)-specific mAb 4C2 (28) was unreactive to the S 11.5-15 ml RV fractions. However, 4C2 did bind to the stabilised Sclamp trimer and the S aggregated species.

EBOV GPΔMLDclamp produced high MW aggregates and the trimeric peak mAU maximum was reduced 2.13-fold relative to the unstabilised antigen (14.7 ml RV GPΔMLDclamp; 14.9 ml

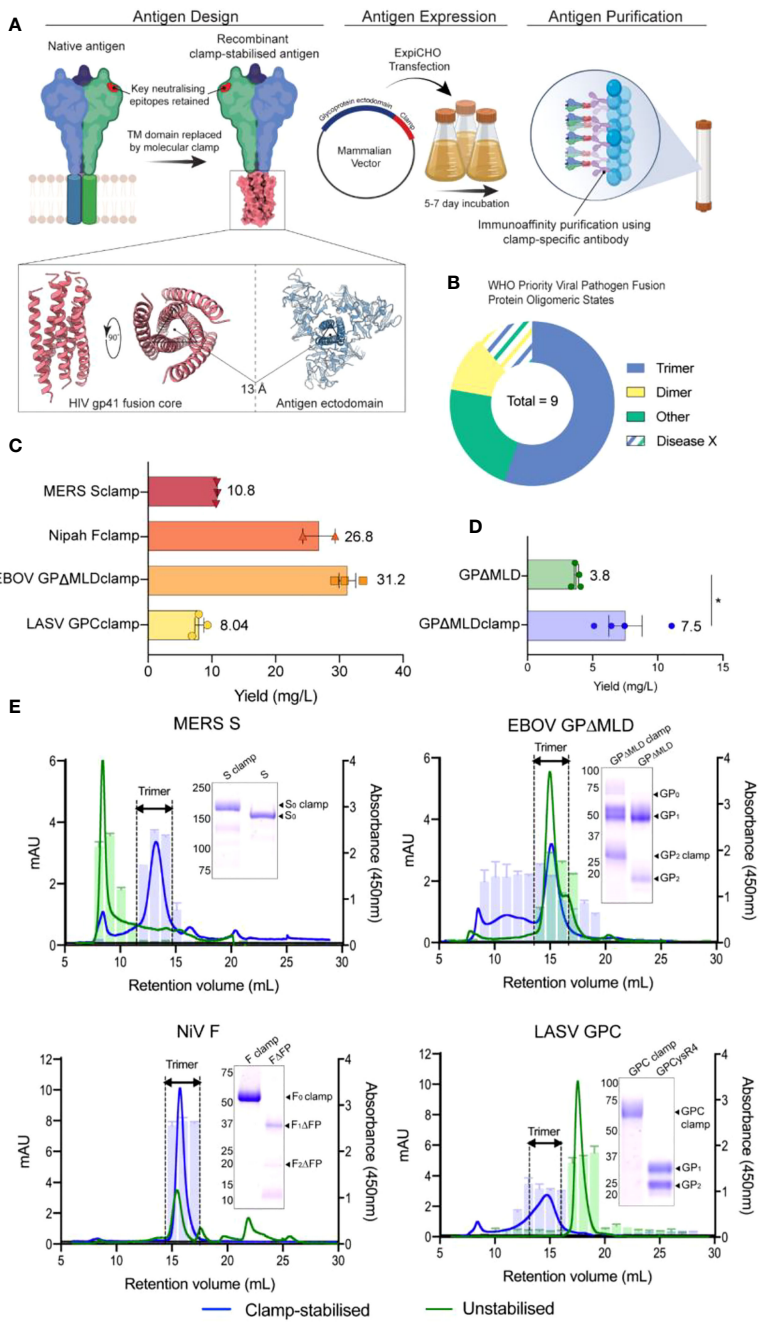


FIGURE 1
Design, generation and characterization of recombinant vaccine antigens. **(A)** The vaccine candidates were generated by genetically appending the clamp coding region in place of the native C-terminal transmembrane (TM) domain sequence. The constructs were cloned into a mammalian expression vector and expressed in ExpiCHO-S cells before the secreted antigens were purified using clamp-mediated immunoaffinity chromatography. **(B)** The WHO list's priority viral pathogens of greatest significance for public health to focus R&D efforts towards. As the molecular clamp is a trimerization domain, the proportion of listed viruses containing a trimeric fusion protein that the clamp may be applied to are shown in blue. Disease X represents a yet unknown pathogen which may or may not also utilize a trimeric fusion protein. For the full list see (54). **(C)** The yield of the recombinant antigens from transient CHO-S suspension culture and clamp-mediated immunoaffinity chromatography. Error bars show SD from multiple expression runs. **(D)** Comparison of the yield recovered from anti-ectodomain affinity purification using EBOV mAb KZ52 of the clamp-stabilised and unstabilsed GPΔMLD constructs. The asterisk (*) indicates statistical significance determined by an unpaired, two-tailed t test ($p = 0.0267$). **(E)** The oligomerization states of the clamp-stabilised (blue) and unstabilised (green) antigens are presented as size exclusion chromatography UV traces (mAU) with prefusion specific monoclonal antibody reactivity presented as bar graphs (absorbance 450nm). Antibodies used were KZ52 for EBOV GP, 37.7H for LASV GPC, 5B3 for NiV F and 4C2 for MERS-CoV S. Coomassie blue-stained SDS-PAGE gels of the purified antigens are inserted.

RV GPΔMLD) (Figure 1E). Trimer AUC comparisons between the EBOV constructs were not conducted as the GPΔMLD trimeric peak exhibited an overlapping, lower MW 'shoulder' consistent with monomeric dissociation (16.2 ml RV). However, aggregated protein accounted for the majority of the GPΔMLDclamp product (56% AUC, 0-13.5 ml RV) while only contributing 12% AUC to the total of GPΔMLD (0-13.5 ml RV).

NiV Fclamp exhibited a single major peak corresponding to trimeric oligomerisation (15.7 ml RV) and was reactive to mAb 5B3 (Figure 1E). The AUC of the FΔFP trimeric peak (40% AUC, 14-17 ml RV) was reduced 2.38-fold relative to Fclamp (94% AUC, 14-18.5 ml RV), with no mab 5B3 reactivity associated with the FΔFP trimer. Evidence of degradation and proteolysis was also observed in both SDS-PAGE and SE-FPLC profiles for the unstabilised FΔFP antigen.

Incorporation of the molecular clamp into GPCclamp (LASV) significantly enhanced trimerisation (64% AUC, 12.5-16.5 ml RV), with recombinant GPCysR4 presenting almost entirely as dissociated monomers (94% AUC, 16-19.5 ml RV). Both the monomeric GPCysR4 and the trimeric GPCclamp reacted with mAb 37.7H (Figure 1E).

Epitope presentation on the recombinant antigens was further examined by ELISA using broad panels of characterised mAbs targeted to key subdomains (Supplementary Table S1). The clamp-stabilised and control antigens typically bound the mAbs with comparable nanomolar apparent dissociation constants (kD). Notable discrepancies included MERS-CoV S, anti-'stem' domain antibody, G4 (34), which demonstrated a 10-fold improvement of binding to the clamp-stabilised MERS-CoV antigen. NiV F prefusion-specific mAbs 5B3 and mAb66 bound Fclamp but were unreactive to FΔFP, consistent with previous reports that the unstabilised F antigen assumes a postfusion conformation (18). Epitope stability after storage in PBS at 4°C, 25°C or 40°C was assessed to examine the thermostability of the antigens in the absence of formulation stabilisers (Figure S1). For each clamped antigen it was found that the mAbs maintained low nanomolar affinities under each condition, including 4 weeks of storage at 40°C; the most stringent condition tested.

Negative stain transmission electron microscopy (TEM) was conducted to further investigate the oligomeric state and conformation for size-excluded trimer fractions of each clamped antigen (Figure 2). For MERS-CoV Sclamp, TEM showed a homogenous preparation of prefusion stabilised trimeric S protein. This was further confirmed using single particle analysis (SPA) on Relion 3.1 software which showed 2D class averages displaying S clamp trimer with a three-fold symmetry and a short-tail domain which indicated either the stem region of the spike or the molecular clamp (Figure 2A). Both micrographs and SPA of trimeric fractions of EBOV GPΔMLDclamp showed small, oblong and heart-shaped particles, indicative of the GP trimer (Figure 2B). Micrographs and 2D classifications of LASV GPC showed small spherical ~10 nm particles representative of the small globular LASV GPC trimer (Figure 2C).

As a proof of concept, we imaged NiV Fclamp using cryo-EM to further validate the authenticity of the prefusion structure. NiV Fclamp proteins were complexed with prefusion specific mAb 5B3 Fab fragments and imaged by cryo-EM. From the resulting micrographs, a total of 40,833 particles were used for SPA of the NiV F clamp 5B3 complexes. Using the gold-standard Relion 3.1 software pipeline, a cryo-EM map of NiV Fclamp 5B3 complex was generated with a final resolution of 3.3 Å. Here we observed a three-fold symmetry, showing three Fab fragments bound to a trimeric NiV F structure (Figures 2D, E). The previously solved atomic model of NiV F complexed with 5B3 (PDB 6TYS) (42) fit the cryo-EM map with a high degree of similarity. Of note, several regions of the NiV Fclamp antigen were not resolved. Densities corresponding to the stem domain of NiV F and the clamp domain were not observed in the reconstructions (Figure 2E). It is likely that the stem and clamp regions of the antigen are hyperflexible and therefore result in a high degree of movement, leading to a heterogenous sample and poor resolution around these regions. Furthermore, we observed a preferential top-down orientation of particles across cryo-EM micrographs, obscuring the stem and the clamp domains, potentially contributing to the loss of densities around these regions.

3.2 Vaccine immunogenicity and serum neutralisation in rodent models

We next determined the antigen-specific reactivity of sera elicited to the recombinant vaccine candidates. For the NiV, LASV and MERS-CoV vaccines, BALB/c mice (n = 8) were immunised with 5 μg of recombinant antigen intradermally (ID), adjuvanted with 3 μg of Quil-A® (QA; Brenntag Biosector), or intramuscularly (IM), adjuvanted with 50 μg of Alhydrogel (Alum; Croda), three times at 21-day intervals. BALB/c mice (n = 5) were immunised with the EBOV antigens with the same dose and regimen, however only QA was tested as an adjuvant in these groups. All vaccinations were well tolerated. Twenty-one days after the final dose, blood was collected by cardiac puncture and sera reactivities to the cognate immunogens were assayed by ELISA (Figure 3A). The clamp-stabilised and unstabilised control antigens elicited broadly consistent IgG titres between immunogen pairs except in the case of LASV GPCclamp, which induced a significantly higher IgG titre than GPCysR4 under both adjuvant conditions (QA, p = <0.0001; Alum, p = <0.0001) (Figure 3A). There was a trend of QA-adjuvanted antigens eliciting higher IgG titres than Alum-adjuvanted equivalents, but this only reached statistical significance for LASV GPCysR4 (p = <0.0001) and MERS-CoV S (p = 0.004).

To differentiate fusion protein ectodomain reactivity from off-target immunity elicited to the molecular clamp, immunised serum was assayed against a clamp-stabilised influenza virus H3clamp antigen (Figure 3C). Similarly, the respective

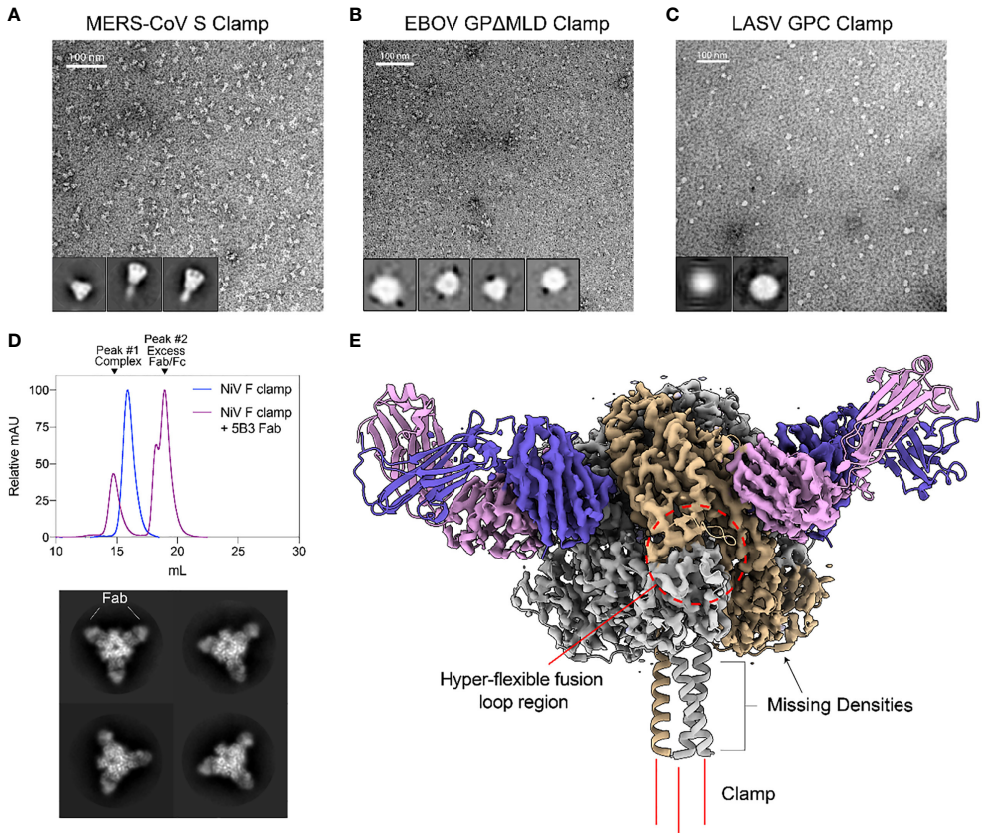


FIGURE 2
Structural validation and characterization of molecular clamped-stabilized antigens. **(A)** Representative negative stain TEM micrographs and 2D classes of MERS S Clamp, **(B)** EBOV GPΔMLD Clamp, and **(C)** LASV GPC Clamp. Clamp antigens were stained with 1% (w/v) uranyl acetate and imaged on a Hitachi HT7700 Transmission Electron Microscope at 120 kV. 2D classes were generated using Relion 3.1 software. **(D)** SEC of 5B3 Fab complexed or uncomplexed NiV F Clamp antigen separated on Superose 6 Increase 10/300GL column. Each data set is normalized to its maximal value. Below, representative 2D class averages of 5B3 complexed NiV F clamp. **(E)** 3.3 Å resolution cryo-EM structure of 5B3 complexed NiV F Clamp with the previously solved atomic structure (PDB 5EVM, PDB 6U1T) fitted in ribbon form. NiV F clamp is coloured in grey with a single monomer in gold. 5B3 heavy chains are coloured purple and light chains in pink. Unsolved clamp domain shown as three red lines at the C-terminus.

'unclamped' control antigens were assayed to determine viral ectodomain IgG titres for MERS-CoV, LASV and EBOV (Figure 3B). For NiV, both foldon stabilised F and the FΔFP control (18) were used to determine prefusion and postfusion ectodomain-specific reactivity, respectively. This analysis indicated that the clamp is immunogenic, but the majority of the IgG titres were elicited to viral ectodomains in all cases except for LASV GPCclamp (clamp reactivity: 96.92%, Alum-adjuvanted; 61.45%, QA-adjuvanted). For EBOV, NiV and MERS-CoV clamped antigens, clamp-specific reactivity ranged from 3.8% - 21.5%.

The neutralisation potency of the immunised sera was then determined using viral isolates in BSL4 biocontainment facilities (Figure 3D) and pseudoviruses expressing wild-type viral surface antigens for LASV and NiV (Figure S2). Immunisation with adjuvant-matched, clamped and control MERS-CoV antigens induced comparably potent virus neutralising responses relative to PBS controls (Sclamp, QA and Alum, p = <0.0001; S, QA and Alum, p = <0.0001), with QA inducing neutralisation to a greater degree than Alum for Sclamp (p = 0.0024). For NiV, immunisation with clamp-stabilised F elicited a significantly enhanced neutralising response relative to FΔFP (QA and Alum, p = <0.0001), consistent with reports that the majority of neutralising epitopes are presented on the prefusion antigen conformation (41, 42, 56). Stabilised EBOV GPΔMLDclamp elicited significantly higher levels of neutralising antibodies (nAbs) in comparison to the GPΔMLD control. For LASV, neither GPCclamp nor GPCysR4 were able to induce a significant neutralising response in both the viral and pseudoviral systems irrespective of the adjuvant used.

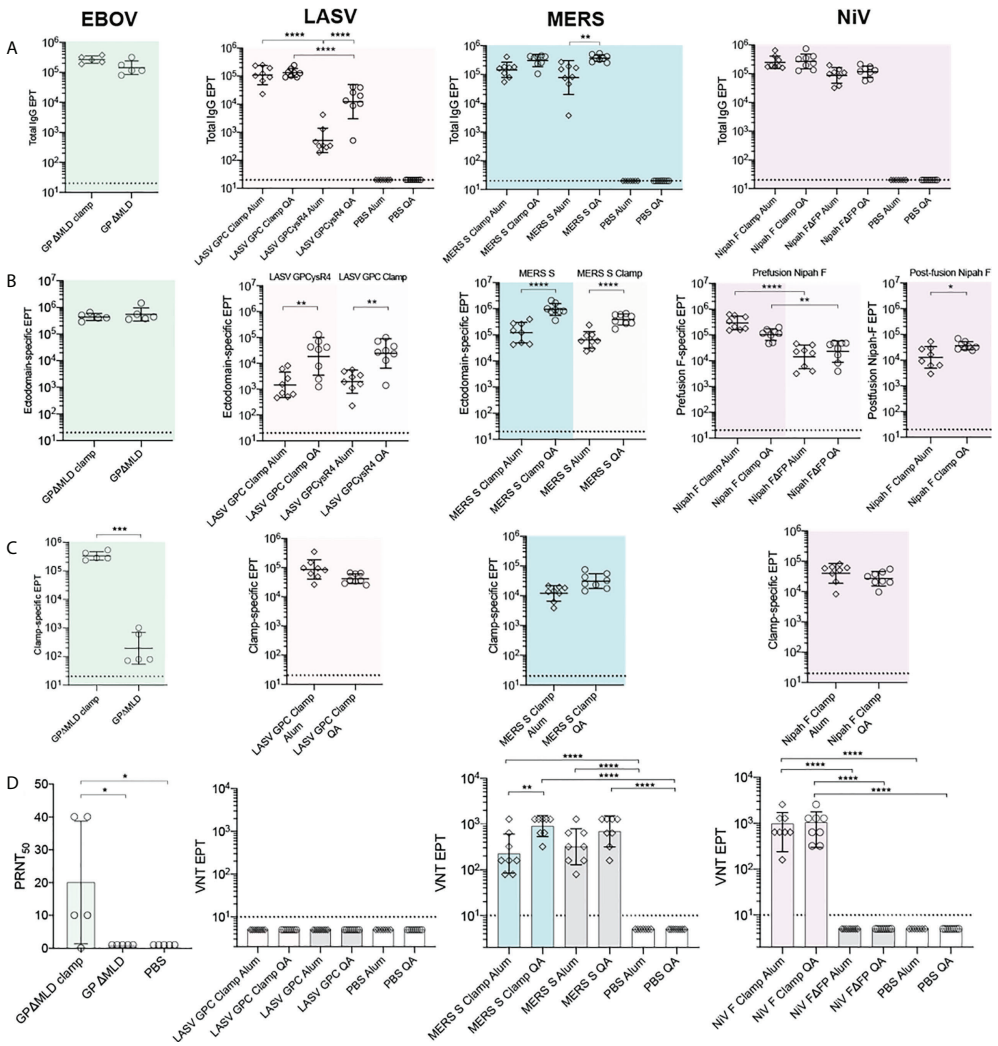


FIGURE 3
Immunogenicity and neutralization of virus. Serum reactivity in BALB/c immunised mice was assayed by ELISA and is presented as end-point titres (EPTs). **(A)** Sera groups were assayed against their cognate antigen (Total IgG), **(B)** an ectodomain only antigen (Ectodomain-specific) and **(C)** an influenza virus hemagglutinin-clamp antigen (Clamp-specific) (n = 5 for EBOV groups; n = 8 for LASV, MERS-CoV and NiV groups). **(D)** Neutralisation against live virus. Serum neutralization of EBOV was determined as the serum dilution factor required to neutralise 50% of the virus ($PRNT_{50}$). For LASV, MERS-CoV and NiV, virus neutralisation titre (VNT) was determined as the inverse serum concentration at which CPE was not observed. *P* values were determined by one-way ANOVA using Tukey's multiple comparison test. Statistical significance is indicated as: $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*). Error bars show SD.

3.3 Viral challenge models

For MERS-CoV and EBOV, the ability of the vaccine candidates to protect against lethal viral challenge was examined in well-characterised rodent models. The MERS-CoV study used *hDPP4* knock-in C57BL/6 mice (n = 5) (45), immunised twice, three weeks apart, with 1 or 5 μg of MERS-CoV Sclamp and 25 μl of squalene adjuvant MF59 (Seqirus). Infected and uninfected PBS placebo controls and an unadjuvanted 5 μg Sclamp cohort were also included. Finally, a formalin-inactivated MERS-CoV (1.25 μg) positive control was added, which has been shown to enhance lung disease in this model (45).

Three weeks after the final dose, mice were intranasally challenged with 1 x 10^4 pfu of MA-MERS-CoV. Three days post-challenge at the peak of infection the mice were sacrificed, and the lungs were assessed histologically and assayed for viral load (Figure 4A). Viral RNA in the lungs of mice immunised with 5 μg MERS-CoV Sclamp + MF59 was below the limit of detection (LOD = 40 mean Ct). This constituted a ≈65,000-fold reduction in viral load relative to 1 μg MERS-CoV Sclamp + MF59 (mean

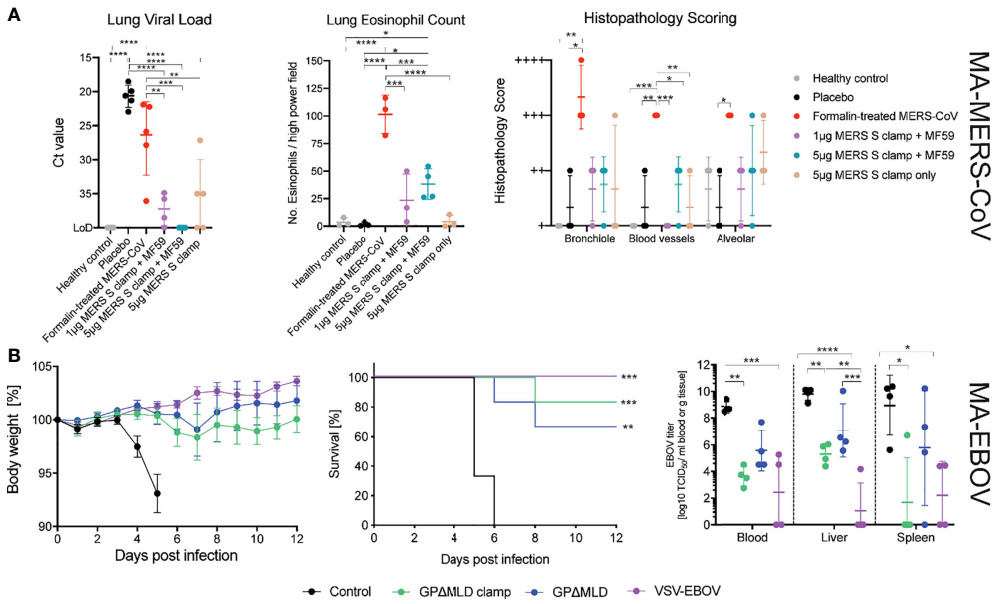


FIGURE 4
MERS-CoV and EBOV challenge models. **(A)** Congenic C57BL/6 DPP4 knock-in mice were given prime and boost immunisations before challenge with 10^4 PFU of MA-MERS. The tissues were then assessed for infiltration of the virus and histopathology. Asterisks indicate statistical significance calculated by a two-way ANOVA with Tukey's multiple comparisons test. **(B)** Syrian golden hamsters were immunised with clamped influenza virus H3 hemagluttinin (control), clamp-stabilised and unstabilised recombinant GPΔMLD or VSV-EBOV and challenged with 1,000 LD_{50} of MA-EBOV (n = 10). Body weights and survival of the hamsters was monitored following the challenge (n = 6). Four days post-challenge, hamsters were euthanized for virological analysis of the blood and tissues (n = 4). Statistics are calculated using a two-way ANOVA grouped analysis with Tukey's multiple comparisons test. Survival curves were analysed with a Mantel-Cox test. Statistical significance is indicated as: $p < 0.0001$ (****), $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

Ct = 37), and a ≈500,000-fold reduction relative to the placebo control (mean Ct = 21).

Disease severity was assessed by histopathological staining for eosinophil infiltration and scoring of lung pathology (Figure 4A). Mice were also assessed for weight loss on days 1-3 post infection however, no significant change in weight was observed for any of the groups (data not shown). The formalin inactivated MERS-CoV induced enhanced eosinophil infiltration into the lungs and increased the percentage of lung tissue affected. Eosinophil infiltration and the percentage of lung tissue affected were higher in MERS-CoV Sclamp/MF59 vaccinated animals compared to placebo. However, both measures of lung inflammation for Sclamp immunised mice were consistently lower than animals vaccinated with the formalin inactivated virus. For the 1 µg and 5 µg MERS-CoV Sclamp + MF59 the average number of eosinophils per high-power viewing field at 400× magnification in lung sections stained with Congo Red (Sigma-Aldrich) was 14 and 27, respectively, compared to 90 for the formalin treated virus. Additionally, the percentage of lung tissue affected was consistently lower for mice vaccinated with MERS-CoV Sclamp at either dose + MF59 compared to formalin treated virus.

We then determined the protective efficacy of vaccination with the recombinant EBOV subunit antigens in a lethal EBOV challenge model. Syrian golden hamsters were immunised twice at a 21-day interval with 5 µg of EBOV GPΔMLDclamp or GPΔMLD, adjuvanted with 3 µg of QA. Negative control animals were immunised with 5 µg of QA-adjuvanted influenza virus H3clamp, and positive control animals were immunised once with 100,000 PFU of VSV-EBOV (48, 57). Twenty-one days after the final immunisation, all hamsters were challenged with 1000 LD_{50} of MA-EBOV (Figure 4B).

Both the GPΔMLDclamp and GPΔMLD subunit vaccines were protective, with 5 of 6 and 4 of 6 hamsters surviving, respectively. Body weights remained stable throughout the observation window; never fluctuating by more than 10% for all groups except for the H3clamp control (Figure 4B). Virological analysis of the blood and tissues 4 days post-infection (n = 4) showed a significant reduction of viral titres from animals immunised with GPΔMLDclamp and VSV-EBOV relative to the control group (Figure 4B, viral titre). Viral clearance associated with GPΔMLD vaccination was not significant. Contrasting titres across all samples for animals immunised with either GPΔMLDclamp or VSV-EBOV revealed insignificant differences. Mean viral titres in the liver

and spleen were approximately 100-fold higher for animals immunised with GPΔMLDclamp than were observed in the VSV-EBOV cohort, although titres in the blood were 5-fold lower by this comparison. In the spleen, the disparity was attributable to one hamster in the GPΔMLDclamp group, with the remaining three fully clearing the virus from this organ. Three of 4 hamsters from the VSV-EBOV group showed no signs of virus in the liver, while all other animals presented substantial titres.

4 Discussion

The response to the SARS-CoV-2 pandemic has shown that with current technologies and sufficient resources, the time required to progress from vaccine design to human trials can be narrowed from years to months. Multiple vaccine modalities have produced effective candidates while offering different strengths and limitations, and future EID outbreak risk may be best mitigated by having vaccine development pathways established across a variety of platforms. Previously, we have shown the application of the molecular clamp in a SARS-CoV-2 spike subunit vaccine and demonstrated safety and efficacy in preclinical (9) and early clinical trials (26). While that study was ultimately discontinued due to interference with certain HIV diagnostics (58), it demonstrated that this platform technology can be rapidly and effectively mobilised in a pandemic. To resolve this issue, work is underway to explore modifying the current clamp and investigating the use of alternate motifs.

Using the clamp domain as an affinity tag, we showed that pure yields of the subunit antigens can be recovered without target-specific reagents (Figures 1C–E). Purification of recombinant proteins by immunoaffinity chromatography typically requires the introduction of affinity tags [such as C-myc (59) or FLAG tags (60)] or using antibodies specific to the antigen itself. The latter approach requires screening infected individuals or immunising animal models with the antigen of interest and identifying monoclonal antibodies that are amenable to antigen purification (61). This process can take months, extending the time required to generate subunit antigens in an emergency setting. However, the clamp domain's affinity tag function eliminates this step. The yields of the stabilised antigens from small-scale transient mammalian culture ranged from 8 - 31.2 mg L^{-1} (Figure 1C). We have previously demonstrated that the level of antigen production can be improved by roughly 50-fold by transitioning to stable clonal cell lines in industry-scale bioreactors. In that instance, upscaling production also improved product homogeneity (9). It is anticipated that clamp-mediated, industry-scale production is also compatible with the vaccine candidates described here.

The production of unclamped but otherwise equivalent control antigens was only successful for EBOV GPΔMLD, suggesting that in the absence of other stabilising modifications the clamp was necessary for expression of the MERS-CoV, NiV and LASV antigens. This was despite validation of anti-ectodomain purification using the clamped counterpart of each control antigen (EBOV, Figure 1D; MERS-CoV, NiV and LASV, data not shown). For EBOV, where direct clamped and unclamped comparisons were possible, antigen yield recoverable by anti-ectodomain immunoaffinity purification was roughly doubled when the clamp was present (Figure 1D). Importantly, including the clamp domain enhanced trimerisation for MERS-CoV, NiV and LASV (Figure 1E) to better reflect the native oligomeric state of the fusion antigens. For EBOV GPΔMLDclamp, aggregation was observed that reduced the overall proportion of trimer to below the corresponding control. However, while a greater proportion of the EBOV control antigen was trimeric, it also exhibited evidence of dissociating into monomers that was absent from GPΔMLDclamp. Aggregation is a common issue in recombinant protein expression and can occur for a variety of reasons including culture conditions, protein yield and intrinsic antigen instability (62). To circumvent the presence of aggregate for GPΔMLDclamp, an additional purification step can be introduced to isolate the trimeric fractions by gel filtration, for example. This is routinely performed to ensure the homogeneity of recombinant proteins and is a scalable process (62).

Remarkably, each of the four clamp-stabilised vaccine candidates were found to be stable at 40°C for at least 4 weeks in PBS without formulation stabilisers. (Figure S1). This suggests that optimised formulations of these vaccines have potential to be distributed without stringent temperature controls. Preserving the safety and efficacy of vaccine doses by maintaining unbroken cold chains during vaccine deployment can account for up to 80% of costs (63), so by reducing or eliminating dependence on cold chain distribution clamp-stabilised vaccines could significantly improve availability for at-risk populations. The impact of vaccine thermostability should not be understated, with vaccine access having been described as "the greatest challenge for protection of the human population against serious infectious disease" (64).

Robust murine IgG responses were elicited to the LASV, MERS-CoV, NiV and EBOV clamped subunit vaccines (Figure 3). This corresponded to significant viral neutralisation *in vitro* for three of the four vaccine candidates (MERS-CoV Sclamp, EBOV GPΔMLDclamp and NiV Fclamp). However, neutralisation was not observed for LASV GPCclamp immunised sera despite the antigen being shown to bind known nAbs 37.7H, 25.10C and 12.1F (Table S1). Notably, LASV GPCclamp was unprocessed at the internal GPC1-GPC2 cleavage site (Figure 1E, inlayed), which has been described as necessary for GPC to properly adopt its prefusion conformation and could affect immunogenicity despite the observed nAb reactivity (65). Another contributing factor may be putative shielding of the GPC protein surface with extensive

glycosylation (66). GPC shielding is consistent with the observation that LASV GPCclamp generated higher IgG titres to the clamp domain relative to the ectodomain, despite the clamp accounting for approximately 1/10th of the total antigen by molecular weight. Therapeutic antibodies and convalescent sera have shown variable efficacy for treating LASV disease which may indicate why the generation of protective humoral immunity from a LASV vaccine remains elusive despite considerable efforts (67). Cai *et al.* (2020) (68) showed that immunisation with a codon deoptimised, whole virus LASV vaccine conferred full protection in guinea pigs challenged with guinea pig-adapted LASV while only 1 of 16 animals produced detectable nAbs. In the LASV case, protection appears to be conferred by alternate immune mechanisms such as antibody-dependent cellular cytotoxicity (ADCC) (69). EBOV also shrouds epitopes on GP with glycans but is sensitive to nAbs at key immunogenic sites (36, 70, 71). For EBOV, total IgG titres have been proposed as a stronger correlate of protection than *in vitro* neutralisation or ADCC (72), which may indicate why protection from lethal viral challenge was observed (Figure 4B) despite weak virus neutralisation in plaque assays. Taken together, these results indicate that it may be worth persisting with this candidate LASV vaccine despite the absence of *in vitro* neutralisation.

While the lack of replicating elements and the absence of off-target pathogen associated molecular patterns confers greater safety profiles to subunit vaccines, adjuvants are typically necessary to achieve protective B and T cell responses (22). The LASV and MERS-CoV immunised sera showed significant enhancement of ectodomain-specific IgG titres for QA-adjuvanted doses relative to Alum for both control and clamped antigens, indicating that immunogenicity could be further augmented by exploring different adjuvants (Figure 3B). Ideally, a vaccine will induce both humoral and cellular immunity to generate stronger protection and exploit the differing susceptibilities of the viral targets (LASV, for example). While only IgG induction and neutralisation were assessed here, the previously mentioned SARS-CoV-2 Sclamp study showed that, in addition to robust humoral immunity, strong CD8 T-cell responses were elicited in mice when immunised with SARS-CoV-2 Sclamp adjuvanted with MF59 (9). However, previous research comparing a variety of adjuvants with different subunit vaccines has shown that the relationship between antigen, adjuvant and the induction of specific immunity can be unpredictable (73). This was illustrated by NiV Fclamp showing a trend of Alum enhancing ectodomain-specific immunity over QA, contrary to the observations for MERS-CoV and LASV (Figure 3B). Future studies should assess the effects of different adjuvants on the induction of alternate immune pathways with these vaccines.

The recent resounding success of mRNA vaccines for mitigating the impact of COVID-19 has launched mRNA vaccines into the wider consciousness, with many speculating that they may represent the future of vaccinology (74). The major advantages of the platform have been efficacy and the ability to generate vaccines with unprecedented speed (5, 8, 11). The current limitations of mRNA vaccines are their formulation stability and associated dependency on ultracold storage and transport (75) and uncertainty about long-term safety and immunological memory due to the nascence of the technology. Each of these limitations may yet be surmounted with continued investigation and development of the technology and its delivery vehicles. However, fusion antigens translated *in vivo* from mRNA vaccines still require that the correct antigen conformation be presented to the immune system. For example, both the Pfizer BNT162b2 and the Moderna mRNA-1273 vaccines contain codon substitutions that introduce prolines to constrain the prefusion conformation of the expressed vaccine antigen (76, 77). In both cases, the necessary structural understanding to rationally target the K986P and V987P substitutions was greatly accelerated by adapting established insights from prior structural investigation of the MERS-CoV spike protein (16). Assessing the application of the molecular clamp to mRNA vaccines presents another interesting line of enquiry and may offer a generic stabilisation strategy for *in vivo* mRNA expression of novel EID fusion antigens without prior characterisation of analogous proteins.

Here we have demonstrated the application of the molecular clamp platform technology to generate subunit vaccines for viruses from four divergent families. Conventional development of subunit vaccines can take longer than other classes of vaccine, offsetting the benefits of an established safety record and capacity for economical production and deployment. By exploiting the clamp domain's dual functions of stabilising immunologically important epitopes on trimeric fusion proteins while acting as an affinity tag, the subunit vaccine discovery phase can be shortened substantially, bringing development times in line with those of other platforms and offering an additional tool for emergency response.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was reviewed and approved by The University of Queensland Animal Ethics Committee, RML Animal Care and Use Committee (ACUC) and the Committee on the Use of Live Animals in Teaching and Research (CULATR), HKU.

Author contributions

Conceptualisation and invention of the clamp technology by PY, KC and DW. Clamped vaccine candidate and EBOV control antigen preparation, characterisation, immunisation and serum EPT studies by AY, AI, CS and CM. Additional animal handling and immunisation by SC. Electron microscopy by NM, AI and CS. Live virus neutralisation assays by JB and GM. MERS-CoV challenge studies by KL, HL, K-HK, SL and PW. EBOV challenge studies by WF and AM. Pseudovirus assays and analysis by AI, NT and DB. Writing-original draft AY. Writing-review and editing AI, CS, DW and KC. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by NHMRC Project grant APP1144025 and NHMRC Development grant APP1125107. This study was also funded in part by the Intramural Research Program, NIAID, NIH. NT and DB were funded by The Pirbright Institute's BBSRC institute strategic programme grant BBS/E/I/00007031, with NT receiving studentship support from BB/T008784/1.

Acknowledgments

Purified, recombinant NiV sF glycoprotein in its postfusion conformation (NiV FΔFP) was prepared and provided by Lianying Yan and Christopher Broder, Uniformed Services University, Bethesda, Maryland, USA. Recombinant LASV GPCysR4 control antigen was provided by Dr. Erica Ollmann Saphire, Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, CA, USA. MF59 adjuvant was provided by Seqirus Inc. Holly Springs, NC 27540 USA. Figure 1A was created using BioRender.com.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.963023/full#supplementary-material

References

1. Maslow JN. The cost and challenge of vaccine development for emerging and emergent infectious diseases. *Lancet Global Health* (2018) 6(12):e1266–e7. doi: 10.1016/S2214-109X(18)30418-2

2. Bloom DE, Black S, Rappuoli R. Emerging infectious diseases: A proactive approach. *Proc Natl Acad Sci U.S.A.* (2017) 114(16):4055–9. doi: 10.1073/pnas.1701410114

3. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* (2020) 395(10223):470–3. doi: 10.1016/S0140-6736(20)30185-9

4. Pronker ES, Weenen TC, Commandeur H, Claassen EHJHM, Osterhaus ADME. Risk in vaccine research and development quantified. *PLoS One* (2013) 8(3):e57755. doi: 10.1371/journal.pone.0057755

5. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The covid-19 vaccine development landscape. *Nat Rev Drug Discovery* (2020) 19(5):305–6. doi: 10.1038/d41573-020-00073-5

6. Organisation WH. Who issues its first emergency use validation for a covid-19 vaccine and emphasizes need for equitable global access who.int. In: *World health organisation* Geneva (2020). Available at: https://www.who.int/news/item/31-12-2020-who-issues-its-first-emergency-use-validation-for-a-covid-19-vaccine-and-emphasizes-need-for-equitable-global-access.

7. van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, et al. Chadox1 ncov-19 vaccine prevents sars-Cov-2 pneumonia in rhesus macaques. *Nature* (2020) 586(7830):578–82. doi: 10.1038/s41586-020-2608-y

8. Corbett KS, Flynn B, Foulds KE, Francica JR, Boyoglu-Barnum S, Werner AP, et al. Evaluation of the mrna-1273 vaccine against sars-Cov-2 in nonhuman primates. *N Engl J Med* (2020) 383(16):1544–55. doi: 10.1056/NEJMoa2024671

9. Watterson D, Wijesundara DK, Modhiran N, Mordant FL, Li Z, Avumegah MS, et al. Preclinical development of a molecular clamp-stabilised subunit vaccine for severe acute respiratory syndrome coronavirus 2. *Clin Transl Immunol* (2021) 10(4):e1269–e. doi: 10.1002/cti2.1269

10. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the Chadox1 ncov-19 vaccine against sars-Cov-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* (2020) 396(10249):467–78. doi: 10.1016/s0140-6736(20)31604-4

11. Walsh EE, Frenck RW, Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. Safety and immunogenicity of two rna-based covid-19 vaccine candidates. *New Engl J Med* (2020) 383(25):2439–50. doi: 10.1056/NEJMoa2027906

12. Tebas P, Yang S, Boyer JD, Reuschel EL, Patel A, Christensen-Quick A, et al. Safety and immunogenicity of ino-4800 DNA vaccine against sars-Cov-2: A preliminary report of an open-label, phase 1 clinical trial. *EClinicalMedicine* (2021) 31:100689. doi: 10.1016/j.eclinm.2020.100689

13. Rey FA, Lok SM. Common features of enveloped viruses and implications for immunogen design for next-generation vaccines. *Cell* (2018) 172(6):1319–34. doi: 10.1016/j.cell.2018.02.054

14. McLellan JS, Chen M, Joyce MG, Sastry M, Stewart-Jones GB, Yang Y, et al. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* (2013) 342(6158):592–8. doi: 10.1126/science.1243283

15. Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza a hemagglutinins. *Science* (2011) 333(6044):850–6. doi: 10.1126/science.1205669

16. Pallesen J, Wang N, Corbett KS, Wrapp D, Kirchdoerfer RN, Turner HL, et al. Immunogenicity and structures of a rationally designed prefusion mers-cov spike antigen. *Proc Natl Acad Sci United States America* (2017) 114(35):E7348–E57. doi: 10.1073/pnas.1707304114

17. Frey G, Chen J, Rits-Volloch S, Freeman MM, Zolla-Pazner S, Chen B. Distinct conformational states of hiv-1 Gp41 are recognized by neutralizing and non-neutralizing antibodies. *Nat Struct Mol Biol* (2010) 17(12):1486–91. doi: 10.1038/nsmb.1950

18. Chan YP, Lu M, Dutta S, Yan L, Barr J, Flora M, et al. Biochemical, conformational, and immunogenic analysis of soluble trimeric forms of henipavirus fusion glycoproteins. *J Virol* (2012) 86(21):11457–71. doi: 10.1128/jvi.01318-12

19. Mire CE, Chan YP, Borisevich V, Cross RW, Yan L, Agans KN, et al. A cross-reactive humanized monoclonal antibody targeting fusion glycoprotein function protects ferrets against lethal nipah virus and hendra virus infection. *J Infect Dis* (2020) 221(Suppl 4):S471–s9. doi: 10.1093/infdis/jiz515

20. Dang HV, Cross RW, Borisevich V, Bornholdt ZA, West BR, Chan YP, et al. Broadly neutralizing antibody cocktails targeting nipah virus and hendra virus fusion glycoproteins. *Nat Struct Mol Biol* (2021) 28(5):426–34. doi: 10.1038/s41594-021-00584-8

21. Graham BS, Gilman MSA, McLellan JS. Structure-based vaccine antigen design. *Annu Rev Med* (2019) 70:91–104. doi: 10.1146/annurev-med-121217-094234

22. Moyle PM, Toth I. Modern subunit vaccines: Development, components, and research opportunities. *ChemMedChem* (2013) 8(3):360–76. doi: 10.1002/cmdc.201200487

23. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, et al. Structure and immune recognition of trimeric pre-fusion hiv-1 env. *Nature* (2014) 514 (7523):455–61. doi: 10.1038/nature13808

24. Wijesundara DK, Avumegah MS, Lackenby J, Modhiran N, Isaacs A, Young PR, et al. Rapid response subunit vaccine design in the absence of structural information. *Front Immunol* (2020) 11:592370. doi: 10.3389/fimmu.2020.592370

25. McMillan CLD, Cheung STM, Modhiran N, Barnes J, Amarilla AA, Bielefeldt-Ohmann H, et al. Development of molecular clamp stabilized hemagglutinin vaccines for influenza a viruses. *NPJ Vaccines* (2021) 6(1):135. doi: 10.1038/s41541-021-00395-4

26. Chappell KJ, Mordant FL, Li Z, Wijesundara DK, Ellenberg P, Lackenby JA, et al. Safety and immunogenicity of an Mf59-adjuvanted spike glycoprotein-clamp vaccine for sars-Cov-2: A randomised, double-blind, placebo-controlled, phase 1 trial. *Lancet Infect Dis* (2021) 21(10):1383–94. doi: 10.1016/S1473-3099(21)00200-0

27. Hastie KM, Zandonatti MA, Kleinfelter LM, Heinrich ML, Rowland MM, Chandran K, et al. Structural basis for antibody-mediated neutralization of lassa virus. *Sci (New York NY)* (2017) 356(6341):923–8. doi: 10.1126/science.aam7260

28. Li Y, Wan Y, Liu P, Zhao J, Lu G, Qi J, et al. A humanized neutralizing antibody against mers-cov targeting the receptor-binding domain of the spike protein. *Cell Res* (2015) 25(11):1237–49. doi: 10.1038/cr.2015.113

29. Ying T, Prabakaran P, Du L, Shi W, Feng Y, Wang Y, et al. Junctional and allele-specific residues are critical for mers-cov neutralization by an exceptionally potent germline-like antibody. *Nat Commun* (2015) 6:8223. doi: 10.1038/ncomms9223

30. Walls AC, Xiong X, Park YJ, Tortorici MA, Snijder J, Quispe J, et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* (2019) 176(5):1026–39.e15. doi: 10.1016/j.cell.2018.12.028

31. Yu X, Zhang S, Jiang L, Cui Y, Li D, Wang D, et al. Structural basis for the neutralization of mers-cov by a human monoclonal antibody mers-27. *Sci Rep* (2015) 5:13133. doi: 10.1038/srep13133

32. Chen Z, Bao L, Chen C, Zou T, Xue Y, Li F, et al. Human neutralizing monoclonal antibody inhibition of middle East respiratory syndrome coronavirus replication in the common marmoset. *J Infect Dis* (2017) 215(12):1807–15. doi: 10.1093/infdis/jix209

33. Wang L, Shi W, Chappell JD, Joyce MG, Zhang Y, Kanekiyo M, et al. Importance of neutralizing monoclonal antibodies targeting multiple antigenic sites on the middle East respiratory syndrome coronavirus spike glycoprotein to avoid neutralization escape. *J Virol* (2018) 92(10):e02002–17. doi: 10.1128/jvi.02002-17

34. Wang L, Shi W, Joyce MG, Modjarrad K, Zhang Y, Leung K, et al. Evaluation of candidate vaccine approaches for mers-cov. *Nat Commun* (2015) 6 (1):7712. doi: 10.1038/ncomms8712

35. Robinson JE, Hastie KM, Cross RW, Yenni RE, Elliott DH, Rouelle JA, et al. Most neutralizing human monoclonal antibodies target novel epitopes requiring both lassa virus glycoprotein subunits. *Nat Commun* (2016) 7:11544. doi: 10.1038/ncomms11544

36. Bornholdt ZA, Turner HL, Murin CD, Li W, Sok D, Souders CA, et al. Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak. *Science* (2016) 351(6277):1078–83. doi: 10.1126/science.aad5788

37. Maruyama T, Rodriguez LL, Jahrling PB, Sanchez A, Khan AS, Nichol ST, et al. Ebola Virus can be effectively neutralized by antibody produced in natural human infection. *J Virol* (1999) 73(7):6024–30. doi: 10.1128/JVI.73.7.6024-6030.1999

38. Qiu X, Alimonti JB, Melito PL, Fernando L, Ströher U, Jones SM. Characterization of Zaire ebolavirus glycoprotein-specific monoclonal antibodies. *Clin Immunol* (2011) 141(2):218–27. doi: 10.1016/j.clim.2011.08.008

39. Corti D, Misasi J, Mulangu S, Stanley DA, Kanekiyo M, Wollen S, et al. Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. *Science* (2016) 351(6279):1339–42. doi: 10.1126/science.aad5224

40. Wilson JA, Hevey M, Bakken R, Guest S, Bray M, Schmaljohn AL, et al. Epitopes involved in antibody-mediated protection from Ebola virus. *Science* (2000) 287(5458):1664–6. doi: 10.1126/science.287.5458.1664

41. Avanzato VA, Oguntuyo KY, Escalera-Zamudio M, Gutierrez B, Golden M, Kosakovsky Pond SL, et al. A structural basis for antibody-mediated neutralization of nipah virus reveals a site of vulnerability at the fusion glycoprotein apex. *Proc Natl Acad Sci U.S.A.* (2019) 116(50):25057–67. doi: 10.1073/pnas.1912503116

42. Dang HV, Chan YP, Park YJ, Snijder J, Da Silva SC, Vu B, et al. An antibody against the f glycoprotein inhibits nipah and hendra virus infections. *Nat Struct Mol Biol* (2019) 26(10):980–7. doi: 10.1038/s41594-019-0308-9

43. Marzi A, Chadinah S, Haddock E, Feldmann F, Arndt N, Martellaro C, et al. Recently identified mutations in the Ebola virus-makona genome do not alter pathogenicity in animal models. *Cell Rep* (2018) 23(6):1806–16. doi: 10.1016/j.celrep.2018.04.027

44. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. *Am J Epidemiol* (1938) 27(3):493–7. doi: 10.1093/oxfordjournals.aje.a118408

45. Li K, Wohlford-Lenane CL, Channappanavar R, Park J-E, Earnest JT, Bair TB, et al. Mouse-adapted mers coronavirus causes lethal lung disease in human Dpp4 knockin mice. *Proc Natl Acad Sci* (2017) 114(15):E3119–E28. doi: 10.1073/pnas.1619109114

46. Li K, Wohlford-Lenane C, Perlman S, Zhao J, Jewell AK, Reznikov LR, et al. Middle East respiratory syndrome coronavirus causes multiple organ damage and lethal disease in mice transgenic for human dipeptidyl peptidase 4. *J Infect Dis* (2016) 213(5):712–22. doi: 10.1093/infdis/jiv499

47. Bray M, Davis K, Geisbert T, Schmaljohn C, Huggins J. A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J Infect Dis* (1998) 178(3):651–61. doi: 10.1086/515386

48. Marzi A, Robertson SJ, Haddock E, Feldmann F, Hanley PW, Scott DP, et al. Vsv-ebov rapidly protects macaques against infection with the 2014/15 Ebola virus outbreak strain. *Science* (2015) 349(6249):739–42. doi: 10.1126/science.aab3920

49. Baer A, Kehn-Hall K. Viral concentration determination through plaque assays: Using traditional and novel overlay systems. *J Vis Exp* (2014) 93):e52065. doi: 10.3791/52065

50. Pyankov OV, Setoh YX, Bodnev SA, Edmonds JH, Pyankova OG, Pyankov SA, et al. Successful post-exposure prophylaxis of Ebola infected non-human primates using Ebola glycoprotein-specific equine igg. *Sci Rep* (2017) 7:41537. doi: 10.1038/srep41537

51. Thakur N, Gallo G, Elreafey AME, Bailey D. Production of recombinant replication-defective lentiviruses bearing the sars-cov or sars-Cov-2 attachment spike glycoprotein and their application in receptor tropism and neutralisation assays. *Bio Protoc* (2021) 11(21):e4249. doi: 10.21769/BioProtoc.4249

52. Thakur N, Conceicao C, Isaacs A, Human S, Modhiran N, McLean RK, et al. Micro-fusion inhibition tests: Quantifying antibody neutralization of virus-mediated cell-cell fusion. *J Gen Virol* (2021) 102(1):jgv001506. doi: 10.1099/jgv.0.001506

53. Isaacs A, Cheung STM, Thakur N, Jaberolansar N, Young A, Modhiran N, et al. Combinatorial f-G immunogens as nipah and respiratory syncytial virus vaccine candidates. *Viruses* (2021) 13(10):1942. doi: 10.3390/v13101942

54. Organisation WH. Prioritizing diseases for research and development in emergency contexts who.int. In: *World health organisation*. Geneva (2020). Available at: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts#:~:text=At%20present%2C%20the%20priority%20diseases,disease%20and%20Marburg%20virus%20disease.

55. Chappell KJ, Watterson D, Young PR. *Chimeric molecules and uses thereof.* (International Bureau: World Intellectual Property Organization) (2018). Patent number: WO2018176103A1. Available at: https://patents.google.com/patent/WO2018176103A1/en

56. Loomis RJ, Stewart-Jones GBE, Tsybovsky Y, Caringal RT, Morabito KM, McLellan JS, et al. Structure-based design of nipah virus vaccines: A generalizable approach to paramyxovirus immunogen development. *Front Immunol* (2020) 11:842. doi: 10.3389/fimmu.2020.00842

57. Tsuda Y, Safronetz D, Brown K, LaCasse R, Marzi A, Ebihara H, et al. Protective efficacy of a bivalent recombinant vesicular stomatitis virus vaccine in the Syrian hamster model of lethal Ebola virus infection. *J Infect Dis* (2011) 204 Suppl 3(Suppl 3):S1090–S7. doi: 10.1093/infdis/jir379

58. Normile D. Development of unique Australian covid-19 vaccine halted. *Science* (2020), 80–2020. doi: 10.1126/science.abg1208

59. Evan GI, Lewis GK, Ramsay G, Bishop JM. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol* (1985) 5(12):3610–6. doi: 10.1128/mcb.5.12.3610-3616.1985

60. Hopp TP, Prickett KS, Price VL, Libby RT, March CJ, Pat Cerretti D, et al. A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* (1988) 6(10):1204–10. doi: 10.1038/nbt1088-1204

61. Moser AC, Hage DS. Immunoaffinity chromatography: An introduction to applications and recent developments. *Bioanalysis* (2010) 2(4):769–90. doi: 10.4155/bio.10.31

62. Cromwell ME, Hilario E, Jacobson F. Protein aggregation and bioprocessing. *AAPS J* (2006) 8(3):E572–9. doi: 10.1208/aapsj080366

63. Pelliccia M, Andreozzi P, Paulose J, D'Alicarnasso M, Cagno V, Donalisio M, et al. Additives for vaccine storage to improve thermal stability of adenoviruses from hours to months. *Nat Commun* (2016) 7(1):13520. doi: 10.1038/ncomms13520

64. Pollard AJ, Bijker EM. A guide to vaccinology: From basic principles to new developments. *Nat Rev Immunol* (2021) 21(2):83–100. doi: 10.1038/s41577-020-00479-7

65. Hastie KM, Saphire EO. Lassa virus glycoprotein: Stopping a moving target. *Curr Opin Virol* (2018) 31:52–8. doi: 10.1016/j.coviro.2018.05.002

66. Sommerstein R, Flatz L, Remy MM, Malinge P, Magistrelli G, Fischer N, et al. Arenavirus glycan shield promotes neutralizing antibody evasion and protracted infection. *PLoS Pathog* (2015) 11(11):e1005276. doi: 10.1371/journal.ppat.1005276

67. Salami K, Gouglas D, Schmaljohn C, Saville M, Tornieporth N. A review of lassa fever vaccine candidates. *Curr Opin Virol* (2019) 37:105–11. doi: 10.1016/j.coviro.2019.07.006

68. Cai Y, Ye C, Cheng B, Nogales A, Iwasaki M, Yu S, et al. A lassa fever live-attenuated vaccine based on codon deoptimization of the viral glycoprotein gene. *mBio* (2020) 11(1):e00039–20. doi: 10.1128/mBio.00039-20

69. Abreu-Mota T, Hagen KR, Cooper K, Jahrling PB, Tan G, Wirblich C, et al. Non-neutralizing antibodies elicited by recombinant lassa-rabies vaccine are critical for protection against lassa fever. *Nat Commun* (2018) 9(1):4223. doi: 10.1038/s41467-018-06741-w

70. Lee JE, Saphire EO. Neutralizing ebolavirus: Structural insights into the envelope glycoprotein and antibodies targeted against it. *Curr Opin Struct Biol* (2009) 19(4):408–17. doi: 10.1016/j.sbi.2009.05.004

71. Misasi J, Gilman MS, Kanekiyo M, Gui M, Cagigi A, Mulangu S, et al. Structural and molecular basis for Ebola virus neutralization by protective human antibodies. *Science* (2016) 351(6279):1343–6. doi: 10.1126/science.aad6117

72. Sullivan NJ, Martin JE, Graham BS, Nabel GJ. Correlates of protective immunity for Ebola vaccines: Implications for regulatory approval by the animal rule. *Nat Rev Microbiol* (2009) 7(5):393–400. doi: 10.1038/nrmicro2129

73. Isaacs A, Li Z, Cheung STM, Wijesundara DK, McMillan CLD, Modhiran N, et al. Adjuvant selection for influenza and rsv prefusion subunit vaccines. *Vaccines (Basel)* (2021) 9(2):71. doi: 10.3390/vaccines9020071

74. Extance A. *Mrna vaccines: Hope beneath the hype thebmj* (2021). Available at: https://www.bmj.com/content/375/bmj.n2744.

75. Crommelin DJA, Anchordoquy TJ, Volkin DB, Jiskoot W, Mastrobattista E. Addressing the cold reality of mrna vaccine stability. *J Pharm Sci* (2021) 110 (3):997–1001. doi: 10.1016/j.xphs.2020.12.006

76. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mrna vaccine against sars-Cov-2 - preliminary report. *N Engl J Med* (2020) 383(20):1920–31. doi: 10.1056/NEJMoa2022483

77. Vogel AB, Kanevsky I, Che Y, Swanson KA, Muik A, Vormehr M, et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* (2021) 592:283–89. doi: 10.1038/s41586-021-03275

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 30 August 2022
DOI 10.3389/fimmu.2022.942907

OPEN ACCESS

EDITED BY
Adriana Harbuzariu,
Emory University, United States

REVIEWED BY
Sajjad Ahmad,
Abasyn University, Pakistan
Hamidreza Majidiani,
Neyshabur University of Medical Sciences, Iran

*CORRESPONDENCE
David Requena
drequena@rockefeller.edu

SPECIALTY SECTION
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

RECEIVED 13 May 2022
ACCEPTED 02 August 2022
PUBLISHED 30 August 2022

CITATION
Soto LF, Romaní AC, Jiménez-Avalos G, Silva Y, Ordinola-Ramirez CM, Lopez Lapa RM and Requena D (2022) Immunoinformatic analysis of the whole proteome for vaccine design: An application to *Clostridium perfringens*. *Front. Immunol.* 13:942907. doi: 10.3389/fimmu.2022.942907

Immunoinformatic analysis of the whole proteome for vaccine design: An application to *Clostridium perfringens*

Luis F. Soto[1], Ana C. Romaní[1], Gabriel Jiménez-Avalos[2], Yshoner Silva[3], Carla M. Ordinola-Ramirez[3], Rainer M. Lopez Lapa[3,4] and David Requena[5]*

[1]Escuela Profesional de Genética y Biotecnología, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Peru, [2]Departamento de Ciencias Celulares y Moleculares, Laboratorio de Bioinformática, Biología Molecular y Desarrollos Tecnológicos, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia (UPCH), Lima, Peru, [3]Departamento de Salud Pública, Facultad de Ciencias de la Salud, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru, [4]Instituto de Ganadería y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru, [5]Laboratory of Cellular Biophysics, The Rockefeller University, New York, NY, United States

Clostridium perfringens is a dangerous bacterium and known biological warfare weapon associated with several diseases, whose lethal toxins can produce necrosis in humans. However, there is no safe and fully effective vaccine against *C. perfringens* for humans yet. To address this problem, we computationally screened its whole proteome, identifying highly immunogenic proteins, domains, and epitopes. First, we identified that the proteins with the highest epitope density are Collagenase A, Exo-alpha-sialidase, alpha n-acetylglucosaminidase and hyaluronoglucosaminidase, representing potential recombinant vaccine candidates. Second, we further explored the toxins, finding that the non-toxic domain of Perfringolysin O is enriched in CTL and HTL epitopes. This domain could be used as a potential sub-unit vaccine to combat gas gangrene. And third, we designed a multi-epitope protein containing 24 HTL-epitopes and 34 CTL-epitopes from extracellular regions of transmembrane proteins. Also, we analyzed the structural properties of this novel protein using molecular dynamics. Altogether, we are presenting a thorough immunoinformatic exploration of the whole proteome of *C. perfringens*, as well as promising whole-protein, domain-based and multi-epitope vaccine candidates. These can be evaluated in preclinical trials to assess their immunogenicity and protection against *C. perfringens* infection.

KEYWORDS

immunoinformatics, Clostridium perfringens, epitope, toxin, vaccine, molecular dynamics

Introduction

Clostridium perfringens is a Gram-positive bacterium frequently associated with systemic and enteric diseases (1). In humans, *C. perfringens* is one of the most common food-poisoning causing bacteria, responsible for 1,000,000 cases per year in the US (2). Over a thousand cases result in gas gangrene, which can take from hours to weeks to develop depending on the tissue oxygen levels (3). It has a 10-30% mortality rate when treated, but 100% when untreated (4). Moreover, it is a pathogen known as a biological warfare weapon (5, 6). Therefore, it is necessary to develop preventive tools, like the identification of proteins and domains that can be used for molecular diagnostics and vaccines.

The histotoxic infections caused by this bacterium include gas gangrene in contaminated wounds, and several symptoms of human gastrointestinal diseases by either food- or non-food-borne *C. perfringens* infection (7, 8). *C. perfringens* isolates are classified into five toxinotypes, based on the production of four major toxins: α, β, ϵ and ι (1). *C. perfringens* type A is the main toxinotype that infects humans, producing gas gangrene, food poisoning, and non-foodborne gastrointestinal disease (9). Its main mechanism of cell invasion depends on the formation of a pore in the host cell membrane. The phospholipase C (cpa) and perfringolysin O (pfo) are involved in histotoxic infections, while the enterotoxin (etx), the *β* toxin (cpb) and the *β*-like toxin, the epsilon toxin (cpe) are involved in intestinal diseases (9, 10).

When *C. perfringens* toxins enter host cells, they are cut into small peptides by the proteasomes (11). Similarly, the whole bacterium can also be phagocytized and degraded by the endolysosomes into peptides. In both scenarios, the HLA (human leukocyte antigen) molecules (class I and class II, respectively) bind to these peptides and display them on the cell surface. Then, the HLA-peptide complexes are recognized by the TCR receptor in the surface of CD8+ and CD4+ immature T-cells, respectively, triggering an adaptive immune response. They will mature into cytotoxic T (CTL) and T-helper (HTL) lymphocytes. CTLs will produce a cytotoxic response against the infected cells, whereas HTL will stimulate the proliferation of antigen-specific B cells by clonal expansion, generating specific antibodies against the pathogen (12).

Vaccination is the most cost-effective method to prevent diseases (13). Although there is a vaccine available against *C. perfringens* for sheep and goats (14), there is no one approved for humans (5). Traditional vaccine development approaches are based on whole attenuated or dead microorganisms, and inactivated bacterial toxins. Nevertheless, they present the risk of potential reactivation or recombination of the vaccine strain, as well as offering limited protective effectiveness and immunity versus newer technologies (13). An alternative is subunit-based vaccines, consisting in one or more domains of antigenic proteins. Generally, these protein domains should be easily accessible, such as the external region of membrane proteins. As example, the NVX-CoV2373, a protein subunit-based vaccine producing the recombinant Spike protein, have demonstrated to neutralize the virus in different organism models and in humans (15, 16).

A new kind of vaccine, called the multi-epitope vaccine, has gained popularity in recent years (17). It consists of a novel protein connecting immunogenic epitopes. This kind of vaccine provides certain advantages over classical vaccines and single-epitope vaccines, such as a broader spectrum of pathogen variants, an optimized design that elicits both CTL, HTL, and B cell responses, and reduced adverse effects (18). It relies on the identification of epitopes through computational prediction and experimental testing. Epitopes are selected based on different criteria and assembled into a construct for their delivery to the immune system machinery. Multi-epitopes vaccines have already shown good results when tested in human clinical trials. For example, EMD640744 and Reniale have shown immunologic efficacy against advanced solid tumors (19) and in reducing tumor progression (20), respectively. Also, multi-epitope vaccines developed to trigger cross-immunity against different strains of influenza have shown great immunogenicity (21). Additionally, in mice, a multi-epitope vaccine has shown protective immunity against Toxoplasma gondii (22).

Nowadays, immunoinformatics tools help to massively screen protein sequences. They allow the computational identification of antigenic proteins and epitopes, reducing development time and cost (23). They rely on experimental data, which is available in databases like the Immune Epitope Database (IEDB) (24). It collects data of antibodies and CTL-, HTL- and B-epitopes, detected or evaluated in humans and other animal species (24, 25). Furthermore, there are predictors of CTL- and HTL-epitopes based on artificial intelligence tools and trained on experimental information. Among them, we have NetMHCpan I and II (26, 27) and MHCFlurry (28), which are currently the best according to independent evaluations (29). Additionally, Epitope-Evaluator performs comparative analysis of the outputs of these predictors, allowing an easy and graphical identification of highly antigenic proteins, as well as conserved promiscuous epitopes (30). Protein modeling and molecular dynamics (MD) allow studying structural characteristics of antigenic proteins, such as flexibility, disorder degree, and solvent accessibility, among others. These features are related to the immune response elicited (31–34). This could be because flexible and disordered antigens have more different conformations available at the moment of binding to the immune system molecules, maximizing favorable interactions (33).

In the present study, we predicted immunogenic epitopes from the 2721 proteins that comprise the known proteome of *C. perfringens* type A, to identify vaccine candidates following 3 different approaches. First, we used epitope prediction to identify

the proteins containing the highest number of epitopes that may elicit a good immunogenic response when used as recombinant vaccines. Second, we analyzed only the toxins and determined which non-toxic regions of these proteins are rich in HTL-epitopes and could be used as a vaccine, without the histotoxic damage. Third, using the best candidate epitopes, we generated synthetic constructs and studied their structural characteristics such as the flexibility and the accessibility of the epitopes through molecular dynamics. As result, we are presenting novel candidates for further testing as potential vaccines.

Material and methods

Data retrieval

Protein sequences and selection

We downloaded amino acid sequences of the 2721 proteins reported for *C. perfringens* Type A in the UniProt database, using the reference proteome with ID:UP000000818 (Supplementary Table 1). We explored the proteome, following three approaches (Figure 1).

In the first approach, we sought to identify proteins with most epitopes. So, the "epitope density" of each protein was calculated, which was defined as the number of epitopes of the protein divided by the length (in aa) of the protein. This was calculated using the tool "Epitope Density" from Epitope-Evaluator (https://fuxmanlab.shinyapps.io/Epitope-Evaluator/) (30). Only proteins longer than 100aa with "evidence at protein level" according to the UniProt database were considered in this approach (Figure 1A).

In the second approach, the epitopes within toxins were analyzed, using the following protein sequences: perfringolysin O (P0C2E9), enterotoxin A (Q8XKY4), enterotoxin B (Q8XKP0), enterotoxin D (Q8XMT2), beta2-toxin (Q93MD0) and phospholipase C (P0C216), as previously reported (10, 35). To propose a subunit-based vaccine that induces an appropriate humoral response, we further studied the HTL-epitopes in the non-toxic domains of these proteins. Among them, only the non-toxic domain of Perfringolysin O is well characterized, so we further studied this region. This protein is one of the most immunogenic toxins of *C. perfringens* (35), and its non-toxic domain is between the amino acid position 1 and 363, while the toxic and binding domains are reported to be between 363 and 472 (35, 36). This analysis was performed using the "Epitope Location" tool of Epitope-Evaluator (30) and the tridimensional model obtained by AlphaFold2 (Figure 1B).

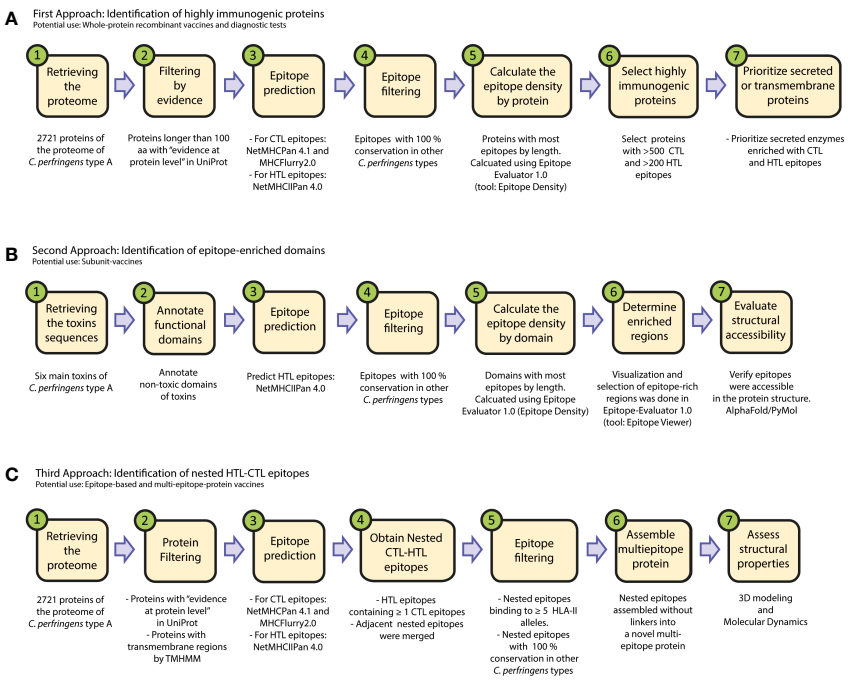


FIGURE 1
Flowchart of the immunoinformatic exploration of pathogens for vaccine development. We present three workflows to identify: **(A)** immunogenic proteins for protein-based vaccines, **(B)** protein domains enriched in HTL epitopes for subunit vaccines, and **(C)** nested epitopes for the design of novel multi-epitope protein vaccines. The proteins obtained in these three approaches might be used as well in immunodiagnostic tests.

In the third approach, we started with all the epitopes in the proteins of *C. perfringens* Type A and applied consecutive filters, as described below (see Figure 1C).

Selection of HLA class-I and -II alleles

Each population has a different distribution of HLA class I and II alleles (37, 38). Therefore, to maximize the potential use of our construct across populations, the HLA "supertype" alleles were considered as they are representative of the most frequent HLAs class I and II alleles worldwide. For HLA class -I, these are: HLA-A*01:01 (A1), HLA-A*26:01 (A1), HLA-A*02:01 (A2), HLA-A*03:01 (A3), HLA-A*24:02 (A24), HLA-B*07:02 (B7), HLA-B*08:01 (B8), HLA-B*27:05 (B27), HLA-B*39:01 (B27), HLA-B*40:01 (B44), HLA-B*58:01 (B58) and HLA-B*15:01 (B62) (39). And for HLA class-II: DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*11:01, DRB1*13:02 and DRB3*01:01 and DRB3*02:02 (9, 10, 40).

Prediction and selection of candidate epitopes

Prediction of CTL- and HTL-epitopes

Epitopes of 8, 9, and 10 amino acids (aa) were predicted in *C. perfringens* proteins for the HLA class-I supertype alleles using NetMHCpan v4.1 (27) and MHCFlurry 2.0 (28), selecting based on the consensus of these predictors. For HLA class-II, epitopes of 15 aa were predicted using the NetMHCIIpan v4.0 (26). A threshold of rank ≤ 2% was used for all these software. With the prediction results, the epitope promiscuity (number of HLA alleles each epitope is predicted to bind to) was calculated, identifying epitopes that can bind to many alleles.

Discarding epitopes present in the host proteome

A previous study has shown that epitopes present in both proteins of the pathogen and proteins of the host may trigger an auto-immune response (41). To prevent this, all predicted epitopes matching human proteins with 100% identity and sequence coverage were removed.

In the first and second approaches described in 2.1.1, the whole protein was discarded. In the third approach, just the epitope was discarded. All the proteoforms of the human proteome (UniProt ID: UP000005640) were used as reference.

Multi-epitope vaccine design

Identification of nested epitopes

To keep the artificial multi-epitope protein small while maximizing the number and quality of epitopes in the construct, highly immunogenic overlapping candidate epitopes were selected, which we called "nested epitopes". These were defined as linear HTL-epitopes containing linear CTL-epitopes in its sequence. Previous studies have also used this approach to induce both CD4+ and CD8+ T-cell responses (40, 41).

Filtering proteins by the number of transmembrane helices and UniProt evidence

To further filter the list of nested epitopes, only predicted epitopes belonging to proteins having two or more predicted external transmembrane regions with epitopes within were considered. Transmembrane helices were predicted using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (42, 43), and the presence of epitopes in external regions was checked with a custom script in Python. Additionally, only proteins with the following evidence categories according UniProt were considered: "Experimental evidence at protein level", "Experimental evidence at transcript level" and "Protein inferred from homology". The annotation found in UniProt was independently verified for correctness.

Selection of nested epitopes with high immunogenicity

From the proteins filtered above, only those nested epitopes predicted to bind 5 or more HLA-II supertype alleles were selected. Adjacent nested epitopes with overlap were merged to reduce the number of peptides while extending the predicted immunogenic regions. Non-overlapping nested epitopes were discarded.

Determination of highly conserved epitopes

Pathogens frequently mutate as an adaptation mechanism to environmental and immunological pressure, generating multiple variants (44). Selecting conserved regions may extend the validity of vaccines over time and confer protection against different strains. For this purpose, we performed BLASTp of our nested epitopes against *C. perfringens*. Epitopes with 90% of conservation or higher in the alignment were selected as candidate epitopes.

In addition, the conservation of each candidate nested epitope among the five *C. perfringens* toxinotypes was calculated. For each of the proteins having any of our candidate nested epitopes, a sequence alignment of the protein and its corresponding homologous proteins in the other toxinotypes was made. First, to find the homologous proteins, a Blast alignment was performed between the *C. perfringens* type A protein and the whole proteome of each toxinotype (Supplementary Table 2). Next, multiple global alignments were performed by protein using ClustalX2 (45). The alignments were visualized using CLC Sequence Viewer, where the regions corresponding to nested epitopes were extracted.

Design of multi-epitope vaccine candidates

Multi-epitope constructs were designed by concatenating the candidate epitopes in different orders. However, when assembling the constructs, neoepitopes that can bind to the HLA supertype alleles may appear at the interface of two candidate epitopes joined. Therefore, we need to know which epitope connections are allowed. This problem was modeled using a directed graph, where the nested candidate epitopes were represented as nodes, connected by directed edges representing the order of precedence in which the epitopes can be concatenated. To determine which directed edges are allowed in the graph, all the candidate epitopes were concatenated in a "pyramidal order", which is a single sequence containing all possible connections (see Figure 4A). This sequence was submitted to NetMHCIIpan to predict if there are strong binder epitopes located in the interface between two nested epitopes. Thus, the edges (connections) harboring unwanted neoepitopes were removed from the directed graph. Then, all the semi and complete Hamiltonian paths were found using EpiSorter, a python-based toolset for multi-epitope assembly (Figure 4B). Finally, those candidate epitopes either disconnected from the rest of the graph or with no edges enabling the generation of a complete path were discarded.

Additionally, neoepitopes at the interface of two candidate epitopes may be present in the host proteome, which may result in unwanted immune responses potentially leading to autoimmune reactions or tolerance. To prevent this, the multi-epitope constructs were sliced in all the possible fragments of 15 aa and we evaluated if they were present in any human protein using BLASTp against the human proteome. This was performed using BioPython and the following parameters: program = "blastp", database = "nr", entrez_query = "txid9606ORGN, expect = 20000, alignments = 100". For any 15-mer, if scores of 100% of coverage and 100% of identity were obtained, the corresponding construct was discarded.

Structural analysis of the vaccine constructs

Prediction of physicochemical properties

The physicochemical properties of the multi-epitope vaccines were calculated using ProtParam (https://web.expasy.org/protparam/) (46), which computes the molecular weight, theoretical isoelectric point, grand average of hydropathicity (GRAVY), among other metrics.

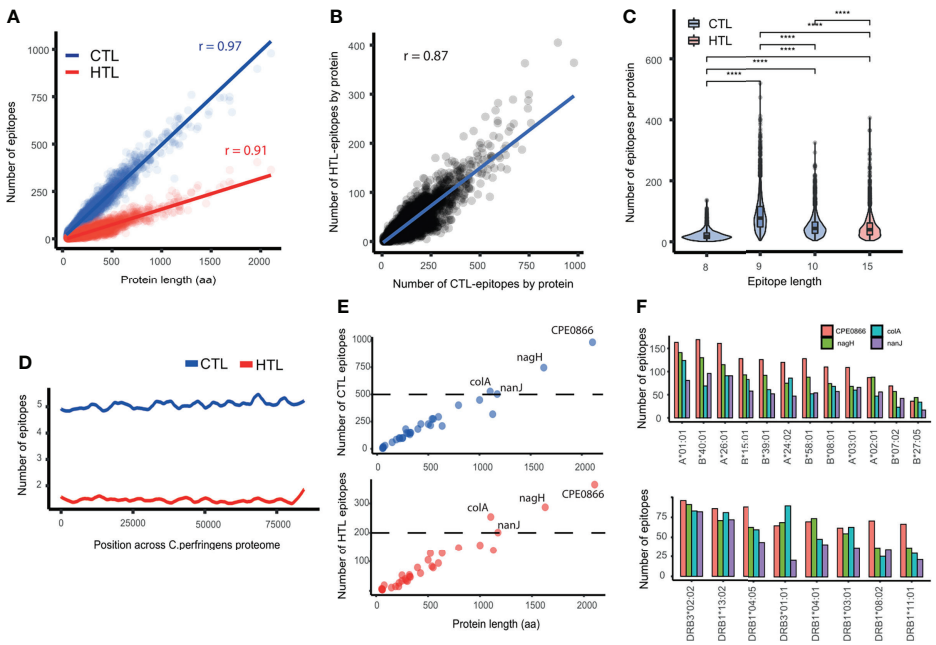


FIGURE 2
Description of the predicted epitopes in *C. perfringens*. **(A)** Correlation between the number of epitopes and protein length. Blue: CTL-epitopes. Red: HTL-epitopes. **(B)** Correlation between the number of HTL- and CTL-epitopes. **(C)** Comparison of the number of epitopes per protein, for different epitope lengths of CTL- (8, 9 and 10 aa) and HTL- (15 aa) epitopes. **(D)** Lasso regression of the number of CTL- (blue) and HTL- (red) epitopes across the proteome of *C. perfringens*. **(E)** Correlation between protein length and number of CTL- and HTL-epitopes in the protein (top and bottom, respectively). Proteins with more than 500 CTL- and 200 HTL-epitopes (above the dashed line) are labeled. **(F)** Bar plot showing the number of CTL- and HTL-epitopes (top and bottom, respectively) by HLA allele within, for the four proteins highlighted on panel **(E)**. **** means p-value < 0.0001.

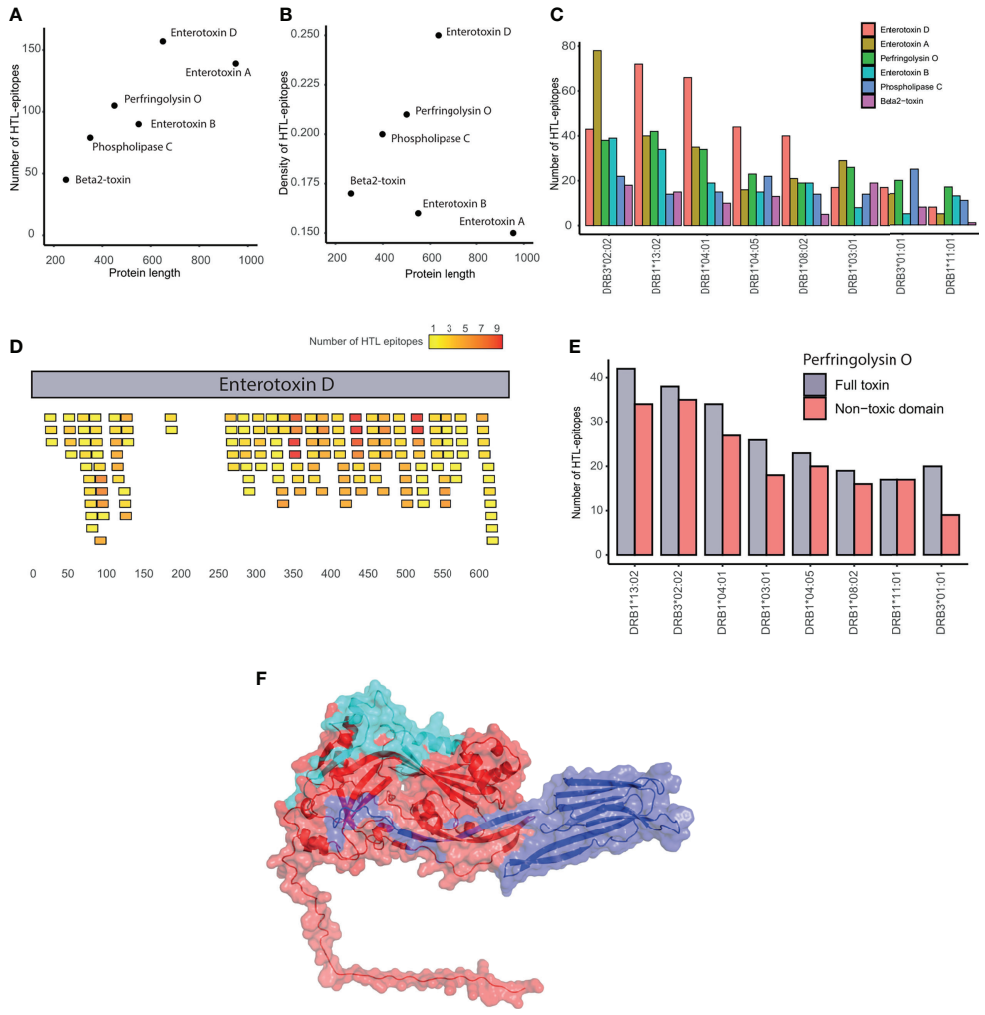


FIGURE 3
Evaluation of epitopes in *C. perfringens* toxins. **(A)** Correlation between the number of HTL-epitopes and protein length. **(B)** Correlation between HTL-epitope density and protein length. **(C)** Number of HTL-epitopes within toxins binding each of the HLA-II supertype alleles. **(D)** Location of the HTL-epitopes along the enterotoxin **(D)** Epitopes are colored in a gradient from yellow to red, representing the number of HLA alleles they bind. **(E)** Number of HTL-epitopes in the whole Perfringolysin O toxin (in gray) and its non-toxic domain (red) that are predicted to bind each of the HLA-II supertype alleles. **(F)** Structure of the Perfringolysin O. The toxic-domain is represented in blue. The non-toxic domain in red, highlighting the most promiscuous epitopes in cyan.

Structural modeling

The structure of the multi-epitope proteins were predicted with AlphaFold v. 2.1.0 (47), using all the databases to search for templates and the model "monomer_casp14" to infer the 3D coordinates. For each multi-epitope construct, its possible 3D structures were ranked by mean pLDDT and the best five structures were refined by restrained energy minimization with AMBER99SB, as implemented in the AlphaFold2 pipeline. Thus, selecting the structure with highest confidence for each multi-epitope construct. Then, the two multi-epitope proteins with the highest mean pLDDT were selected to analyze their structural conformation through MD simulation.

Molecular dynamics

We performed 1.2 μs of MD for each multi-epitope structure. First, the structures were submitted to the PDB2PQR web server (48) to add the corresponding hydrogens at pH 7.4. Then, MD was carried out using NAMD 2.14 (49) with the CHARMM36 force field (50). Accordingly, NaCl ions were included on the surface of the protein based on its Coulombic potential using the package cIonize 2.0 (51). Next, ions located 20 Å farther from the protein were removed. Finally, the protein was solvated in a size-optimized box with 15 Å of padding and a salt concentration of 0.154 M, using the Autoionize Plugin v.1.5 (https://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/).

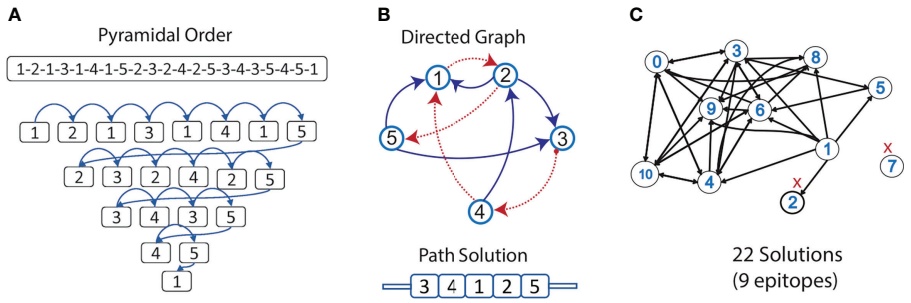


FIGURE 4
Selection of epitopes for the design of a multi-epitope construct. **(A)** An example of the pyramidal order for 5 epitopes, showing how they should be concatenated into a new protein, to evaluate the presence of neoepitopes in all the connections in just a single prediction step. **(B)** Allowed connections (without neoepitopes) are represented in the directed graph. The Hamiltonian path (in red) exemplifies a solution containing all the nodes. **(C)** Graph representing the nested epitopes (nodes) and its allowed connections (edges), selected in this study for the construction of the multi-epitope construct. Epitopes discarded from the design are marked in red.

Initially, only the water atoms were minimized for 5000 steps using the conjugate gradient algorithm. Then, MD of these atoms was performed at 0 K for 30 ps. Next, the whole system was minimized for 5000 steps, equilibrating the temperature to 298.15 K and the pressure 1 bar, as previously described (52). For this process, the Langevin thermostat and the Nosé-Hoover Langevin barostat were used in the NPT ensemble. Briefly, the system was heated from 50 K to 298 K, increasing the temperature by 4 K every 10 ps and applying harmonic restraints to the backbone with a force constant of 5 kcal/mol. Subsequently, the restraints were reduced by 10% every 0.05 ns. Finally, 1.2 μs of unrestrained MD of the NPT ensemble was performed at 298.15 K. Changes of temperature, potential energy, and density along the simulation were examined to verify convergence. All the processes described were performed in periodic boundary conditions with an integration time of 2 fs/timestep and Particle Mesh Ewald (PME) grid spacing of 1.0 Å. The cut-off for non-bonded interactions was set at 12 Å.

Assessing conformational convergence

For each trajectory, the alpha carbon root mean square deviation (Cα-RMSD) of all frames was calculated with an in-house Tcl script that uses VMD functions and the module "bigdcd". The frame 0 was used as reference point. The structural compactness was quantified by the radius of gyration (Rg), which was calculated using an in-house Tcl script similar to the one described above. Trajectories where the RMSD and Rg did not converge were discarded. The remaining trajectories were trimmed to only include the interval where the RMSD and Rg converged.

It is known that two completely different structures could have the same Cα-RMSD when compared against the same reference. Therefore, we complemented our graphical approach with a Cα-RMSD-based clustering in Wordom v.0.22 (53). This method assigns conformations to the same cluster if every pair in the group has a Cα-RMSD less or equal to a given threshold (cluster diameter) (54), which we set as 2.5 Å. Therefore, finding a cluster much bigger than the rest suggests that a dominant conformation was produced by the simulation, suggesting conformational convergence.

Final refinement and evaluation of the structural quality

The centroid of the most populated cluster of each remaining multi-epitope MD was subjected to a two-step final refinement. First, main-chain and fast all-atom energy minimizations were conducted using the web server ModRefiner (55), without a reference structure. Second, the protein was solvated, followed by a minimization and MD of only water atoms using the methodology described above. Then, the output was minimized for 5000 steps in explicit solvent with NAMD and the CHARMM36 force field, applying harmonic restraints to all protein atoms with a force constant of 1 kcal/mol. The refined structures were uploaded to the web server MolProbity to construct Ramachandran plots. In addition, the web implementations of ProSA (56) and ERRAT (57) were used to further assess the structural quality.

Analysis of epitope flexibility and accessibility in MEP_12 structure

For each MD of the multi-epitope constructs, the centroids of the five most populated clusters were compared against each other. First, to have all the structures in the same coordinate system, all the centroids were RMS-aligned against the initial AlphaFold2 model in Pymol v.2.4.1 (58). Then, the AlphaFold2

model was removed and Pymol selection algebra was used to obtain a list of residues forming the same secondary structure in all the centroids.

To identify the most rigid and accessible epitope within the multi-epitope construct, the convergent trimmed trajectories were analyzed, identifying the most flexible protein regions by computing the Cα-RMSF of the trimmed trajectory in VMD. The trimmed trajectory was loaded with a step of 5 and aligned against frame 0, then the Cα-RMSF measured using these modified coordinates. Finally, the Solvent Accessible Surface Area (SASA) per epitope was measured during the trimmed trajectory using an in-house Tcl script. Then, the module "bigdcd" was used to load the complete trimmed trajectory and examine the changes of SASA along it, assessing the convergence of each epitope. For each epitope whose SASA converged, the mean SASA over the entire trajectory was computed and 95% confidence intervals calculated by computing the statistical inefficiency with the block averaging approach (59).

Evaluation of MEP_12 innate response potential

The potential of MEP_12 to trigger an innate response was tested based on its interaction with TLR1/TLR2 and TLR4/MD2. The representative conformation of MEP_12, from MD Ca-RMSD-based clustering, was docked against the ectodomains of TLR1/TLR2 (PDB ID: 2Z7X) and TLR4/MD-2 (TLR4/MD-2, PDB ID: 3FXI) using Haddock v.2.4 (60). The structures of these TLRs were submitted to PDBfixer to complete the missing sections and remove irrelevant heteroatoms (61). As TLR ectodomains are glycosylated *in vivo* (62), the covalently attached glycans of the structures were kept in order to obtain a TLR-MEP_12 interaction closer to real conditions.

As information about the possible interaction sites was not available, a blind docking approach was followed. Haddock ab-initio mode was used to scan the surface of MEP_12 and the corresponding TLR ectodomain, to find the most favorable interacting pose. This was performed during the rigid docking phase. Fifty thousand structures were computed for this step, to ensure that the whole surface of each protein will be sampled. Then, the best 500 structures were selected based on the Haddock scores, which were calculated from the semi-flexible simulated annealing and the final energy minimization refinement. The solutions of both docking steps were clustered by Fraction of Common Contacts (FCC), using 0.6 Å as threshold. Cluster-mean Haddock scores were computed using the structures with the four lowest values. Then, the clusters were ranked, where those with lower mean Haddock scores were the most favorable.

The most favorable clusters that had overlapping error bars (one standard deviation) of their Haddock-scores were selected. Protein-protein interaction analysis of the epitopes was performed using a custom Python pipeline, evaluating the plausibility of the solution based on prior structural data. After selecting one cluster, its most favorable structure was submitted to the web servers PDBsum and PRODIGY to map the interacting residues and to compute the protein-protein binding energy, respectively (63, 64).

Simulating the immune response

The immune response profile by immunization with the multi-epitope vaccine was simulated using C-IMMSIM (https://kraken.iac.rm.cnr.it/C-IMMSIM/). Three vaccinations with a dose of 1000 unit of the multi-epitope vaccine and 100 of adjuvant without LPS were administrated as described in Figure 10A. The most frequent HLA alleles were considered: HLA-A*0101, HLA-B*08:01, HLA-B*15:01, DRB1-11:01 and DRB3-02:02. One thousand timesteps of 8 hours were simulated, representing ~11 months. Vaccine doses were administered four weeks apart, as commonly recommended (65), corresponding to days 3, 30 and 60. Pathogen challenge was introduced on day 111 by inoculating 1000 units of those proteins of *C. perfringens type A* harboring epitopes of the vaccine, with a pathogen multiplication factor of 0.2 (Figure 10A).

Multi-epitope sequence design and codon optimization

Codon optimization was performed to improve the expression efficiency of the vaccine construct in *Escherichia coli* for production. The codon usage table of *E. coli* K12 strain, available in the Codon Usage Database (https://www.kazusa.or.jp/codon/), was used for reverse translation. The CAIcal SERVER (http://genomes.urv.es/CAIcal/) was used to calculate the Codon Adaptation Index (CAI). The cDNA sequence obtained was analyzed with NEBcutter (http://nc2.neb.com/NEBcutter2/), identifying cleavage sites of commercially available restriction enzymes.

Results

Identification of proteins with high epitope density

A total of 429809 CTL- and 121450 HTL-epitopes were predicted from the proteome of *C. perfringens* Type A. The number of CTL- and HTL-epitopes predicted were strongly correlated with the protein length (CTL: $p < 2.2e\text{-}16$, $r = 0.97$; HTL: $p < 2.2e\text{-}16$, $r = 0.91$, Pearson correlation) (Figure 2A). There is also a positive correlation between the number of CTL- and HTL-epitopes predicted by protein ($p < 2.2e\text{-}16$, $r = 0.87$, Pearson correlation) (Figure 2B), indicating that *C. perfringens* proteins containing a higher number of predicted CTL-epitopes tend to contain more HTL-predicted epitopes as well.

Regarding the epitope length, we found a significantly greater number of 9-mer than 8-mer or 10-mer predicted

CTL-epitopes (Wilcoxon test, p value < 2.2e-16), as biologically expected. Additionally, the number of 9-mer CTL epitopes predicted by protein was higher than the number of HTL-epitopes (Figure 2C). These observations were statistically significant even when normalizing by protein length (Wilcoxon test, p < 2.2e-16) (Supplementary Figure 1A).

The distribution of predicted epitopes across the proteome was explored by randomly concatenating the sequences of all the 2721 proteins and calculating the epitope density along it. Local peaks were detected, showing that the predicted epitopes were not evenly distributed across the proteome and that not all proteins have the same epitope density (Figure 2D).

Epitopes that were either duplicated (i.e. appearing in two or more different proteins) or present in any human protein were removed, resulting in 121244 (99.83%) HTL- and 427944 (99.56%) predicted CTL-epitopes left (Supplementary Table 3). This comprises 59578, 237664 and 130702 CTL-epitopes predicted of 8, 9 and 10 aa, respectively (Supplementary Figure 1B). Regarding the epitope promiscuity (i.e. the number of alleles an epitope can bind to), we found that the majority of the CTL-epitopes predicted (312032 out of 427944) bind to only one HLA-I allele. Similarly, 67080 of the 121244 HTL-epitopes were predicted to bind just one HLA-II. Noteworthy, there were 2 promiscuous CTL-epitopes binding to 11 HLA-I alleles, and 33 HTL-epitopes binding to 8 HLA-II alleles (Supplementary Figures 1C, D).

The evidence status of the proteins with at least 1 epitope predicted was retrieved from UniProt, obtaining 32 with "evidence at protein level", 1 with "evidence at transcript level", 1064 with "inferred from homology", 1623 "predicted", and 1 "uncertain". The set of proteins with "evidence at protein level", when compared with the set of proteins "inferred from homology", showed no significant difference in the number of epitopes (p = 0.33) but in epitope density (p < 0.02) (Supplementary Figures 1E, F).

Among the proteins with "evidence at protein level", 25 show an epitope density above 0.5 (Supplementary Table 4). Notoriously, each of the proteins Collagenase A, Exo-alpha-sialidase, alpha-n-acetylglucosaminidase and hyaluronoglucosaminidase contain more than 500 CTL- and 200 HTL-epitopes (Figure 2E), and more than 15 epitopes per HLA supertype allele (Figure 2F).

Evaluation of the HTL-epitopes in *C. perfringens* toxins

The six toxins evaluated have a different number (45–157) but similar density (0.15-0.25) of HTL-epitopes. Notably, enterotoxin D contains the highest number of HTL-epitopes even not being the largest toxin. Contrarily, beta2-toxin showed the lowest number of HTL-epitopes (Figure 3A). In terms of HTL-epitope density, the enterotoxins A and D showed the lowest and highest values, respectively (Figure 3B).

All the toxins, except the beta2-toxin, contained several epitopes by HLA allele. Among the HLA-II alleles used, DRB3*02:02 and DRB1*11:01 were found to recognize the highest and the lowest number of epitopes from toxins (Figure 3C). Regions enriched with promiscuous HTL-epitopes within the toxins were identified. For example, the C-terminal region of the enterotoxin D contains the most promiscuous epitopes of the protein, while the region between positions 130 and 256aa has almost no epitopes (Figure 3D). Similarly, the HTL-epitopes of the enterotoxins A and B were mainly located at the N-terminal region. Additionally, the non-toxic domain of Perfringolysin O contains at least nine predicted epitopes per HLA-II allele (Figure 3E). Furthermore, among the toxins, we identified 12 HTL-epitopes binding five or more HLA-II alleles, covering the majority of HLA-II supertype alleles (Table 1). We also found that four of these promiscuous epitopes are located in the external region of the non-toxic domain of the Perfringolysin O (Figure 3F). Altogether, it suggests that this domain can be considered as a potential subunit-based vaccine.

Design of multi-epitope vaccine candidates

We identified 112714 nested epitopes from the whole proteome, comprising 112714 predicted HTL-epitopes containing 145854 predicted CTL-epitopes (Supplementary Table 5). Next, from the 2685 proteins containing nested epitopes, 266 proteins had at least two external transmembrane regions with at least one nested epitope. Then, the proteins annotated as "predicted proteins" or "uncertain proteins" were removed, resulting in 105 proteins containing 1884 nested epitopes. From this set, only nested epitopes that can bind to at least five HLA-II alleles were selected, obtaining 42. And it was possible to merge 29 of these 42 nested epitopes into 11 overlapped nested epitopes (Table 2), which represent the candidates for the design of the muti-epitope construct (Figure 1C). The conservation analysis showed that these candidate epitopes were highly conserved among the sequences of the different strains of C. perfringens (Table 2). Moreover, these epitopes are 100% conserved among the five toxinotypes (Supplementary Figure 2).

These 11 overlapped nested epitopes were used to build the directed graph. By finding the allowed directed edges, epitopes "7" was discarded as it was disconnected from the rest of the graph. Also, epitope "2", because there were no edges permitting the generation of a complete path (Figure 4C). From the subgraph containing the nine remaining overlapped nested epitopes, we obtained 21 different Hamiltonian paths, representing multi-epitope constructs (Supplementary

10.3389/fimmu.2022.942907

TABLE 1 Promiscuous HTL-epitopes in *C. perfringens* toxins.

Epitopes	Alleles	Number of Alleles	Proteins
PENIKIIANGKVVVD	DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*11:01, DRB1*13:02, DRB3*02:02	7	PHOSPHOLIPASE C
PKYIVIHDTDNRQAG	DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*13:02, DRB3*01:01	6	ENTEROTOXIN D
RKPININIDLPGLKG-NPKYIVIHDTDNRQA	DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*13:02, DRB3*02:02	6	PERFRINGOLYSIN O - ENTEROTOXIN D
MLEEFKYDPNQQLKS-LEEFKYDPNQQLKSF	DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*13:02, DRB3*01:01, DRB3*02:02	6	BETA2 TOXIN
LKSFEILNSQKIDNK	DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*11:01, DRB1*13:02, DRB3*02:02	6	BETA2 TOXIN
KYIVIHDTDNRQAGA	DRB1*03:01-DRB1*04:01-DRB1*04:05, DRB1*13:02-DRB3*01:01	5	ENTEROTOXIN D
KRKPININIDLPGLK	DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*13:02, DRB3*02:02	5	PERFRINGOLYSIN O
EIRKVIKDNATFSTK-IRKVIKDNATFSTKN-NDNINIDLSNSNVAV-EMLEEFKYDPNQQLK-EEFKYDPNQQLKSFE	DRB1*03:01, DRB1*04:01, DRB1*13:02, DRB3*01:01, DRB3*02:02	5	PERFRINGOLYSIN O - PERFRINGOLYSIN O - ENTEROTOXIN A - BETA2 TOXIN - BETA2 TOXIN
GEIFNIDGKEGSWYK	DRB1*03:01, DRB1*08:02, DRB1*11:01, DRB1*13:02, DRB3*01:01	5	ENTEROTOXIN B
ENIKIIANGKVVVDK-NIKIIANGKVVVDKD	DRB1*0301, DRB1*08:02, DRB1*11:01, DRB1*13:02, DRB3*02:02	5	PHOSPHOLIPASE C
WNEKYSSTHTLPART-NEKYSSTHTLPARTQ-GSNYGVIGTLRNNDK-ASKSYITIVNEGSNN-SKSYITIVNEGSNNG	DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*11:01, DRB3*02:02	5	PERFRINGOLYSIN O - PERFRINGOLYSIN O - ENTEROTOXIN D - ENTEROTOXIN D - ENTEROTOXIN D
KQGIVKVNSALNMRS-KSFEILNSQKIDNKE	DRB1*04:01, DRB1*04:05, DRB1*08:02, DRB1*13:02, DRB3*02:02	5	ENTEROTOXIN D - BETA2 TOXIN

TABLE 2 Characteristics of the predicted epitopes selected for the construction of the multi-epitope protein.

Epitope	Overlaped epitope	Nested_epitope	HLA-II alleles	Protein name	Protein ID	Position	Epitope	HLA-I Alleles	Conservation
Ep_0	IDGKEYKIANNALIGEGK	IDGKEYKIANNALIG	DRB1*13:02, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01	FtsX domain-containing protein	Q8XM39	453	EYKIANNALI	A*24:02	160/161
							KEYKIANNAL	B*40:01	
							EYKIANNAL	A*24:02, B*08:01, B*39:01	
							YKIANNALI	B*39:01	
							KEYKIANNA	B*40:01	
		DGKEYKIANNALIGE	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB3*01:01, DRB1*04:01			454	EYKIANNALI	A*24:02	
							KEYKIANNAL	B*40:01	
							EYKIANNAL	A*24:02, B*08:01, B*39:01	
							YKIANNALI	B*39:01	
							KEYKIANNA	B*40:01	
		GKEYKIANNALIGEG	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB3*01:01, DRB1*04:01			455	EYKIANNALI	A*24:02	
							KEYKIANNAL	B*40:01	
							EYKIANNAL	A*24:02, B*08:01, B*39:01	

(Continued)

TABLE 2 Continued

Epitope	Overlaped epitope	Nested_epitope	HLA-II alleles	Protein name	Protein ID	Position	Epitope	HLA-I Alleles	Conservation
							YKIANNALI	B*39:01	
							KEYKIANNA	B*40:01	
		KEYKIANNALIGEGK	DRB1*13:02, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01			456	EYKIANNALI	A*24:02	
							KEYKIANNAL	B*40:01	
							EYKIANNAL	A*24:02, B*08:01, B*39:01	
							YKIANNALI	B*39:01	
							KEYKIANNA	B*40:01	
Ep_1	LYEKGFLHAKTIVADSS	LYEKGFLHAKTIVAD	DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01	Cardiolipin synthase	P0C2E2	387	FLHAKTIV	B*08:01	89/89
							LHAKTIVA	B*39:01	
							FLHAKTIVA	A*02:01, B*08:01	
		YEKGFLHAKTIVADS	DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01			388	FLHAKTIV	B*08:01	
							LHAKTIVA	B*39:01	
							FLHAKTIVA	A*02:01, B*08:01	
		EKGFLHAKTIVADSS	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01			389	FLHAKTIV	B*08:01	
							LHAKTIVA	B*39:01	
							FLHAKTIVA	A*02:01, B*08:01	
Ep_2*	EGKIVVIIDNSPSVIIL	EGKIVVIIDNSPSVI	DRB1*13:02, DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01	Stage V sporulation protein AF	Q8XLQ7	245	VIIDNSPSV	A*02:01, A*26:01	85/86
							IIDNSPSVI	A*02:01	
		GKIVVIIDNSPSVII	DRB1*13:02, DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01			246	VIIDNSPSV	A*02:01, A*26:01	
							IIDNSPSVI	A*02:01	
		KIVVIIDNSPSVIIL	DRB1*13:02, DRB3*02:02, DRB1*03:01, DRB3*01:01, DRB1*04:01			247	VIIDNSPSV	A*02:01, A*26:01	
							IIDNSPSVI	A*02:01	
							DNSPSVIIL	B*39:01	
Ep_3	GAERFVLISTDKAVNPT	GAERFVLISTDKAVN	DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01	Polysacc synt 2 domain-containing protein	Q8XN75	406	FVLISTDKAV	A*02:01	30/32
							ERFVLISTDK	B*27:05	
							VLISTDKAV	A*02:01	
							AERFVLIST	B*40:01	
		AERFVLISTDKAVNP	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01			407	FVLISTDKAV	A*02:01	
							ERFVLISTDK	B*27:05	
							VLISTDKAV	A*02:01	
							AERFVLIST	B*40:01	
		ERFVLISTDKAVNPT	DRB1*13:02, DRB1*11:01, DRB1*04:05, DRB1*08:02, DRB1*04:01			408	FVLISTDKAV	A*02:01	
							ERFVLISTDK	B*27:05	
							VLISTDKAV	A*02:01	
Ep_4	IKENEFVVDGSTRLSDL	IKENEFVVDGSTRLS	DRB1*04:01, DRB3*01:01, DRB3*02:02, DRB1*03:01, DRB1*13:02	Probable hemolysin-related protein	Q8XPD3	339	FVVDGSTRL	A*02:01, A*26:01	41/41
		KENEFVVDGSTRLSD	DRB1*04:01, DRB3*01:01, DRB3*02:02, DRB1*03:01, DRB1*13:02			340	FVVDGSTRL	A*02:01, A*26:01	
		ENEFVVDGSTRLSDL	DRB1*13:02, DRB3*02:02, DRB1*03:01, DRB3*01:01, DRB1*04:01			341	FVVDGSTRL	A*02:01, A*26:01	
							DGSTRLSDL	B*08:01	

(Continued)

TABLE 2 Continued

Epitope	Overlaped epitope	Nested_epitope	HLA-II alleles	Protein name	Protein ID	Position	Epitope	HLA-I Alleles	Conservation
Ep_5	RHKDKIYIDTSPVNNLI	RHKDKIYIDTSPVNN	DRB1*04:01, DRB1*04:05, DRB3*01:01, DRB3*02:02, DRB1*03:01, DRB1*13:02	TraG-D C domain-containing protein	Q93M96	158	KIYIDTSPV	A*02:01	70/82
		HKDKIYIDTSPVNNL	DRB1*13:02, DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01			159	YIDTSPVNNL	A*02:01	
							KIYIDTSPV	A*02:01	
							IDTSPVNNL	B*40:01	
		KDKIYIDTSPVNNLI	DRB1*13:02, DRB3*02:02, DRB1*03:01, DRB3*01:01, DRB1*04:01			160	YIDTSPVNNL	A*02:01	
							KIYIDTSPV	A*02:01	
							DTSPVNNLI	A*26:01	
							IDTSPVNNL	B*40:01	
Ep_6	ASATYYIDEDSKIKTA	ASATYYIDEDSKIKT	DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01	FtsX domain-containing protein	Q8XM39	331	ATYYIDEDSK	A*03:01	126/128
							TYYIDEDSKI	A*24:02	
							YIDEDSKIK	A*01:01	
							YYIDEDSKI	A*24:02	
		SATYYIDEDSKIKTA	DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01			332	ATYYIDEDSK	A*03:01	
							TYYIDEDSKI	A*24:02	
							YIDEDSKIK	A*01:01	
							YYIDEDSKI	A*24:02	
Ep_7*	VPDNIVSNLKPIANKI	VPDNIVSNLKPIANK	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*03:01, DRB1*08:02	FtsX domain-containing protein	Q8XM39	490	VSNLKPIANK	A*03:01	136/147
							VPDNIVSNL	B*07:02, B*08:01, B*39:01	
							SNLKPIANK	A*03:01	
		PDNIVSNLKPIANKI	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*03:01, DRB1*08:02			491	VSNLKPIANK	A*03:01	
							SNLKPIANK	A*03:01	
							NLKPIANKI	B*08:01	
Ep_8	LDYKFILDTNYIEAKL	LDYKFILDTNYIEAK	DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01	Spore germination protein KA	Q8XMP0	191	FILDTNYIEA	A*02:01	42/43
							ILDTNYIEAK	A*03:01	
							ILDTNYIEA	A*03:01	
							KFILDTNYI	A*24:02	
							YKFILDTNY	B*15:01	
		DYKFILDTNYIEAKL	DRB3*02:02, DRB1*04:05, DRB1*03:01, DRB3*01:01, DRB1*04:01			192	FILDTNYIEA	A*02:01	
							ILDTNYIEAK	A*03:01	
							ILDTNYIEA	A*03:01	
							KFILDTNYI	A*24:02	
							DTNYIEAKL	A*26:01	
							YKFILDTNY	B*15:01	
Ep_9	LDDFITIEKANNSYTF	LDDFITIEKANNSYT	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*08:02, DRB1*04:01	Cardiolipin synthase	Q8XP94	265	ITIEKANNSY	A*01:01, A*26:01, B*15:01, B*58:01	114/115
							TIEKANNSY	A*01:01, A*26:01, B*15:01	
		DDFITIEKANNSYTF	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*08:02, DRB1*04:01			266	ITIEKANNSY	A*01:01, A*26:01, B*15:01, B*58:01	
							IEKANNSYTF	B*40:01	
							KANNSYTF	B*58:01	
							TIEKANNSY		

(Continued)

TABLE 2 Continued

Epitope	Overlaped epitope	Nested_epitope	HLA-II alleles	Protein name	Protein ID	Position	Epitope	HLA-I Alleles	Conservation
								A*01:01, A*26:01, B*15:01	
Ep_10	SDNDYVIVNTEGGEFD	SDNDYVIVNTEGGEF	DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01	UPF0182 protein CPE0011	Q8XPF2	461	VIVNTEGGEF	B*15:01	173/174
							IVNTEGGEF	A*26:01, B*15:01	
		DNDYVIVNTEGGEFD	DRB1*13:02, DRB1*11:01, DRB3*02:02, DRB1*04:05, DRB1*08:02, DRB1*04:01			462	VIVNTEGGEF	B*15:01	
							IVNTEGGEF	A*26:01, B*15:01	

Epitopes "2" and "7" (with asterisk) were not included in the final design. Conservation is represented as the number of sequences where epitope is conserved over the total number of sequences analyzed.

Table 6). Using BLASTp, we obtained that only a 5 aa fragment of an epitope (DTNYI) matched with a human protein (ATP-dependent DNA helicase HFM1) with a 100% coverage and identity. Therefore, no construct was discarded.

Structural analysis of the multi-epitope constructs

Physicochemical characterization

We evaluated the physicochemical properties of the constructs using ProtParam (Supplementary Table 7). Each one had 150 aa, a molecular weight of 16.83 kDa and a predicted pI of 4.61, indicating an acidic nature. The number of negatively and positively charged residues computed at pH 7 were 27 and 17, respectively. The aliphatic index (relative volume occupied by the aliphatic side chains) was 93, indicating a thermostable nature. The construct had a GRAVY value of -0.363.

Modelling the constructs

Models of each multi-epitope construct were prepared in AlphaFold2, and its mean pLDDT were computed. Values under 50 were obtained for all models (Supplementary Figure 3A), suggesting structural disorder (66). After modeling the 21 multi-epitope constructs, only MEP_6 and MEP_12 were selected, as they showed the highest mean pLDDT.

The structures of the multi-epitope (MEP) MEP_6 and MEP_12 had the highest mean pLDDT values, being the most reliable ones. Besides the similarity in their mean pLDDT, the distribution of per-residue pLDDT notably differed. MEP_6 had 94.7% of its residues with pLDDT under 50 and 5.3% between 50 and 70 (Figure 5A). In contrast, MEP_12 had 68% of residues with pLDDT under 50, 31.3% between 50 and 70, and 0.7% between 70 and 90 (Figure 5B).

In addition, the per-residue pLDDT seemed to be non-dependent on the epitope from which the residues belong. For instance, the residues of epitope "4" had pLDDTs under 50 in MEP_6, whereas values between 50 and 70 in MEP_12 (Figure 5).

Assessing the conformational convergence

The confidence of our best candidate models (MEP_6 and MEP_12) were improved by performing extensive MDs to refine them, and their conformational changes were evaluated. The RMSD and Rg over each trajectory were measured to assess their conformational convergence.

The RMSD of MEP_6 increased during the last 600 ns of MD, which is indicative of a conformation that is still changing (Figure 6A). The conformational instability was corroborated by an unstable Rg, which continuously decreased during all the trajectory (Figure 6B). Therefore, MEP_6 was discarded of further analysis as it did not reach conformational convergence.

In contrast, the RMSD and Rg of MEP_12 reached a steady state during the last 500 ns (Figures 6C, D) suggesting conformational convergence of the trajectory. It was verified by an RMSD-based clustering of frames of the last 500 ns, using 2.5 Å as threshold. More than 70% of the conformations adopted during this lapse were highly similar among them and were grouped in Cluster 1, confirming the convergence (see Figure 6E; Supplementary Table 8).

As the centroid of Cluster 1 represents the preferred conformation of MEP_12, it was retrieved and considered as the most probable average structure of MEP_12 in solution. Remarkably, although displaced due to the MD, the β-strands formed by the same residues are present both in the centroid and in the initial AphaFold2 model (Figure 6F).

Evaluating the structural quality

A two-steps refinement of the centroid was performed using ModRefiner and NAMD, obtaining a structure that remained close to the original conformation (Supplementary Figure 3B). The quality of the refined structure was assessed using the web servers MolProbity, ProSA and ERRAT. We observed that all the

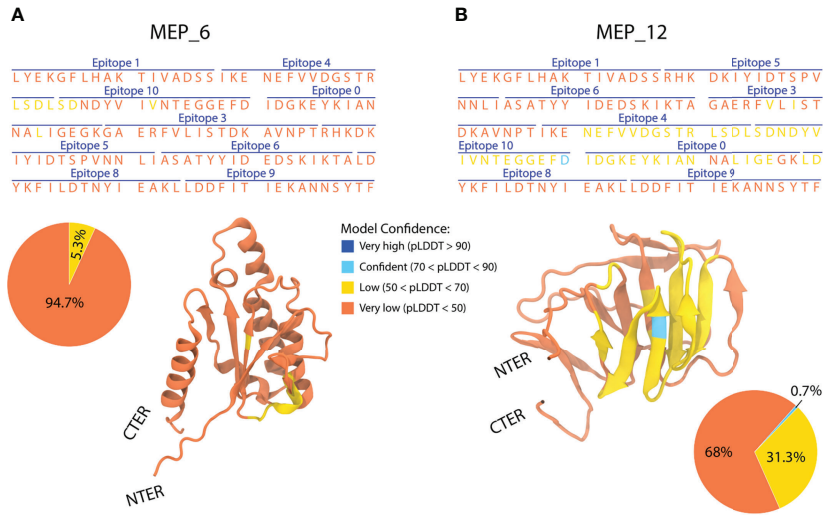


FIGURE 5
Amino acid sequence, epitope sorting and AlphaFold2 3D model of the multi-epitope constructs MEP_6 **(A)** and MEP_12 **(B)**. The confidence value (pLDDT) is categorized in 4 groups (orange, yellow, cyan, and blue), representing their percentage in pie charts.

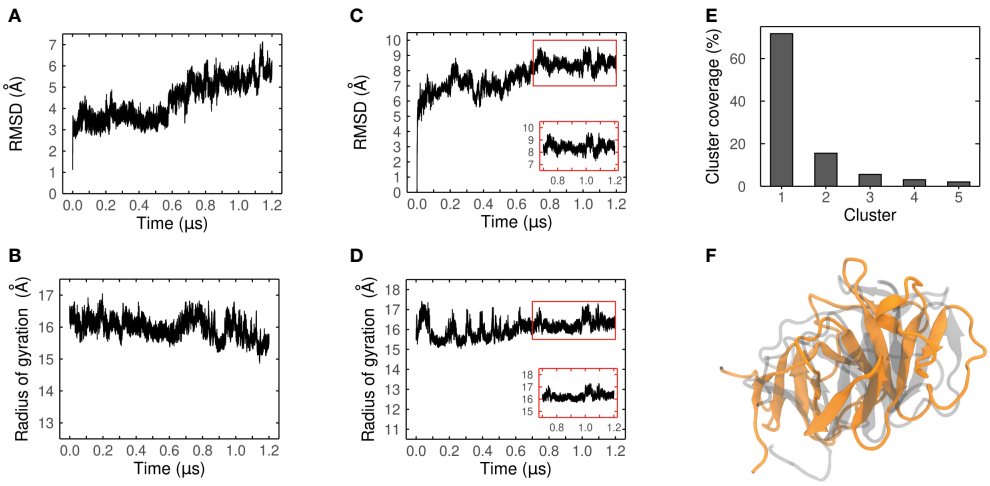


FIGURE 6
Convergence analysis of the MD simulation. Ca-RMSD values and radius of gyration (Rg) of MEP_6 **(A, B)** and MEP_12 **(C, D)** during a simulation of 1.2 μs. Inset plots in **(C, D)** show the last 500 ns where the RMSD and Rg converged. **(E)** Coverage (%) of the five most populated clusters obtained from the RMDS-based clustering. **(F)** Structural alignment between the AlphaFold2 model (gray) and the centroid of the most populated cluster (in orange) of MEP_12.

residues in the model were correctly oriented, as the Ramachandran plot indicated that 100% of residues were in allowed regions, not showing any outliers (Figure 7A). Furthermore, 96.6% of the residues were observed in favored regions. ProSA predicted a Z-score of -6.55 for our model, which is within the range of values observed in other proteins of similar length, obtained from NMR or X-ray crystallography (Figure 7B). An ERRAT overall score of 94.531 was obtained, indicating that this model has a good resolution. Notably, only windows centered at residues 96 and 127 had error values above 99% (Figure 7C). All the quality statistics were better in the refined structure than the original model (Supplementary Figure 4).

Analysis of MEP_12 trajectory

An ensemble approach was followed to assign the secondary structures of the aminoacidic residues. The centroids of the five

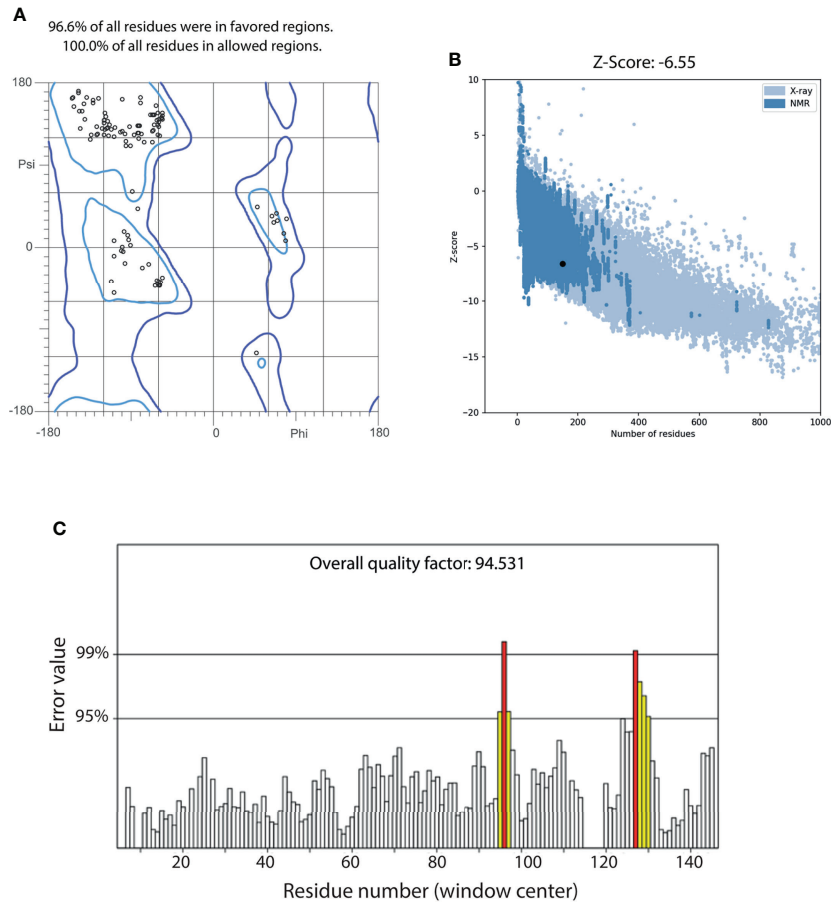


FIGURE 7
Quality assessment of the refined centroid of MEP_12 most populated cluster. **(A)** Ramachandran plot of MEP_12 centroid, indicating the percentage of residues in favored (light blue) and allowed (blue) regions. **(B)** Scatterplot of the Z-scores of MEP_12 centroid (black dot) and structures with experimental evidence obtained from NMR (blue) and X-ray crystallography (light blue). **(C)** ERRAT plot of MEP_12 centroid. Bars represent the error value (white: error < 95%, yellow: 95%, error < 99%, red: error > 99%) of a nine-residue sliding window. The overall quality factor indicates the percentage of protein residues with error values lower than 95%.

most populated clusters were compared, as they are distinct and representative conformations of MEP_12 in solution. Then, sets of residues forming the same secondary structure in all the centroids were retrieved. We reasoned that if those sets are folded in the same way, even in different but representative conformations, there is a high chance that they might adopt that conformation when synthesized. Notably, comparing the five centroids, we found that 12 β-strands were formed by the same sets of residues (Figure 8A). The number of residues forming these β-strands represent 41.33% of the MEP_12. The other residues were mostly in loops in all the centroids evaluated. No consistent alpha helices were found.

By analyzing the Cα-RMSF, epitopes "3", "8" and "10" were found to be the least flexible ones (Figure 8B). The rigidity of "3" can be explained because part of this epitope forms a parallel β-sheet with a portion of epitope "6" (Figure 8D). Next, the mean SASA over the steady state of the MD was computed, excluding epitope "4" as it was not stable over the MD (Supplementary Figure 5). Epitopes "3", "10", and "8" showed the lowest mean SASA values, in decreasing order (Figure 8C). Altogether, this indicates that epitope "3" is the most accessible among the rigid epitopes of MEP_12.

Evaluation of MEP_12 innate response potential

Blind docking of MEP_12 against TLR1/TLR2 resulted in four clusters containing 3.6% of the refined structures. The Haddock-scores of these clusters showed overlapping error bars (one standard deviation) (Figure 9A), not allowing to discern which cluster is the best.

The TLR1/TLR2 crystal structure had missing residues in the N-termini of both TLRs. As the cluster "1" contained binding

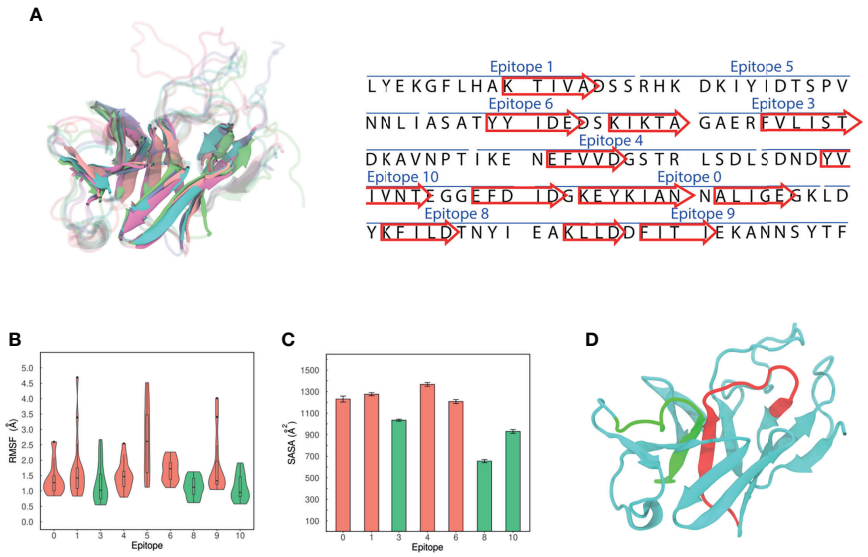


FIGURE 8
Structural characteristics of MEP_12. **(A)** Alignment of the 3D structures of MEP_12 from all the centroids of the five most populated clusters. Cyan, purple, orange, green, and pink cartoons correspond to the centroids of clusters 1, 2, 3, 4 and 5 respectively. The twelve β-strands in the structure (left) are represented as arrows in the sequence (right). **(B)** Violin plots showing the distribution of Ca-RMSF values. The lowest Ca-RMSF distributions are colored in green. **(C)** Mean SASA by epitope. The error bars represent 95% confidence intervals. **(D)** Modeled structure of MEP_12, showing epitopes "3" and "6" in red and green, respectively.

modes of MEP_12 interacting with TLR2 in a region that might be occupied by those missing residues, it was discarded (Supplementary Figures 6A, B). All the other clusters showed MEP_12 docked to similar regions of TLR1 (Supplementary Figure 6A). The best binding mode of the cluster "2" showed contacts with epitopes "3", "4" and "6". It showed four hydrogen bonds and one salt bridge that stabilized the interaction (Figure 9B). The best binding modes of the cluster "3" and "4" showed contacts were with epitopes "6" and "0"; and "0", "1" and "9", respectively (Supplementary Table 9).

Blind docking of MEP_12 against TLR4/MD-2 resulted in six clusters containing 5.2% of the refined structures. The Haddock-scores of the best four clusters had overlapping error bars (one standard deviation) (Figure 9C). From this set, only cluster "5" contained solutions in which MEP_12 interacted exclusively with the co-receptor MD-2, while the others showed MEP_12 interacting with any of the two chains of TLR4 (Supplementary Figure 6C). Therefore, the solutions of cluster "5" were discarded.

Noteworthy, unlike TLR1/TLR2, the TL4/MD-2 complex is symmetric (Supplementary Figure 6D). In that sense, we noticed that the binding modes of clusters "3" are the same but reflected along the symmetry axis (Supplementary Figure 6D). Thus, these two can be seen as the same cluster. The best binding mode of the cluster "3-6" showed contacts with epitopes "0", "1", "3" and "10". It showed 3 hydrogen bonds and 2 salt bridges and had a binding energy of -9.9 kcal/mol (Figure 9D). The best binding mode of the cluster "1" showed contacts with epitopes "1", "5", "8 and "9" (Supplementary Table 9).

Simulating the immune response

The simulation of immunizing with the multi-epitope protein showed that the second and third doses generated significantly higher responses than the first one, as expected. HTL populations were higher in the second and third dose than in the first one, suggesting activation of the memory cells (Figure 10B). However, the CTL population was higher during the first dose, indicating an early immune response (Figure 10C). The B-cell subpopulations, including memory B-cells and Plasma B lymphocytes (PLB cells), showed considerable expansion after each dose reaching the highest peak at day 60 (Figure 10D). After the challenge, the response generated by B-cells was the highest in the simulation, indicating an appropriate production of antibodies (Figure 10E). Moreover, an early production of IgM was detected, which changed to IgG after the antigen administration (fourth response) (Figure 10F). Regarding the innate system, NK (natural killer) activity was found to be constant during the three doses and showed increased activity during the challenge (Figure 10G).

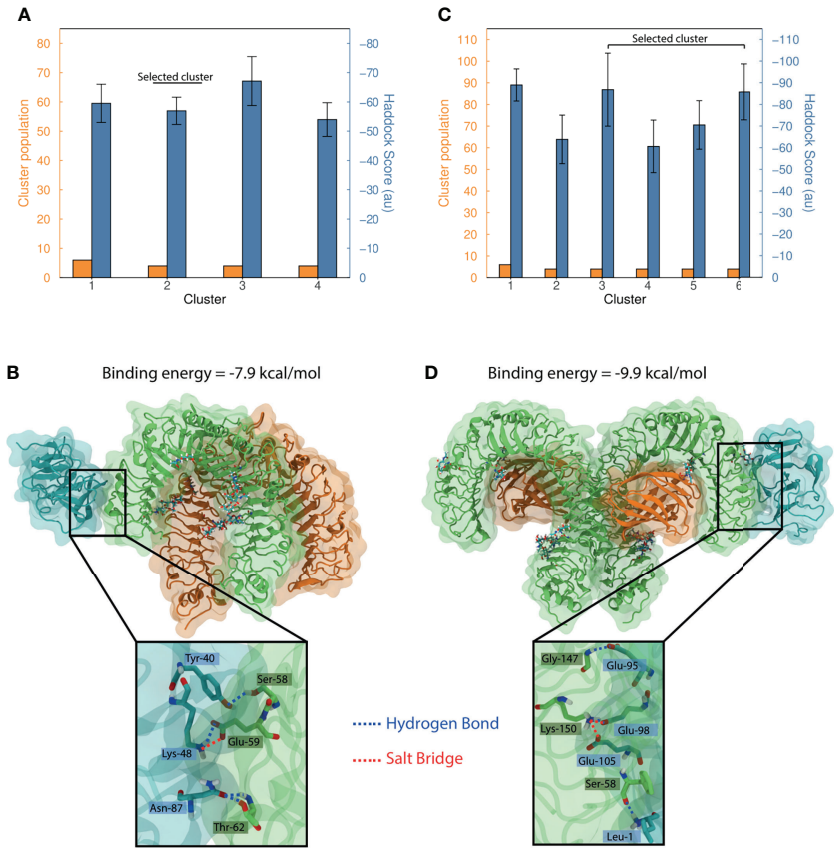


FIGURE 9
Protein docking of MEP_12 against TLR1/TLR2 and TLR4/MD-2. Bar plots summarizing the dockings **(A)** MEP_12-TLR1/TLR2 and **(C)** MEP_12-TLR4/MD2. The number of structures and the mean Haddock-score by cluster are shown in orange and blue, respectively. Whiskers in the mean Haddock-score bars represent one standard deviation. Structural representation of the most favorable binding mode of the cluster selected, of dockings **(B)** MEP_12-TLR1/TLR2, and **(D)** MEP_12-TLR4/MD2; showing the binding energy computed in PRODIGY. MEP_12 is colored in cyan; TLR1 and TLR4 in green; and TLR2 and MD2 in orange. Glycosylated residues and attached glycans (cyan) are shown as sticks, and non-carbon atoms are colored following the CPK convention. The inset plots show a closeup of the residues involved in polar interactions (as cyan and green sticks). The hydrogen bonds and salt bridges are represented by blue and red lines, respectively, connecting the interacting atoms which are labeled indicating amino acid and position.

Codon optimization of MEP_12 DNA sequence

The optimized sequence of the 450 nt cDNA vaccine construct had a Codon Adaptation Index (CAI) of 0.772, and a %GC of 48.9%. This sequence contains 111 restriction sites for 55 commercial enzymes (Supplementary Table 10).

Discussion

C. perfringens is one of the most common food-poisoning causing bacteria, causing a major impact on human health worldwide. It is also a reported biological warfare agent. These reasons make developing a vaccine an urgent matter. Most experimental studies have focused on *C. perfringens* toxins to find vaccines protective against gas gangrene. It has been shown that the alpha toxin protect against *C. perfringens* type A (67, 68). This toxin has been produced in endospores of recombinant *Bacillus subtilis* and tested as vaccine in mice, resulting in protection against gas gangrene (69). The β-toxin has also been evaluated as a possible vaccine against *C. perfringens* type C in piglets (70). However, none of them have been tested in humans. There is still no approved vaccine for humans nowadays (5).

While other studies had focused only on specific proteins of *C. perfringens* (71, 72), we consider that analyzing the whole proteome could reveal novel more immunogenic proteins. We followed three approaches to propose vaccine candidates: a whole-protein based vaccine, a protein-subunit based vaccine, and a multi-epitope protein vaccine. The first approach consist in identifying natural proteins of the pathogen enriched with both CTL and HTL epitopes to be used as recombinant vaccines. With the second approach, we further studied the sequence and

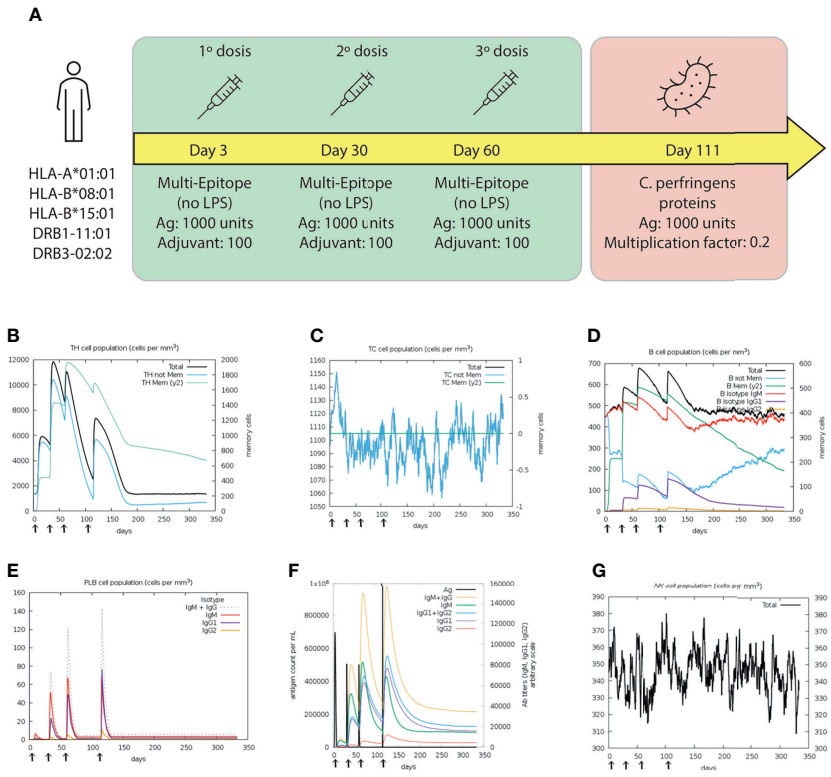


FIGURE 10
Results of the C-IMMSIM simulation for 350 days. **(A)** Schematic illustration of the vaccination trial, with three doses of the multi-epitope vaccine (green box) at days 3, 30, 60; and the challenge (red box) at day 111. Dynamics of **(B)** HTL and **(C)** CTL populations. Memory and not memory cells are represented with light-blue and green lines, respectively. **(D)** B cell populations, grouped by immunoglobulin isotype production. **(E)** Population of Plasma B lymphocytes producing IgM, IgG1 and IgG2. **(F)** Antigen concentration and relative antibodies responses. **(G)** Total population of NK cells. The first, second and third doses were inoculated.

structure of the main toxins, identifying regions exposed to the extra-cellular medium and rich in HTL epitopes that could generate immune responses against gas gangrene. Lastly, in the third approach, we identified the best CTL and HTL epitopes of the whole proteome and performed several immunoinformatics and structural analyses to assemble them into a multi-epitope construct, aiming to combat *C. perfringens* infection. The immunogenic proteins obtained by computational methods in these three approaches may represent good vaccine candidates, and they could be used as well for the development of immunodiagnostic tests (73, 74).

The immunoinformatic exploration of the whole proteome resulted in 429809 CTL- and 121450 HTL- predicted epitopes (Supplementary Table 3), which could be further used in both computational and experimental studies seeking a better understanding of the immunogenic characteristics of *C. perfringens* proteins, the developing of diagnostic tests, and the design of peptide-based vaccines and therapies. A strong positive correlation between the number of CTL- and HTL-epitopes predicted with the protein length was found, indicating that longer proteins contain a higher number of epitopes. This has been previously reported by an independent study in viral proteins, showing that protein length is positively correlated with its number of CD8+ T-cell epitopes within (75), supporting our observation. Another observation is that the 9-mer CTL-epitopes were more promiscuous, binding more HLA-I supertype alleles (Supplementary Figure 1B). These two observations might be correlated with the fact that most of the experimental data, and consequently most software training data, is based on 9-mer CTL-epitopes. A similar pattern was observed for 15 aa HTL-epitopes. These observations match with was is biologically expected, as they are the canonical epitope sizes which have been commonly observed in MHC-epitope experimental data (29, 76, 77).

The Collagenase A (UniProt ID: P43153), Exo-alpha-sialidase (UniProt ID: Q8XMY5), alpha n-acetylglucosaminidase (UniProt ID: Q8XM24) and hyaluronoglucosaminidase (UniProt ID: P26831) were identified as the top 4 proteins with more CTL- and HTL-epitopes. The Collagenase A, encoded by the gene colA, is an extracellular proteolytic enzyme that degrades extracellular matrix and plays a role in the pathogenesis of gangrene (78). This enzyme has hemorrhagic and dermonecrotic activities,

andintravenous inoculation of Collagenase A has shown to be lethal for mice (79). However, little is known about its role as immunogen. The Exo-alpha-sialidase, encoded by the gene nanJ, is the largest of the three sialidases produced by *C. perfringens*. It is involved in the intestinal virulence by increasing the binding affinity of the sialidases to their targets, enhancing pathogen adherence to intestinal cells. Sialidase inhibitors, such as Siastatin B or N-acetyl-2,3-dehydro-2-deoxyneuraminic acid (NADNA), have been tested *in vitro* and proposed as possible therapeutics against intestinal infection of *C. perfringens*; however, *in vivo* validation is still needed (80, 81). The hyaluronoglucosaminidase, encoded by the gene nagH, is a carbohydrate-active enzyme that acts on the connective tissue during the gas gangrene (82, 83). The alpha n-acetylglucosaminidase is an enzyme with strong preferences for carbohydrate motifs, found on the class III mucins within the gastric mucosa (81). Although there is no study testing alpha-acetylglucosaminidase, vaccination with beta-acetylglucosaminidase has shown to generate protection against necrotic enteritis in chickens (84). Similarly, a peptide-based vaccine comprising of several epitopes from these mucolytic enzymes produced reduction of lesions caused by necrotic enteritis in chicks. Altogether, these studies suggest that these four enzymes can be used as potential protein-based vaccines (85). The use of epitope-enriched proteins has been previously suggested as a strategy for the rational selection of immunogens, considering B-cell epitopes (86) and T-cell epitopes (73). Proteins with high epitope density are expected to maximize the probability of immune cells presentation and activation. Protein-based vaccines have been widely studied and used, with some advantages like having relatively low production costs, and not causing severe side effects unlike attenuated vaccines (87, 88).

Next, we studied the six most studied *C. perfringens* type A toxins in more detail. All of them, except for the beta2 toxin, showed at least 10 HTL-epitopes by HLA-II supertype allele (Figure 2C). These observations suggest that toxins are generally good candidates for protein-based vaccines, as shown in previous experimental studies (89–93). Our results support the idea that toxins could generate an appropriate humoral response to protect, mainly, against gangrene or histotoxic damage before infection (67, 94). Noteworthy, we show the relevance of not only predicting promiscuous epitopes but studying their location within the proteins as well. For instance, we found that the most promiscuous epitopes of enterotoxin D are in SH3B domains, which are correlated with promoting pathogen survival and invasion by binding to host receptors. This indicates that not all the extension, but specific regions of the protein are rich in epitopes. The use of non-toxic domains of toxins has shown to be advantageous for vaccine development as they present similar immunogenicity than the entire toxins without undesired toxicity (95–97). For example, the use of the non-toxic domain HC50 of the Botulinum neurotoxin type A induces a strong anti-HC50 IgG antibody response, neutralizing the circulating neurotoxin in mice (98). And, immunization with the non-toxic fragment of the C-domain of phospholipase C produces antibodies against this toxin, providing protection against gas gangrene in mice (5). Similarly, our analysis suggests that the non-toxic domain of Perfringolysin O is a promissory candidate for experimental testing as a protein-subunit based vaccine.

Numerous studies have used bioinformatics software to propose epitope-based vaccines against *C. perfringens*. A previous study predicted B-cell epitopes in the epsilon toxin of *C. perfringens* types B and D. However, they are not the main toxinotypes of *C. perfringens* affecting human health (99) and the epsilon toxin is not present in *C. perfringens* type A. Another study predicted 15 unique epitopes in the toxin NetF using mouse rather than human MHC alleles. Although the authors suggest that NetF could be a good vaccine candidate and the epitopes found can be used in multi-epitope vaccines, this needs to be evaluated in humans (100). Furthermore, the NetF protein has been associated with gastroenteritis and enterocolitis in canine and foals (100), but its role in humans is not well studied. Another study has predicted T- and B-cell epitopes in the fructose 1,6-bisphosphate aldolase (101). Nonetheless, they did not design a multi-epitope construct, which might be more effective than single-epitope vaccines (18, 102). Other bioinformatic studies of *C. perfringens* aimed to find only candidate epitopes in toxins or few proteins (71, 103). However, exploring the whole proteome, as done in the present study, allows the computational identification of a broader set of potential epitopes that may trigger better immune responses, as seen in other pathogens like SARS-CoV-2 (104).

A rational workflow was elaborated to design a multi-epitope vaccine, consisting of (1) using different immunoinformatics tools to predict epitopes from the whole proteome of *C. perfringens* (2), considering the epitope location in the structure of their native proteins, and (3) merging and assembling nested epitopes in potential constructs, and (4) evaluating the structure and dynamics of the constructs by MD. Of the programs available for T-cell epitope prediction, we opted to use NetMHCpan and MHCFlurry to predict CTL-epitopes, and NetMHCIIpan for HTL-epitopes, as they have shown the best performance against competitors (27, 28). Proteasome cleavage and TAP transport are also important processes of the intracellular presentation pathway. Predictors of proteasome developed, being the most popular netChop and ProteaSMM. However, benchmarking studies have shown that these predictors still need to improve their sensitivity and specificity (105). Moreover, it has been reported that current cleavage predictions based on *in vitro* data do not correlate with *in vivo* data (106). Therefore, we decided not to discard constructs based on these predictors. TAP transport predictors face the same limitations, with very few predictors available and the lack of unbiased benchmarking studies (107).

Targeting HLA supertype alleles maximize the potential usability of our multiepitope in different populations. This choice

was a tradeoff between wider usability and higher local specificity, and we decided to prioritize the first one as currently *C. perfringens* infection is a widely spread pathogen without universal vaccine. Higher local specificity would be more relevant in other scenarios, such as if broad-range vaccines become available, or if the disease becomes endemic, or if there is a high incidence of cases in a specific geographic location, among others. This will imply a change in the HLA allele selection strategy, giving more importance to certain HLA alleles abundant in the specific geographic areas affected by the disease.

Keeping the synthetic protein small is important to reduce synthesis and production costs (i.e. facilitate its purification in inclusion bodies), as well as to prevent toxicity in the organism used for production (98, 108). Using nested epitopes allows to maximize the number of epitopes in the multi-epitope construct without increasing the construct length. Recent studies have used nested epitopes to construct multi-epitope proteins, but experimental testing is still needed to verify the advantage provided by this strategy (104, 109). From the set of 112714 nested epitopes, several filters were applied to select the best candidates. One of these filters is to prioritize epitope promiscuity, which has been associated with contributing to epitope immunodominance, as promiscuous epitopes are recognized by multiple HLA alleles (110). The use of promiscuous epitopes allows to cover a larger number of HLA-alleles (i.e. a larger proportion of the target population) without increasing the number of epitopes in the vaccine construct (111). We also filtered by epitope conservation among all the reported protein variants of *C. perfringens*. This may result in covering a broader range of current, and potentially future, pathogen variants. Thus, a high epitope conservation could lead to a better protection as the immune response tends to focus on conserved epitopes when individuals are exposed to different strains (112, 113). In our construct, the conservation analysis of the overlapped nested epitopes showed 100% conservation. Thus, even though *C. perfringens* type A is the most common cause of gas gangrene, our construct may confer extended protection against other toxinotypes as well. Our construct of 150 aa is made of 9 overlapped HTL-CTL epitopes, comprising 24 HTL-epitopes containing 34 CTL-epitopes. This suggests a better cellular immune response than previous multiepitope constructs, which were longer and had less epitopes in its sequence (114). Whilst the study of Aldakheel *et al.* (114) attempted to target all HLA-I and -II alleles, we opted to focus on the HLA supertype alleles. This allowed us to need less epitopes in our design to match all the target alleles. Additionally, having less HLAs to target allows epitopes with better immunogenicity to be selected from the prediction. Thus, we successfully covered all the HLA supertype alleles, while the previous study just covered 10 of the 12 HLA-I and 5 of the 8 HLA-II supertypes alleles. Additionally, our design is about 1/3 in length (150 versus 415 aa), due to our strategy of using nested HTL-CTL epitopes in extracellularly exposed regions.

The 9 overlapped nested epitopes in our construct belong to the FTsX domain-containing protein (UniProt ID: Q8XM39), the cardiolipin synthase (UniProt ID: P0C2E2), the Polysaccharide synthase 2 domain-containing protein (UniProt ID: Q8XN75), the probable hemolysin-related protein (UniProt ID: Q8XPD3), the TraG-D C domain-containing protein (UniProt ID: Q93M96), the spore germination protein KA (UniProt ID: Q8XMP0) and CPE0011 (UniProt ID: Q8XPF2). Our prediction indicates that all of them are membrane proteins with a high content of HTL- and CTL-epitopes. Moreover, these proteins present functions related with the cell membrane structure (Cardiolipin synthase), survival (Spore germination protein KA) and host colonization (Hemolysin-related protein) (115–117). External regions of transmembrane proteins involved in the infective process are frequently considered vaccine targets, as antibodies can efficiently neutralize them (118).

Assembling epitopes to make a new protein can be performed with or without linkers, or adding adjuvant sequences (119). Multiple linker and adjuvant sequences have been reported in the literature (120–122). Linkers are used to reduce the occurrence of neoepitopes (123), with the downside of increasing the construct length and the cost of protein synthesis. There are very few software to optimize the use of linkers (112, 124, 125), and this approach is frequently used because finding an appropriate epitope sorting without linkers is time-consuming and computationally demanding. Here, we used a novel algorithm and software based on graph theory (EpiSorter), assembling all the epitopes without the need of linkers and without unwanted neoepitopes shared with human proteins, avoiding autoimmune or tolerance responses. Thus, we believe that using linkers can be avoided and, instead, we recommend prioritizing the number and promiscuity of the epitopes selected.

Several immunoinformatics studies have reported structural models of their multi-epitope constructs using only one or two software, complemented with refinement steps (126–128). Nonetheless, modeling novel proteins is a difficult task as they usually do not have close homologues. The situation becomes even more complex if the novel protein to model consists of highly flexible linear epitopes. To tackle this problem, we employed the state-of-the-art protein-structure prediction software AlphaFold2 (47). Nowadays, this is the first time this software has been used to predict the 3D structure of multi-epitope constructs.

The high proportion of residues of the multi-epitope structures with low pLDDT suggests that the models obtained by AlphaFold were not completely reliable. However, it also says that the multi-epitope protein is highly disordered and flexible, characteristics that may favor its binding to immune proteins. Disordered regions in proteins are characterized by a lack of a stable tertiary structure and high flexibility (129). Moreover, it has been pinpointed that flexibility in protein antigens positively affects their binding affinity (130).

The multi-epitope construct selected did not present Ramachandran outliers and shows an ERRAT score

comparable to scores of high-resolution structures (131). Moreover, these quality statistics are better than the ones obtained in various similar studies designing multi-epitope vaccines (119, 127, 132–135). This reaffirms the quality and consistency of our structural approach, although better modeling software is still needed for multi-epitope proteins.

Docking was performed against TLR1/TLR2 and TLR4/MD2, as they have been widely reported to be important in the innate defense against *C. perfringens* infection in chickens and mice (136), although little is known in humans. The lipopeptide and lipopolysaccharide binding sites of TLR1/TLR2 and TLR4/MD-2, respectively, have been structurally characterized (137, 138). As there is neither prior experimental information about the binding sites for MEPs, nor which epitope of our novel MEP might interact with the TLRs, we considered blind docking the most appropriate approach. The docking simulation showed binding between our MEP_12 and the TLR1/TLR2, suggesting an innate immune response in addition to the adaptive immune response generated by the nested epitopes. The docking analyses also showed that epitopes "3" and "10" interact with TLR1/TLR2. These epitopes are relatively rigid, which may result in a more favorable binding as shown in previous studies. Epitopes tend to be more rigid than the rest of the protein (31). And, from a thermodynamic point of view, rigid surfaces have less entropic penalty when interacting with other proteins, resulting in tighter bindings.

The binding region of MEP_12 in TLR4/MD-2 is similar to the one observed in a previous study, where a SARS-CoV-2 candidate vaccine was docked against this receptor (119), even though different protocols were followed. This suggest that this region of TLR4/MD-2 is where antigenic proteins bind. However, experimental studies are needed to validate this observation. Noteworthy, the MEP_12 is smaller and establishes fewer interactions with TLR4/MD-2 than the SARS-CoV-2 candidate vaccine but achieves a similar binding energy, differing in just in 0.4 kcal/mol. This indicates that our design is a more efficient TLR4/MD-2 binder and correlates with the finding that MEP_12 is flexible and disordered. These are two desired characteristics, as they make antigens more efficient at binding immune proteins (33).

Immunization with MEP_12 was simulated, showing that our vaccine candidate can elicit immune responses to clear the antigen on secondary exposure. The challenge with *C. perfringens* proteins after three vaccine injections induced higher levels of IgG than IgM. IgM is the principal isotype in the first response, while IgG is predominant in secondary responses, representing specific pathogen recognition (139). Also, the increased production of immunoglobulins by plasma B cells indicated that memory of *C. perfringens* proteins is present in the immune system (140). Finally, a gene encoding MEP_12 was designed optimizing its sequence for expression and production in *E. coli*, as this organism has been widely used to produce recombinant vaccines (141–143). Additionally, the design can be adapted to other organisms following the methodology we described.

In summary, in the current study, we have performed a thorough immunoinformatic exploration of the whole proteome to generate vaccine candidates against *C. perfringens*. Three approaches were followed to identify (1) the most immunogenic proteins (2), immunogenic non-toxin domains of toxins, and (3) the design a novel protein with the best HTL-CTL nested epitopes, expected to trigger both adaptive and cellular immune responses. These resulted in promising candidates for further experimental *in vitro* and *in vivo* studies. These candidates may help in the prevention of necrotic enteritis, as well as other human diseases caused by *C. perfringens*.

Data availability statement

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

Author contributions

Conceptualization, LFS and DR. Methodology, LFS, DR, GJ-A, and DR. Software, LFS, DR, and GJ-A. Validation, LS and DR. Formal Analysis, LFS and DR. Investigation, LFS, AR, GJ-A, and DR. Resources, DR. Data curation, DR. Writing – original draft preparation, LFS, DR, GJ-A, and DR. Writing-review and editing, LFS, AR, and DR. Visualization, LFS and GJ-A. Supervision, DR. Project administration, DR, RLL, YS, and CO-R. Funding acquisition, DR, RLL, YS, and CO-R. All authors contributed to the article and approved the submitted version.

Funding

The publication of this work was funded by the grant N° 2264962 SNIP N°317435 "Creación del Centro de Promoción de la Investigación y Transferencia Tecnológica" of the University "Toribio Rodriguez de Mendoza de Amazonas".

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.942907/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Density and promiscuity of the CTL- and HTL-epitopes. **(A)** Epitope density in proteins by epitope length (CTL-epitopes: 8, 9 and 10 aa, HTL-epitopes: 15 aa). **(B)** The bar plot represents in the X-axis the number of epitopes identified, for each epitope length (Y-axis). The color indicates how many HLA supertype alleles each epitope binds. In the bar plots **(C)** (for CTL-epitopes) and **D** (for HTL-epitopes), the Y-axis represents the number of epitopes that can bind a certain number of HLA supertype alleles (X-axis). Panels **(E, F)** show the distribution of the number and density, respectively, of HTL-epitopes in the proteins of each of the levels of evidence indicated in the X-axis.

SUPPLEMENTARY FIGURE 2
Conservation of the nested candidate epitopes across C. perfringens toxinotypes. Each subfigure shows the alignment of the nested epitope, the epitope ID and the protein name.

SUPPLEMENTARY FIGURE 3
(A) Mean pLDDT of the 22 multi-epitope (MEP) candidates. **(B)** Structural alignment between the centroid of the most populated cluster of MEP_12, before (orange) and after (cyan) refinement.

SUPPLEMENTARY FIGURE 4
Quality assessment of the centroid of the most populated cluster of MEP_12 before refinement. **(A)** Ramachandran plot for the general case. The percentage of residues in preferred (light blue) and allowed (blue) regions are shown. **(B)** Scatterplot of the Z-scores of the centroid (black dot) and protein structures with experimental evidence (NMR: blue, X-ray crystallography: light blue). **(C)** ERRAT plot of the centroid. Bars represent the error value (white: error < 95%, yellow: 95% error < 99%) of a nine-residue sliding window. The overall quality factor indicates the percentage of protein residues with error values lower than 95%.

SUPPLEMENTARY FIGURE 5
SASA values per epitope. Values over the last 500 ns of convergent simulation are shown for epitopes "0" to "10" (excluding "4") in panels **A** to **I**, respectively. The SASA of all epitopes shown but "5" and "9" achieved convergence. Epitope "8" only achieved convergence during the last 400 ns (enclosed in a red box, close-up in inset plot). So, only this region was considered for its mean SASA computations.

SUPPLEMENTARY FIGURE 6
Comparison of all plausible clusters of docking of MEP_12 against TLR1/TLR2 and TLR4/MD-2. The best binding modes of clusters having Haddock-scores with overlapping error bars (standard deviations) are shown for the dockings **(A)** MEP_12-TLR1/TLR2, and **(C)** MEP_12-TLR4/MD2. TLR1 and TLR2 chains are colored in green and orange, respectively. Attached glycans are shown as sticks in cyan. In **(A)**, the best binding modes of MEP_12 from clusters "1", "2", "3", and "4" are colored in red, cyan, gray, and blue, respectively. In **(C)**, the best binding modes of MEP_12 from clusters "1", "3", "5", and "6" are colored in gray, blue, red, and cyan, respectively. **(B)** Enlarged view of the best binding mode of cluster "1" (red cartoon with transparent surface) in the docking MEP_12-TLR1/TLR2. The N-terminal residues missing in TLR1/TLR2 structure (in gray) were completed by structural alignment to a TLR2 model from the AlphaFold Protein 1300 Database. These interact with MEP_12 cluster "1". **(D)** Best binding modes from cluster "3" and "6" from the docking MEP_12-TLR4/MD2. The black line separating the two sides of TLR4/MD-2 represents a symmetry axis. The binding modes of left side become equivalent to the right side after a 180° right-handed rotation, as shown in **(E)**.

References

1. Hassan KA, Elbourne LDH, Tetu SG, Melville SB, Rood JI, Paulsen IT. Genomic analyses of clostridium perfringens isolates from five toxinotypes. *Res Microbiol* (2015) 166(4):255–63. doi: doi: 10.1016/j.resmic.2014.10.003

2. Grass JE, Gould LH, Mahon BE. Epidemiology of foodborne disease outbreaks caused by clostridium perfringens, United States, 1998–2010. *Foodborne Pathog Dis* (2013) 10(2):131–6. doi: 10.1089/fpd.2012.1316

3. Janik E, Ceremuga M, Saluk-Bijak J, Bijak M. Biological toxins as the potential tools for bioterrorism. *Int J Mol Sci* (2019) 20(5):E1181. doi: 10.3390/ijms20051181

4. Buboltz JB, Murphy-Lavoie HM. *Gas gangrene*. Treasure Island (FL: StatPearls Publishing (2022). Available at: http://www.ncbi.nlm.nih.gov/books/NBK537030/.

5. Titball RW. Clostridium perfringens vaccines. *Vaccine* (2009) 27 Suppl 4: D44–47. doi: 10.1016/j.vaccine.2009.07.047

6. Barras V, Greub G. History of biological warfare and bioterrorism. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis* (2014) 20(6):497–502. doi: 10.1111/1469-0691.12706

7. Sim K, Shaw AG, Randell P, Cox MJ, McClure ZE, Li MS, et al. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin Infect Dis Off Publ Infect Dis Soc Am* (2015) 60(3):389–97. doi: 10.1093/cid/ciu822

8. Heida FH, van Zoonen AGJF, Hulscher JBF, te Kiefte BJC, Wessels R, Kooi EMW, et al. A necrotizing enterocolitis-associated gut microbiota is present in the meconium: Results of a prospective study. *Clin Infect Dis* (2016) 62(7):863–70. doi: 10.1093/cid/ciw016

9. Uzal FA, Freedman JC, Shrestha A, Theoret JR, Garcia J, Awad MM, et al. Towards an understanding of the role of clostridium perfringens toxins in human and animal disease. *Future Microbiol* (2014) 9(3):361–77. doi: 10.2217/fmb.13.168

10. Navarro MA, McClane BA, Uzal FA. Mechanisms of action and cell death associated with clostridium perfringens toxins. *Toxins* (2018) 10(5):212. doi: 10.3390/toxins10050212

11. Kammerl IE, Meiners S. Proteasome function shapes innate and adaptive immune responses. *Am J Physiol Lung Cell Mol Physiol* (2016) 311(2):L328–336. doi: 10.1152/ajplung.00156.2016

12. Yao Y, Yang H, Shi L, Liu S, Li C, Chen J, et al. HLA class II genes HLA-DRB1, HLA-DPB1, and HLA-DQB1 are associated with the antibody response to inactivated Japanese encephalitis vaccine. *Front Immunol* (2019) 10. doi: 10.3389/fimmu.2019.00428

13. Moyle PM, Toth I. Modern subunit vaccines: Development, components, and research opportunities. *ChemMedChem* (2013) 8(3):360–76. doi: 10.1002/cmdc.201200487

14. Uzal FA, Wong JP, Kelly WR, Priest J. Antibody response in goats vaccinated with liposome-adjuvanted clostridium perfringens type d epsilon toxoid. *Vet Res Commun* (1999) 23(3):143–50. doi: 10.1023/A:1006206216220

15. Tian JH, Patel N, Haupt R, Zhou H, Weston S, Hammond H, et al. SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 immunogenicity in baboons and protection in mice. *Nat Commun* (2021) 12(1):372. doi: 10.1038/s41467-020-20653-8

16. Heath PT, Galiza EP, Baxter DN, Boffito M, Browne D, Burns F, et al. Safety and efficacy of NVX-CoV2373 covid-19 vaccine. *N Engl J Med* (2021) 385 (13):1172–83. doi: 10.1056/NEJMoa2107659

17. Koff WC, Schenkelberg T. The future of vaccine development. *Vaccine* (2020) 38(28):4485–6. doi: 10.1016/j.vaccine.2019.07.101

18. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* (2018) 15(2):182–4. doi: 10.1038/cmi.2017.92

19. Lennerz V, Gross S, Gallerani E, Sessa C, Mach N, Boehm S, et al. Immunologic response to the survivin-derived multi-epitope vaccine EMD640744 in patients with advanced solid tumors. *Cancer Immunol Immunother* (2014) 63(4):381–94. doi: 10.1007/s00262-013-1516-5

20. Doehn C, Esser N, Pauels HG, Kießig ST, Stelljes M, Grossmann A, et al. Mode-of-Action, efficacy, and safety of a homologous multi-epitope vaccine in a murine model for adjuvant treatment of renal cell carcinoma. *Eur Urol* (2009) 56 (1):123–33. doi: 10.1016/j.eururo.2008.05.034

21. Romeli S, Hassan SS, Yap WB. Multi-epitope peptide-based and vaccinia-based universal influenza vaccine candidates subjected to clinical trials. *Malays J Med Sci MJMS* (2020) 27(2):10–20. doi: 10.21315/mjms2020.27.2.2

22. Majidiani H, Dalimi A, Ghaffarifar F, Pirestani M. Multi-epitope vaccine expressed in leishmania tarentolae confers protective immunity to toxoplasma gondii in BALB/c mice. *Microb Pathog* (2021) 155:104925. doi: 10.1016/j.micpath.2021.104925

23. Romano P, Giugno R, Pulvirenti A. Tools and collaborative environments for bioinformatics research. *Brief Bioinform* (2011) 12(6):549–61. doi: 10.1093/bib/bbr055

24. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* (2017) 8:278. doi: 10.3389/fimmu.2017.00278

25. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res* (2019) 47(W1):W502–6. doi: 10.1093/nar/gkz452

26. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J Proteome Res* (2020) 19(6):2304–15. doi: 10.1021/acs.jproteome.9b00874

27. 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(1):449–54. doi: 10.1093/nar/gkaa379

28. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst* (2020) 11(1):42–48.e7. doi: 10.1016/j.cels.2020.06.010

29. Zhao W, Sher X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PloS Comput Biol* (2018) 14(11):e1006457. doi: 10.1371/journal.pcbi.1006457

30. Soto LF, Requena D, Bass JIF. Epitope-evaluator: an interactive web application to study predicted T-cell epitopes. *PLoS One* (2022)17(8):e0273577. doi: 10.1371/journal.pone.0273577

31. Kim DG, Choi Y, Kim HS. Epitopes of protein binders are related to the structural flexibility of a target protein surface. *J Chem Inf Model* (2021) 61 (4):2099–107. doi: 10.1021/acs.jcim.0c01397

32. Alaofi AL. Probing the flexibility of zika virus envelope protein DIII epitopes using molecular dynamics simulations. *Mol Simul* (2020) 46(7):541–7. doi: 10.1080/08927022.2020.1738424

33. MacRaild CA, Richards JS, Anders RF, Norton RS. Antibody recognition of disordered antigens. *Struct Lond Engl* (2016) 24(1):148–57. 1993. doi: 10.1016/j.str.2015.10.028

34. Yuan X, Qu Z, Wu X, Wang Y, Liu L, Wei F, et al. Molecular modeling and epitopes mapping of human adenovirus type 3 hexon protein. *Vaccine* (2009) 27 (37):5103–10. doi: 10.1016/j.vaccine.2009.06.041

35. Verherstraeten S, Goossens E, Valgaeren B, Pardon B, Timbermont L, Haesebrouck F, et al. Perfringolysin O: The underrated clostridium perfringens toxin? *Toxins* (2015) 7(5):1702–21. doi: 10.3390/toxins7051702

36. Goossens E, Verherstraeten S, Valgaeren BR, Pardon B, Timbermont L, Schauvliege S, et al. Toxin-neutralizing antibodies protect against clostridium perfringens-induced necrosis in an intestinal loop model for bovine necrohemorrhagic enteritis. *BMC Vet Res* (2016) 12(1):101. doi: 10.1186/s12917-016-0730-8

37. Gonzalez-Galarza FF, McCabe A, Melo dos Santos EJ, Jones AR, Middleton D. A snapshot of human leukocyte antigen (HLA) diversity using data from the allele frequency net database. *Hum Immunol* (2021) 82(7):496–504. doi: 10.1016/j.humimm.2020.10.004

38. Requena D, Médico A, Chacón RD, Ramírez M, Marín-Sánchez O. Identification of novel candidate epitopes on SARS-CoV-2 proteins for South America: A review of HLA frequencies by country. *Front Immunol* (2020) 11:2008. doi: 10.3389/fimmu.2020.02008

39. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: A revised and updated classification. *BMC Immunol* (2008) 9:1. doi: 10.1186/1471-2172-9-1

40. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* (2011) 63(6):325–35. doi: 10.1007/s00251-011-0513-0

41. Berglund L, Andrade J, Odeberg J, Uhlén M. The epitope space of the human proteome. *Protein Sci Publ Protein Soc* (2008) 17(4):606–13. doi: 10.1110/ps.073347208

42. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* (1998) 6:175–82.

43. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes11Edited by f. Cohen. *J Mol Biol* (2001) 305(3):567–80. doi: 10.1006/jmbi.2000.4315

44. Finlay BB, McFadden G. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell* (2006) 124(4):767–82. doi: 10.1016/j.cell.2006.01.034

45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. *Bioinforma Oxf Engl* (2007) 23(21):2947–8. doi: 10.1093/bioinformatics/btm404

46. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol Clifton NJ* (1999) 112:531–52. doi: 10.1385/1-59259-584-7:531

47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) 596 (7873):583–9. doi: 10.1038/s41586-021-03819-2

48. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* (2007) 35(suppl_2):W522–5. doi: 10.1093/nar/gkm276

49. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* (2020) 153(4):044130. doi: 10.1063/5.0014475

50. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat Methods* (2017) 14(1):71–3. doi: 10.1038/nmeth.4067

51. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. *J Comput Chem* (2007) 28(16):2618–40. doi: 10.1002/jcc.20829

52. Hadden JA, Perilla JR. Molecular dynamics simulations of protein-drug complexes: A computational protocol for investigating the interactions of small-molecule therapeutics with biological targets and biosensors. *Methods Mol Biol Clifton NJ* (2018) 1762:245–70. doi: 10.1007/978-1-4939-7756-7_13

53. Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, et al. Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* (2011) 32(6):1183. doi: 10.1002/jcc.21688

54. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: When simulation meets experiment. *Angew Chem Int Ed* (1999) 38 (1–2):236–40. doi: 10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M

55. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* (2011) 101(10):2525–34. doi: 10.1016/j.bpj.2011.10.024

56. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* (2007) 35:W407–10. doi: 10.1093/nar/gkm290

57. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci Publ Protein Soc* (1993) 2(9):1511–9. doi: 10.1002/pro.5560020916

58. Schrödinger L, DeLano WM. PyMOL [Internet]. (2020). Available from: http://www.pymol.org/pymol

59. Molecular modelling: Principles and applications. Available at: https://www.pearson.com/content/one-dot-com/one-dot-com/us/en/higher-education/program.html.

60. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* (2003) 125(7):1731–7. doi: 10.1021/ja026939x

61. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, et al. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput* (2013) 9(1):461–9. doi: 10.1021/ct300857j

62. Weber ANR, Morse MA, Gay NJ. Four n-linked glycosylation sites in human toll-like receptor 2 cooperate to direct efficient biosynthesis and secretion *. *J Biol Chem* (2004) 279(33):34589–94. doi: 10.1074/jbc.M403830200

63. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci Publ Protein Soc* (2018) 27 (1):129–34. doi: 10.1002/pro.3289

64. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* (2016) 32(23):3676–8. doi: 10.1093/bioinformatics/btw514

65. Castiglione F, Mantile F, De Berardinis P, Prisco A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput Math Methods Med* (2012) 2012:842329. doi: 10.1155/2012/842329

66. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature* (2021) 596(7873):590–6. doi: 10.1038/s41586-021-03828-1

67. Williamson ED, Titball RW. A genetically engineered vaccine against the alpha-toxin of clostridium perfringens protects mice against experimental gas gangrene. *Vaccine* (1993) 11(12):1253–8. doi: 10.1016/0264-410X(93)90051-X

68. Nagahama M. Vaccines against clostridium perfringens alpha-toxin. *Curr Pharm Biotechnol* (2013) 14(10):913–7. doi: 10.2174/13892010146661312261243 48

69. Hoang TH, Hong HA, Clark GC, Titball RW, Cutting SM. Recombinant bacillus subtilis expressing the clostridium perfringens alpha toxoid is a candidate orally delivered vaccine against necrotic enteritis. *Infect Immun* (2008) 76 (11):5257–65. doi: 10.1128/IAI.00686-08

70. Springer S, Selbitz H -J. The control of necrotic enteritis in sucking piglets by means of a clostridium perfringens toxoid vaccine. *FEMS Immunol Med Microbiol* (1999) 24(3):333–6. doi: 10.1111/j.1574-695X.1999.tb01302.x

71. Wang Yh. Bioinformatics analysis of NetF proteins for designing a multi-epitope vaccine against clostridium perfringens infection. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* (2020) 85:104461. doi: 10.1016/j.meegid.2020.104461

72. Elhag M, Abubaker M, Ahmad NM, Haroon EM, Alaagib RM, Albagi SOA, et al. Immunoinformatics prediction of epitope based peptide vaccine against listeria monocytogenes fructose bisphosphate aldolase protein. *bioRxiv* (2019), 649111. doi: 10.1101/649111

73. Morales Ruiz S, Bendezu J, Choque Guevara R, Montesinos R, Requena D, Choque Moreau L, et al. Development of a lateral flow test for the rapid detection of avibacterium paragallinarum in chickens suspected of having infectious coryza. *BMC Vet Res* (2018) 14(1):411. doi: 10.1186/s12917-018-1729-0

74. Valdivia-Olarte H, Requena D, Ramirez M, Saravia LE, Izquierdo R, Falconi-Agapito F, et al. Design of a predicted MHC restricted short peptide immunodiagnostic and vaccine candidate for fowl adenovirus c in chicken infection. *Bioinformation* (2015) 11(10):460–5. doi: 10.6026/97320630011460

75. Agranovich A, Maman Y, Louzoun Y. Viral proteome size and CD8+ T cell epitope density are correlated: The effect of complexity on selection. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* (2013) 20:71–7. doi: 10.1016/j.meegid.2013.07.030

76. Trolle T, McMurtrey CP, Sidney J, Bardet W, Osborn SC, Kaever T, et al. The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. *J Immunol Baltim Md 1950* (2016) 196(4):1480–7. doi: 10.4049/jimmunol.1501721

77. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol* (2013) 191(12):5831–9. doi: 10.4049/jimmunol.1302101

78. Van Damme L, Cox N, Callens C, Dargatz M, Flügel M, Hark S, et al. Protein truncating variants of colA in clostridium perfringens type G strains. *Front Cell Infect Microbiol* (2021) 11:645248. doi: 10.3389/fcimb.2021.645248

79. Kameyama S, Akama K. Purification and some properties of kappa toxin of clostridium perfringens. *Jpn J Med Sci Biol* (1971) 24(1):9–23. doi: 10.7883/yoken1952.24.9

80. Li J, Uzal FA, McClane BA. Clostridium perfringens sialidases: Potential contributors to intestinal pathogenesis and therapeutic targets. *Toxins* (2016) 8(11): E341. doi: 10.3390/toxins8110341

81. Boraston AB, Ficko-Blean E, Healey M. Carbohydrate recognition by a large sialidase toxin from clostridium perfringens. *Biochemistry* (2007) 46(40):11352–60. doi: 10.1021/bi701317g

82. Ficko-Blean E, Boraston AB. N-acetylglucosamine recognition by a family 32 carbohydrate-binding module from clostridium perfringens NagH. *J Mol Biol* (2009) 390(2):208–20. doi: 10.1016/j.jmb.2009.04.066

83. Canard B, Garnier T, Saint-Joanis B, Cole ST. Molecular genetic analysis of the nagH gene encoding a hyaluronidase of clostridium perfringens. *Mol Gen Genet MGG* (1994) 243(2):215–24. doi: 10.1007/BF00280319

84. Jiang Y, Kulkarni RR, Parreira VR, Prescott JF. Immunization of broiler chickens against clostridium perfringens-induced necrotic enteritis using purified recombinant immunogenic proteins. *Avian Dis* (2009) 53(3):409–15. doi: 10.1637/8656-021109-Reg.1

85. Duff AF, Vuong CN, Searer KL, Briggs WN, Wilson KM, Hargis BM, et al. Preliminary studies on development of a novel subunit vaccine targeting clostridium perfringens mucolytic enzymes for the control of necrotic enteritis in broilers. *Poult Sci* (2019) 98(12):6319–25. doi: 10.3382/ps/pez448

86. Caradonna TM, Schmidt AG. Protein engineering strategies for rational immunogen design. *NPJ Vaccines* (2021) 6(1):154. doi: 10.1038/s41541-021-00417-1

87. Merlin M, Gecchele E, Capaldi S, Pezzotti M, Avesani L. Comparative evaluation of recombinant protein production in different biofactories: The green perspective. *BioMed Res Int* (2014) 2014:136419. doi: 10.1155/2014/136419

88. Pollet J, Chen WH, Strych U. Recombinant protein vaccines, a proven approach against coronavirus pandemics. *Adv Drug Deliv Rev* (2021) 170:71–82. doi: 10.1016/j.addr.2021.01.001

89. Gao X, Ma Y, Wang Z, Bai J, Jia S, Feng B, et al. Oral immunization of mice with a probiotic lactobacillus casei constitutively expressing the α-toxoid induces protective immunity against clostridium perfringens α-toxin. *Virulence* (2019) 10 (1):166–79. doi: 10.1080/21505594.2019.1582975

90. Pilehchian Langroudi R, Shamsara M, Aghaiypour K. Expression of clostridium perfringens epsilon-beta fusion toxin gene in e. coli and its immunologic studies in mouse. *Vaccine* (2013) 31(32):3295–9. doi: 10.1016/j.vaccine.2013.04.061

91. Cooper KK, Trinh HT, Songer JG. Immunization with recombinant alpha toxin partially protects broiler chicks against experimental challenge with clostridium perfringens. *Vet Microbiol* (2009) 133(1):92–7. doi: 10.1016/j.vetmic.2008.06.001

92. Salvarani FM, Conceição FR, Cunha CEP, Moreira GMSG, Pires PS, Silva ROS, et al. Vaccination with recombinant clostridium perfringens toxoids α and β promotes elevated antepartum and passive humoral immunity in swine. *Vaccine* (2013) 31(38):4152–5. doi: 10.1016/j.vaccine.2013.06.094

93. Zichel R, Mimran A, Keren A, Barnea A, Steinberger-Levy I, Marcus D, et al. Efficacy of a potential trivalent vaccine based on hc fragments of botulinum toxins a, b, and e produced in a cell-free expression system. *Clin Vaccine Immunol CVI* (2010) 17(5):784–92. doi: 10.1128/CVI.00496-09

94. Flores-Díaz M, Alape-Girón A. Role of clostridium perfringens phospholipase c in the pathogenesis of gas gangrene. *Toxicon* (2003) 42(8):979–86. doi: 10.1016/j.toxicon.2003.11.013

95. Verherstraeten S, Goossens E, Valgaeren B, Pardon B, Timbermont L, Haesebrouck F, et al. Non-toxic perfringolysin O and α-toxin derivatives as potential vaccine candidates against bovine necrohaemorrhagic enteritis. *Vet J Lond Engl 1997* (2016) 217:89–94. doi: 10.1016/j.tvjl.2016.09.008

96. Liao CM, Huang C, Hsuan SL, Chen ZW, Lee WC, Liu CI, et al. Immunogenicity and efficacy of three recombinant subunit pasteurella multocida toxin vaccines against progressive atrophic rhinitis in pigs. *Vaccine* (2006) 24 (1):27–35. doi: 10.1016/j.vaccine.2005.07.079

97. Tinker JK, Yan J, Knippel RJ, Panayiotou P, Cornell KA. Immunogenicity of a West Nile virus DIII-cholera toxin A2/B chimera after intranasal delivery. *Toxins* (2014) 6(4):1397–418. doi: 10.3390/toxins6041397

98. Mustafa W, Al-Saleem FH, Nasser Z, Olson RM, Mattis JA, Simpson LL, et al. Immunization of mice with the non-toxic HC50 domain of botulinum neurotoxin presented by rabies virus particles induces a strong immune response affording protection against high-dose botulinum neurotoxin challenge. *Vaccine* (2011) 29(28):4638–45. doi: 10.1016/j.vaccine.2011.04.045

99. Kaushik H, Deshmukh S, Mathur DD, Tiwari A, Garg LC. Recombinant expression of *in silico* identified bcell epitope of epsilon toxin of clostridium perfringens in translational fusion with a carrier protein. *Bioinformation* (2013) 9 (12):617. doi: 10.6026/97320630009617

100. Mehdizadeh Gohari I, Parreira VR, Nowell VJ, Nicholson VM, Oliphant K, Prescott JF. A novel pore-forming toxin in type a clostridium perfringens is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis. *PloS One* (2015) 10(4):e0122684. doi: 10.1371/journal.pone.0122684

101. Goumari MM, Farhani I, Nezafat N, Mahmoodi S. Multi-epitope vaccines (MEVs), as a novel strategy against infectious diseases. *Curr Proteomics* (2020) 17 (5):354–64. doi: 10.2174/1570164617666190919120140

102. Sette A, Livingston B, McKinney D, Appella E, Fikes J, Sidney J, et al. The development of multi-epitope vaccines: Epitope identification, vaccine design and clinical evaluation. *Biologicals* (2001) 29(3):271–6. doi: 10.1006/biol.2001.0297

103. Rodrigues RR, Ferreira MRA, Kremer FS, Donassolo RA, Júnior CM, Alves MLF, et al. Recombinant vaccine design against clostridium spp. toxins using immunoinformatics tools. In: S Thomas, editor. *Vaccine design: Methods and protocols, volume 3 resources for vaccine development*, vol. p . New York, NY: Springer US (2022). p. 457–70.

104. Smith CC, Entwistle S, Willis C, Vensko S, Beck W, Garness J, et al. Landscape and selection of vaccine epitopes in SARS-CoV-2. *BioRxiv Prepr Serv Biol* (2020) 13(1):101. doi: 10.1101/2020.06.04.135004

105. Saxová P, Buus S, Brunak S, Keşmir C. Predicting proteasomal cleavage sites: A comparison of available methods. *Int Immunol* (2003) 15(7):781–7. doi: 10.1093/intimm/dxg084

106. Calis JJA, Reinink P, Keller C, Kloetzel PM, Keşmir C. Role of peptide processing predictions in T cell epitope identification: Contribution of different prediction programs. *Immunogenetics* (2015) 67(2):85–93. doi: 10.1007/s00251-014-0815-0

107. Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med* (2015) 7:119. doi: 10.1186/s13073-015-0245-0

108. Seligmann H. Cost-minimization of amino acid usage. *J Mol Evol* (2003) 56 (2):151–61. doi: 10.1007/s00239-002-2388-z

109. Carrasco Pro S, Sidney J, Paul S, Lindestam Arlehamn C, Weiskopf D, Peters B, et al. Automatic generation of validated specific epitope sets. *J Immunol Res* (2015) 2015:763461. doi: 10.1155/2015/763461

110. Tian Y, da Silva Antunes R, Sidney J, Lindestam Arlehamn CS, Grifoni A, Dhanda SK, et al. A review on T cell epitopes identified using prediction and cell-mediated immune models for mycobacterium tuberculosis and bordetella pertussis. *Front Immunol* (2018) 9:2778. doi: 10.3389/fimmu.2018.02778

111. Longmate J, York J, La Rosa C, Krishnan R, Zhang M, Senitzer D, et al. Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. *Immunogenetics* (2001) 52(3–4):165–73. doi: 10.1007/s002510000271

112. Westernberg L, Schulten V, Greenbaum JA, Natali S, Tripple V, McKinney DM, et al. T-Cell epitope conservation across allergen species is a major determinant of immunogenicity. *J Allergy Clin Immunol* (2016) 138(2):571–578.e7. doi: 10.1016/j.jaci.2015.11.034

113. Ikram A, Anjum S, Tahir M. In silico identification and conservation analysis of b-cell and T-cell epitopes of hepatitis c virus 3a genotype enveloped glycoprotein 2 from Pakistan: A step towards heterologous vaccine design. *Hepat Mon* (2014) 14(6):e9832. doi: 10.5812/hepatmon.9832

114. Aldakheel FM, Abrar A, Munir S, Aslam S, Allemailem KS, Khurshid M, et al. Proteome-wide mapping and reverse vaccinology approaches to design a multi-epitope vaccine against clostridium perfringens. *Vaccines* (2021) 9(10):1079. doi: 10.3390/vaccines9101079

115. Tropp BE. Cardiolipin synthase from escherichia coli. *Biochim Biophys Acta* (1997) 1348(1–2):192–200. doi: 10.1016/S0005-2760(97)00100-8

116. Provoda CJ, Lee KD. Bacterial pore-forming hemolysins and their use in the cytosolic delivery of macromolecules. *Adv Drug Deliv Rev* (2000) 41(2):209–21. doi: 10.1016/S0169-409X(99)00067-8

117. Paredes-Sabja D, Setlow P, Sarker MR. Germination of spores of bacillales and clostridiales species: Mechanisms and proteins involved. *Trends Microbiol* (2011) 19(2):85–94. doi: 10.1016/j.tim.2010.10.004

118. Malik JA, Mulla AH, Farooqi T, Pottoo FH, Anwar S, Rengasamy KRR. Targets and strategies for vaccine development against SARS-CoV-2. *BioMed Pharmacother* (2021) 137:111254. doi: 10.1016/j.biopha.2021.111254

119. Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, et al. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep* (2020) 10(1):10895. doi: 10.1038/s41598-020-67749-1

120. Hou J, Liu Y, Hsi J, Wang H, Tao R, Shao Y. Cholera toxin b subunit acts as a potent systemic adjuvant for HIV-1 DNA vaccination intramuscularly in mice. *Hum Vaccines Immunother* (2014) 10(5):1274–83. doi: 10.4161/hv.28371

121. Kim HJ, Kim JK, Seo SB, Lee HJ, Kim HJ. Intranasal vaccination with peptides and cholera toxin subunit b as adjuvant to enhance mucosal and systemic immunity to respiratory syncytial virus. *Arch Pharm Res* (2007) 30(3):366–71. doi: 10.1007/BF02977620

122. Tamura S, Funato H, Nagamine T, Aizawa C, Kurata T. Effectiveness of cholera toxin b subunit as an adjuvant for nasal influenza vaccination despite pre-existing immunity to CTB. *Vaccine* (1989) 7(6):503–5. doi: 10.1016/0264-410X(89)90273-9

123. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, et al. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *J Immunol Baltim Md* (2002) 168(11):5499–506. doi: 10.4049/jimmunol.168.11.5499

124. Liu C, Chin JX, Lee DY. SynLinker: An integrated system for designing linkers and synthetic fusion proteins. *Bioinforma Oxf Engl* (2015) 31(22):3700–2. doi: 10.1093/bioinformatics/btv447

125. Li G, Huang Z, Zhang C, Dong BJ, Guo RH, Yue HW, et al. Construction of a linker library with widely controllable flexibility for fusion protein design. *Appl Microbiol Biotechnol* (2016) 100(1):215–25. doi: 10.1007/s00253-015-6985-3

126. Kalita P, Lyngdoh DL, Padhi AK, Shukla H, Tripathi T. Development of multi-epitope driven subunit vaccine against fasciola gigantica using immunoinformatics approach. *Int J Biol Macromol* (2019) 138:224–33. doi: 10.1016/j.ijbiomac.2019.07.024

127. Vakili B, Eslami M, Hatam GR, Zare B, Erfani N, Nezafat N, et al. Immunoinformatics-aided design of a potential multi-epitope peptide vaccine against leishmania infantum. *Int J Biol Macromol* (2018) 120:1127–39. doi: 10.1016/j.ijbiomac.2018.08.125

128. Enayatkhani M, Hasaniazad M, Faezi S, Gouklani H, Davoodian P, Ahmadi N, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An *in silico* study. *J Biomol Struct Dyn* (2021) 39(8):2857–72. doi: 10.1080/07391102.2020.1756411

129. Bhattarai A, Emerson IA. Dynamic conformational flexibility and molecular interactions of intrinsically disordered proteins. *J Biosci* (2020) 45 (1):29. doi: 10.1007/s12038-020-0010-4

130. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. *Proc Natl Acad Sci U S A* (1987) 84 (23):8568–72. doi: 10.1073/pnas.84.23.8568

131. Dym O, Eisenberg D, Yeates TO. Detection of errors in protein models. *International Tables for Crystallography* (2006), F. ch. 21.3:520–30. doi: 10.1107/97809553602060000709

132. Khatoon N, Pandey RK, Prajapati VK. Exploring leishmania secretory proteins to design b and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep* (2017) 7(1):8285. doi: 10.1038/s41598-017-08842-w

133. Tahir Ul Qamar M, Rehman A, Tusleem K, Ashfaq UA, Qasim M, Zhu X, et al. Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. *PloS One* (2020) 15 (12):e0244176. doi: 10.1371/journal.pone.0244176

134. Tahir ul Qamar M, Shahid F, Aslam S, Ashfaq UA, Aslam S, Fatima I, et al. Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against SARS-CoV-2. *Infect Dis Poverty* (2020) 9:132. doi: 10.1186/s40249-020-00752-w

135. Tariq MH, Bhatti R, Ali NF, Ashfaq UA, Shahid F, Almatroudi A, et al. Rational design of chimeric multiepitope based vaccine (MEBV) against human T-cell lymphotropic virus type 1: An integrated vaccine informatics and molecular docking based approach. *PloS One* (2021) 16(10):e0258443. doi: 10.1371/journal.pone.0258443

136. Takehara M, Kobayashi K, Nagahama M. Toll-like receptor 4 protects against clostridium perfringens infection in mice. *Front Cell Infect Microbiol* (2021) 11:633440. doi: 10.3389/fcimb.2021.633440

137. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* (2007) 130(6):1071–82. doi: 10.1016/j.cell.2007.09.008

138. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* (2009) 458 (7242):1191–5. doi: 10.1038/nature07830

139. Frank SA. *Immunology and evolution of infectious disease*. Princeton University Press (2020). doi: 10.1515/9780691220161

140. Crotty S, Ahmed R. Immunological memory in humans. *Semin Immunol* (2004) 16(3):197–203. doi: 10.1016/j.smim.2004.02.008

141. Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thöny-Meyer L. Production of glycoprotein vaccines in escherichia coli. *Microb Cell Factories* (2010) 9(1):61. doi: 10.1186/1475-2859-9-61

142. Tripathi NK, Shrivastava A. Recent developments in recombinant protein-based dengue vaccines. *Front Immunol* (2018) 9:1919. doi: 10.3389/fimmu.2018.01919

143. Aguilar-Yáñez JM, Portillo-Lara R, Mendoza-Ochoa GI, García-Echauri SA, López-Pacheco F, Bulnes-Abundis D, et al. An influenza A/H1N1/2009 hemagglutinin vaccine produced in escherichia coli. *PloS One* (2010) 5(7): e11694. doi: 10.1371/journal.pone.0011694

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fmed.2022.973036

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Maha A. El Demellawy,
City of Scientific Research and Technological Applications, Egypt
James Tristan Gordy,
Johns Hopkins University, United States

*CORRESPONDENCE
Mohammad Mehdi Amiri
m_amiri@tums.ac.ir
Christiane Moog
c.moog@unistra.fr
Fazel Shokri
fshokri@tums.ac.ir

SPECIALTY SECTION
This article was submitted to Infectious Diseases – Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

RECEIVED 19 June 2022
ACCEPTED 16 August 2022
PUBLISHED 06 September 2022

CITATION
Maghsood F, Amiri MM, Zarnani AH, Salimi V, Kardar GA, Khoshnoodi J, Mobini M, Ahmadi Zare H, Ghaderi A, Jeddi-Tehrani M, Schmidt S, Laumond G, Moog C and Shokri F (2022) Epitope mapping of severe acute respiratory syndrome coronavirus 2 neutralizing receptor binding domain-specific monoclonal antibodies.
Front. Med. 9:973036.
doi: 10.3389/fmed.2022.973036

Epitope mapping of severe acute respiratory syndrome coronavirus 2 neutralizing receptor binding domain-specific monoclonal antibodies

Faezeh Maghsood[1], Mohammad Mehdi Amiri[1]*, Amir-Hassan Zarnani[1], Vahid Salimi[2], Gholam Ali Kardar[3], Jalal Khoshnoodi[1], Maryam Mobini[1], Hengameh Ahmadi Zare[1], Abbas Ghaderi[4], Mahmood Jeddi-Tehrani[5], Sylvie Schmidt[6], Géraldine Laumond[6], Christiane Moog[6]* and Fazel Shokri[1]*

[1]Department of Immunology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, [2]Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, [3]Immunology, Asthma and Allergy Research Institute, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran, [4]Shiraz Institute for Cancer Research, School of Medicine, Shiraz University of Medical Science, Shiraz, Iran, [5]Monoclonal Antibody Research Center, Avicenna Research Institute, Academic Center for Education, Culture and Research (ACECR), Tehran, Iran, [6]Laboratoire d'ImmunoRhumatologie Moléculaire, Institut National de la Santé et de la Recherche Médicale (INSERM) UMR_S 1109, Institut Thématique Interdisciplinaire (ITI) de Médecine de Précision de Strasbourg, Transplantex NG, Faculté de Médecine, Fédération Hospitalo-Universitaire OMICARE, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Strasbourg, France

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the outbreak led to the coronavirus disease 2019 (COVID-19) pandemic. Receptor binding domain (RBD) of spike (S) protein of SARS-CoV-2 is considered as a major target for immunotherapy and vaccine design. Here, we generated and characterized a panel of anti-RBD monoclonal antibodies (MAbs) isolated from eukaryotic recombinant RBD-immunized mice by hybridoma technology. Epitope mapping was performed using a panel of 20-mer overlapping peptides spanning the entire sequence of the RBD protein from wild-type (WT) Wuhan strain by enzyme-linked immunosorbent assay (ELISA). Several hybridomas showed reactivity toward restricted RBD peptide pools by Pepscan analysis, with more focus on peptides encompassing aa 76–110 and 136–155. However, our MAbs with potent neutralizing activity which block SARS-CoV-2 spike pseudovirus as well as the WT virus entry into angiotensin-converting enzyme-2 (ACE2) expressing HEK293T cells showed no reactivity against these peptides. These findings, largely supported by the Western blotting results suggest that the neutralizing MAbs recognize mainly conformational epitopes. Moreover,

our neutralizing MAbs recognized the variants of concern (VOC) currently in circulation, including alpha, beta, gamma, and delta by ELISA, and neutralized alpha and omicron variants at different levels by conventional virus neutralization test (CVNT). While the neutralization of MAbs to the alpha variant showed no substantial difference as compared with the WT virus, their neutralizing activity was lower on omicron variant, suggesting the refractory effect of mutations in emerging variants against this group of neutralizing MAbs. Also, the binding reactivity of our MAbs to delta variant showed a modest decline by ELISA, implying that our MAbs are insensitive to the substitutions in the RBD of delta variant. Our data provide important information for understanding the immunogenicity of RBD, and the potential application of the novel neutralizing MAbs for passive immunotherapy of SARS-CoV-2 infection.

KEYWORDS

COVID-19, monoclonal antibodies, neutralization, omicron, SARS-CoV-2

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan, China in December 2019 (1). As of the 24th of April 2022, there were over 500 million global cases of coronavirus disease 2019 (COVID-19) and more than 6 million deaths worldwide (2). SARS-CoV-2, known as the third highly pathogenic human CoV belonging to the lineage B beta-coronaviruses, is a zoonotic enveloped virus containing a positive-sense single-stranded RNA, presumably originated from bats due to sharing 96% genome sequence identity with RaTG13, a bat-derived SARS-like CoV (3, 4).

The genome of SARS-CoV-2 encodes several structural and non-structural proteins. Homotrimeric spike (S) glycoprotein on the viral surface is involved in cell attachment, membrane fusion, and viral entry (5, 6). The S protein with a length of 1,273 amino acids (aa), is a clove-shaped, type I transmembrane protein consisting of a signal peptide (1–13), S1 subunit (14–685), and S2 subunit (686–1,273). The S1 subunit is composed of the N-terminal domain (NTD) (18–305), the C-terminal receptor binding domain (RBD) (329–528), subdomain-1 (SD1) (529–589), and SD2 (590–686) (7). The RBD consists of two sub-domains, including a core sub-domain composed of a β-sheet with five antiparallel strands (β1–β4, and β7) in the inner side of the spike protein and receptor-binding motif (RBM) extending from the core sub-domain and consisting of β5 and β6 strands (8, 9). The RBM is responsible for virus binding to its receptor, the angiotensin-converting enzyme-2 (ACE2), by forming a surface to cradle the N-terminal α-helix of ACE2 expressed on the host cell surface (7). The S2 subunit consists of the upstream helix (UH) (687–819), N-terminal fusion peptide (FP) (820–846), heptapeptide repeat sequence 1 (HR1) (912–985), SD3 (1,072–1,139), stem helix (SH) (1,139–1,163), HR2 (1,163–1,212), TM domain (1,213–1,237), and intracellular domain (1,238–1,273). The S2 subunit plays an important role in mediating fusion of viral membrane with host cell membrane and virus entry into target cells (7, 10). The S protein of SARS-CoV-2 shares about 76% amino acid sequence homology with that of SARS-CoV and both use ACE2 as a receptor for viral entry in a similar way (11, 12). Also, the core sub-domain of RBD, rather than RBM, is more conserved between SARS-CoV-2 and SARS-CoV viruses (identity of 86.3% for the core sub-domain versus 46.7% for the RBM sub-domain) (13). Notwithstanding the critical function of both S1 and S2 subunits in viral entry, it has been shown that anti-S1 antibodies bind to the S protein and neutralize the virus more efficiently than anti-S2 antibodies, presumably due to lack of a major neutralizing region on the S2 subunit (14). Despite less conserved residues and high mutation rates in the RBD rather than in other parts of the S protein, RBD is still considered the most important region of SARS-CoV-2 to be targeted by neutralizing antibodies due to difficult access to and tight folding of fusion domains (15).

Neutralizing monoclonal antibodies (MAbs) have been considered as one of the most promising approaches to target SARS-CoV-2 surface protein as a prophylactic/therapeutic alternative of COVID-19 convalescent plasma for the passive immunotherapy (16–19), since they have been effective against other viruses such as rabies, HIV, RSV, EBOV, SARS-CoV, and MERS-CoV (20, 21). Accordingly, considerable efforts have been made to produce MAbs capable of neutralizing SARS-CoV-2. Until now, U.S. Food and Drug Administration (FDA) or European Medicines Agency (EMA) have not approved neutralizing MAbs for COVID-19 infection, although a number of MAbs have been authorized for emergency use

as prophylaxis in adults and children at high risk of severe COVID-19, including sotrovimab, REGN-CoV-2, bebtelovimab, and the combination of bamlanivimab and etesevimab (22–26). Considering the emergence of newly fast-spreading variants and reduced or lack of efficacy of a number of neutralizing MAbs, there is an urgent need for development of efficient neutralizing MAbs that cross-neutralize various variants to control SARS-CoV-2 infection and/or disease progression. In this study, we produced a panel of MAbs directed against the RBD of SARS-CoV-2 using hybridoma technology. These MAbs were then extensively characterized and their neutralization potency against different variants of SARS-CoV-2 was investigated *in vitro*.

Materials and methods

Sources of cell lines, antigens, antibodies, and reagents

Sp2/0-Ag14 (murine myeloma cell line) was purchased from the National Cell Bank of Iran (NCBI) (Pasteur Institute, Tehran, Iran). ACE2-expressing HEK293T cell line used for pseudovirus-based neutralization test (PVNT) was a kind gift from Renap Therapeutics Co. (Tehran, Iran). Vero 76 cell line (Vero C1008; CRL-1586, Clone E6) was obtained from ATCC. Recombinant spike protein of SARS-CoV-2 with C-terminal histidine tag expressed in Baculovirus insect cells and recombinant RBD of SARS-CoV-2 spike protein with C-terminal histidine tag (RBD-His) expressed in HEK293T cells were purchased from Sino Biological Inc. (Beijing, China). Full-length trimeric spike antigen variants were obtained from BioServUK–CalibreScientific (Sheffield, United Kingdom). Rabbit anti-sheep immunoglobulins (Ig), and horse-radish peroxidase (HRP)-conjugated sheep anti-mouse Ig were purchased from SinaBiotech (Tehran, Iran). Mouse monoclonal antibody isotyping reagents, goat anti-mouse IgG, IgA, and IgM isotype-specific polyclonal antibodies, complete and incomplete Freund's adjuvants, 2-Mercaptoethanol (2ME), Tween-20, skimmed milk, pristane and all reagents used for cell culture, including Roswell Park Memorial Institute (RPMI) 1640 medium, fetal bovine serum (FBS), L-glutamine, penicillin and streptomycin, HAT supplement (50X), polyethylene glycol (PEG 1500), and dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (Darmstadt, Germany). Anti-nucleocapsid antibody, Alexa Fluor 647-labeled goat anti-rabbit monoclonal antibody, and Sytox green nucleic acid stain were obtained from GeneTex (CA, United States), Abcam (Cambridge, United Kingdom), and Invitrogen (MA, United States), respectively. Streptococcal protein G (SPG) affinity columns and chemiluminescence ECL Prime solution were obtained from GE Healthcare (Nordrhein-Westfalen, Germany). The linear peptides were synthesized by Pepmic company (Jiangsu, China). Tetramethylbenzidine (TMB) and SARS-CoV-2 neutralizing antibody detection kit were obtained from Pishtaz Teb Co. (Tehran, Iran). Maxisorp microtiter ELISA plates and cell culture plates and flasks were purchased from Nunc (Roskilde, Denmark).

Production of anti-severe acute respiratory syndrome coronavirus 2 receptor binding domain-specific monoclonal antibodies

A total of 6–8 weeks old female BALB/c mice were subcutaneously immunized once with 15 μg and then three times with 7.5 μg of SARS-CoV-2 RBD-His protein in combination with complete and incomplete Freund's adjuvants, respectively, at 2-week intervals. Blood samples were then taken from the tail vein before each injection and sera were prepared by centrifugation. Hyper-immunization was confirmed by enzyme-linked immunosorbent assay (ELISA). Three to 5 days after intravenous (i.v.) injection of antigen, the spleen was harvested, and hybridomas were generated by fusing the extracted splenocytes with Sp2/0-Ag14 myeloma cell line at a 6:1 ratio using PEG 1500. Only hybridomas were grown in the presence of a selective medium containing hypoxanthine, aminopterin, and thymidine (HAT 1X). Hybridomas were screened and sub-cloned by RBD-specific indirect ELISA and limiting dilution method, respectively to obtain anti-RBD, anti-S final clones. MAb producing final clones (5×10^6 cells/mouse) were intraperitoneally injected into pristane-pretreated BALB/c mice. Ascitic fluids were collected, and the SARS-CoV-2 RBD-specific MAbs were purified using a SPG affinity column.

Linear peptide synthesis

Amino acid sequences of SARS-CoV-2 RBD were obtained from GenBank under accession number YP_009724390.1 to design the linear peptides. A panel of linear peptides spanning the entire sequence of SARS-CoV-2 RBD (each linear peptide contained 20 amino acid residues with 5 residues overlapping with the adjacent peptides) with purity of more than 90% were used for Pepscan. Lyophilized peptides were dissolved in deionized water and/or DMSO to obtain a stock solution, according to the manufacturer's instructions. Amino acid sequences of peptide sets are shown in **Table 1**.

Enzyme-linked immunosorbent assay

We conducted seven different types of ELISA to evaluate the levels of anti-RBD Ig; in immunized mouse sera, in hybridoma clone supernatants, and to characterize the anti-RBD MAbs:

TABLE 1 Amino acid sequences of RBD peptide sets used in Pepscan.

No.	Code	Position in spike	Amino acid sequence
1	**P1–20**	319–338	RVQPTESIVRFPNITNLCPF
2	**P16–35**	334–353	NLCPFGEVFNATRFASVYAW
3	**P31–50**	349–368	SVYAWNRKRISNCVADYSV
4	**P46–65**	364–383	DYSVLYNSASFSTFKCYGVS
5	**P61–80**	379–398	CYGVSPTKLNDLCFTNVYAD
6	**P76–95**	394–413	NVYADSFVIRGDEVRQIAPG
7	**P91–110**	409–428	QIAPGQTGKIADYNYKLPDD
8	**P106–125**	424–443	KLPDDFTGCVIAWNSNNLDS
9	**P121–140**	439–458	NNLDSKVGGNYNYLYRLFRK
10	**P136–155**	454–473	RLFRKSNLKPFERDISTEIY
11	**P151–170**	469–488	STEIYQAGSTPCNGVEGFNC
12	**P166–185**	484–503	EGFNCYFPLQSYGFQPTNGV
13	**P181–200**	499–518	PTNGVGYQPYRVVVLSFELL
14	**P196–215**	514–533	SFELLHAPATVCGPKKSTNL
15	**P204–223**	522–541	ATVCGPKKSTNLVKNKCVNF

Orange characters show RBM amino acids. RBD, receptor binding domain; RBM, receptor binding motif.

Evaluation of the levels of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera and primary hybridoma supernatants

Severe acute respiratory syndrome coronavirus 2 RBD or S antigens, at a concentration of 1 μg/ml, were coated in flat-bottom 96-well microtiter plates in PBS (pH 7.4) overnight at 4°C. After washing with PBST (0.05% v/v Tween-20 in PBS) three times, blocking buffer containing 3% w/v skimmed milk in PBST was added to each well for 1 h at 37°C. Subsequently, mouse sera at 1:1,000, 1:5,000, and 1:25,000 dilutions, or hybridoma supernatants at 1:5 dilution were applied onto the plates in the blocking buffer and incubated at 37°C for 1 h, followed by three washes with PBST. HRP-conjugated sheep anti-mouse Ig antibody at 1:2,000 dilution was used to detect the anti-RBD and anti-S levels in RBD-immunized mouse sera and primary hybridoma supernatants. The reaction was developed by adding TMB substrate solution for 15 min and stopped by addition of 1 M H_2SO_4. Then, the optical densities (ODs) of the reactions were measured at 450 and 630 nm using a microplate reader (BioTek, United States).

Peptide-based enzyme-linked immunosorbent assay for receptor binding domain-immunized mouse sera, primary hybridoma supernatants, and monoclonal antibodies

Preliminary epitope screening was performed using five pooled RBD peptide sets each composed of three peptides. ELISA was conducted according to the protocol described in section "Evaluation of the levels of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera and primary hybridoma supernatants" with the following modifications. In brief, 3.5 μg/ml of each RBD peptide in three peptide pool sets, or 1 μg/ml of RBD were coated onto the ELISA plates in PBS and incubated overnight at 4°C. After washing and blocking serum samples at 1:200 dilution or hybridoma supernatants at 1:5 dilution or MAbs at a concentration of 1 μg/ml in the blocking buffer were added and incubated for another 1 h at 37°C. HRP-conjugated sheep anti-mouse Ig antibody was used for detecting peptide-bound antibodies. Color development was performed as described in section "Evaluation of the levels of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera and primary hybridoma supernatants."

Further assessment procedures were performed to evaluate the reactivity against individual immunodominant linear peptides by immobilizing each peptide into ELISA plates with a concentration of 5 μg/ml. ELISA was conducted according to the protocol described above.

Surrogate viral neutralization test

Severe acute respiratory syndrome coronavirus 2 neutralizing antibody detection kit was used for the surrogate viral neutralization test (SVNT) assay according to the manufacturer's instructions. Briefly, ACE2-HRP was mixed with different concentrations of MAbs, serially diluted mouse sera, or peptide-adsorbed mouse sera and added to RBD pre-coated plates. After incubation for 30 min at 37°C, unbound HRP-conjugated antigens were removed by five PBST washes. After adding TMB and stop solution, the OD was measured by a plate reader.

The percentage of inhibition was determined as follows:

$$\text{Inhibition (\%)} = ((\text{negative control OD} - \text{sample OD}) / \text{negative control OD}) \times 100.$$

For peptide adsorption, P76–95, P91–110, and P136–155 were selected as the immunodominant peptides. P2-NP (a linear peptide belonging to SARS-CoV-2 nucleocapsid) and P106–125 were also selected as irrelevant peptide and non-reactive control, respectively. First, the immunized mouse sera were serially diluted at a starting dilution of 1:10 and the ID50 values (inhibition dilution of 50%) were calculated in SVNT. Subsequently, 10 μg/ml of selected peptides were incubated with diluted mouse sera at their ID50 dilution point for 2 h at 37°C to adsorb peptide-specific antibodies. Peptides were separately diluted in the sample diluent at a concentration of 10 μg/ml in the absence of mouse serum as peptide controls to assess whether a similar concentration of peptides could individually inhibit the binding of ACE2-HRP to RBD. Peptide-adsorbed mouse sera, non-adsorbed mouse sera, and peptide controls were then mixed with ACE2-HRP and incubated for 30 min at 37°C in RBD pre-coated plates. ELISA was conducted according to the protocol described above.

Isotype determination

Severe acute respiratory syndrome coronavirus 2 RBD-specific MAbs (1 μg/ml) were coated into 96-well ELISA plates

for 1.5 h at 37°C. After washing with PBST and blocking the remaining binding sites, the plates were incubated with goat anti-mouse IgG, IgA, and IgM isotype-specific polyclonal antibodies (1:1,000) for 20 min at 37°C, followed by washing and incubation with HRP-labeled rabbit anti-sheep Ig (1:3,000) for 20 min at 37°C. After washing the plates, adding TMB solution to the wells, and stopping the enzymatic reaction with 1 M H_2SO_4, ODs were measured as described in section "Evaluation of the levels of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera and primary hybridoma supernatants."

Monoclonal antibodies affinity determination

The binding affinities of SARS-CoV-2 RBD-specific MAbs were determined using an ELISA-based method (27, 28). Briefly, ELISA plates were coated with recombinant SARS-CoV-2 RBD protein (1,000, 500, 250, 125, 62, and 31 ng/ml) for 1.5 h at 37°C. After washing and blocking, serially diluted MAbs were added and incubated for 1 h at 37°C, followed by washing and incubation with HRP-labeled sheep anti-mouse Ig for 1 h at 37°C. After washing, TMB solution was added to the wells, enzymatic reaction was stopped, and ODs were measured as described in section "Evaluation of the levels of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera and primary hybridoma supernatants," followed by calculating affinity constants (KD) following the given equations (27, 28). Briefly, ODs were plotted against logarithmic values of antibody concentration. Antibody concentration resulting in half of the maximum OD ([Ab]t) at each antigen concentration was chosen for the affinity measurement using the equation $KD = (n - 1/2)\ (n[Ab']t - [Ab]t)$. n equals [Ag]/[Ag′], where [Ag] and [Ag′] correspond to the higher and lower concentration of antigen. [Ab′]t and [Ab]t correspond to the antibody concentrations giving 50% of maximum OD at [Ag′] and [Ag], respectively. The mean of the calculations for three non-overlapping antigen concentrations was considered as the final KD value.

Monoclonal antibodies cross-competition assay

To relatively map the epitope location of SARS-CoV-2 RBD-specific MAbs, the ability of unlabeled MAbs (competitors) to compete with HRP-labeled ones was assessed by a competitive ELISA. 96-well ELISA plates were coated with recombinant SARS-CoV-2 RBD protein (1 μg/ml) for 1.5 h at 37°C. After washing the plates with PBST and treating the wells with blocking buffer, 25 μl of each competitor MAb was mixed with 25 μl of each HRP-labeled MAb to reach the final concentration of 5, 20, and 60 μg/ml of competitor and 1 μg/ml of HRP-labeled MAb. The mixtures were added to RBD-coated plates for 1 h at 37°C. After washing, the enzymatic reaction was developed by TMB solution and stopped by 1 M H_2SO_4, followed by measuring ODs. The percentage of competition was determined as follows:

$$\text{Competition (\%)} = ((\text{HRP} - \text{labeled MAb OD alone} - \text{competitor plus HRP} - \text{labeled MAb OD}) / \text{HRP} - \text{labeled MAb OD alone}) \times 100.$$

Reactivity of monoclonal antibodies to severe acute respiratory syndrome coronavirus 2 spike variants

Full-length trimeric spike antigen variants were coated with a concentration of 2 and 0.5μg/ml. After blocking, 1 or 0.25 μg/ml of selected MAbs were added to the plates followed by washing and incubating with HRP-labeled sheep anti-mouse Ig for 1 h at 37°C. The reaction development and stopping were done by TMB and 1 M H_2SO_4, respectively.

Western blotting

Reactivity of selected MAbs against non-reduced or 1% 2ME-reduced form of RBD was assessed by Western blotting. A total of 1 μg of RBD was applied to 10% polyacrylamide gel in SDS sample buffer. Proteins were separated by electrophoresis at 100 V for 1 h and transferred to a 0.45 μm hydrophilic polyvinylidene fluoride (PVDF) membrane at 110 V for 1.5 h. After blocking the membranes in blocking buffer (5% skimmed milk in PBS) overnight at 4°C, the membranes were incubated with 1 μg/ml of MAbs in the blocking solution for 45 min at room temperature (RT). Subsequently, the membranes were washed five times with PBST, then incubated with secondary HRP-conjugated sheep anti-mouse Ig at RT for 45 min, followed by washing five times for 5 min. Finally, positive signals were detected by chemiluminescence ECL Prime solution.

Pseudovirus-based neutralization test

Angiotensin-converting enzyme-2-expressing HEK293T cells were cultured in RPMI-1640 medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, and incubated at 37°C, 5% CO_2 and, 95% humidity. MAbs were diluted in RPMI-1640 medium supplemented with 10% FBS and mixed with the same volume of eGFP-SARS-CoV-2 spike pseudotyped lentivirus corresponding to wild-type (WT, D614G genotype) to reach the final concentration of 15, 5, 0.5, 0.1, and 0.01 μg/ml from each intact MAb in a 96-well plate, followed by incubation at 37°C for 2 h. Subsequently, ACE2-expressing HEK293T cells were detached by trypsin-EDTA 0.25% and added to each well (14×10^3 cells/well). Following incubation for 48 h, the medium was removed. Fluorescence microscopy was used

for imaging and detecting the pseudovirus-infected eGFP-positive cells. Microscopic images were taken from at least four microscopic fields. ImageJ software was used to analyze the microscopic images and calculate the fluorescence positive cells. The inhibitory concentration of 50% (IC50) was defined as the Ab concentration leading to a 50% reduction in the percentage of infected cells. IC50 values were determined as described by Ferrara et al. (29).

Conventional authentic virus-neutralizing assay

Vero 76 cells were plated to a 96-well plate at a density of 12.5×10^3 cells/well, 1 day before the infection. A total of 50 μl of fourfold serially diluted MAbs starting at 20 μg/ml were incubated with 50 μl of WT SARS-CoV-2 strain (D614G genotype), alpha, or omicron strain at MOI of 10, leading to about 30% infected cells and further added to priorly seeded Vero 76 cells. After 48 h, the cells were fixed in methanol for 20 min, washed in PBS, and stained with rabbit anti-nucleocapsid antibody at 1:200 dilution in the perm-wash buffer for 45 min at RT. Subsequently, Alexa Fluor 647-labeled goat anti-rabbit MAb at 1:200 dilution in PBS containing 5% FBS was added and incubated for 45 min at RT. Total cells were detected by Sytox green staining. Total cells (Sytox green positive cells) and infected cells (nucleocapsid positive cells) were counted using SpectraMax MiniMax Imaging Cytometer (Molecular Devices LLC). The percentage of infected cells was recorded and IC50 was calculated as described in section "Conventional authentic virus-neutralizing assay."

Statistical analyses

All data were statistically analyzed by Prism v9 (San Diego, CA, United States) and represented as the mean ± SD. Binding inhibition of ACE2 in SVNT by immunodominant peptides or non-reactive and irrelevant peptides-adsorbed RBD-immunized mouse sera was analyzed by ordinary one-way ANOVA. The *p*-values of less than *0.05, **0.01, ***0.001, and ****0.0001) were considered statistically significant. The quantitative cut-off value for positive reactivity of hybridoma supernatants in Pepscan analysis was defined as the mean OD of negative samples plus 3 SDs. Inhibition rates of MAbs and sera were calculated based on the decrease in the fluorescence positive cells (for pseudotype-based neutralization assay) or the decrease in nucleocapsid positive cells (for live SARS-CoV-2-based neutralization assay). IC50 value (inhibition concentration 50%) for MAbs or ID50 value (inhibition dilution 50%) for sera were calculated using non-linear regression, i.e., (inhibitor) vs. response (four parameters). Correlation between SVNT, PVNT, and conventional virus neutralization test (CVNT) was analyzed by Spearman test. Fold change was computed as the ratio of the changes between final IC50 values of MAbs against variants of concern (VOCs) (Y) and the original IC50 values of MAbs against WT, D614G genotype (X) over the initial value using the following equation: Fold change $= (Y - X)/X$.

Results

Titration of anti-receptor binding domain and anti-spike antibodies in receptor binding domain-immunized mouse sera

We employed RBD protein with C-terminal histidine tag (RBD-His) as an immunogen to immunize BALB/c mice (the study design is depicted in **Figure 1**). The mice were primed with RBD-His in combination with complete Freund's adjuvant on day 0 and boosted with RBD-His and incomplete Freund's adjuvant at weeks 4, 6, and 8. Bleeding was done before each injection and 4 weeks after the last booster dose (**Figure 1**). Titration of hyperimmune mouse sera against RBD and S proteins by ELISA showed a mean half-maximal effective serum titer of 6,000 and 7,000 against spike and RBD, respectively, 3–5 days after i.v. booster administration of RBD (**Figures 2A,B**).

Pepscan analysis of receptor binding domain-immunized mouse serum

To map the linear epitopes recognized by mouse polyclonal anti-RBD antibodies, a series of 15 overlapping peptides covering the RBD sequence of the S protein were used as coating antigens in ELISA (**Figures 1B**, **2C,D**). First, three pooled peptide sets (peptide pools A, B, and C) were employed for Pepscan. Preliminary Pepscan analysis indicated one dominant peptide pool (pool B), including amino acids 76–155 (**Figure 2C**). Further Pepscan assay performed on individual linear peptides showed relatively high reactivity to peptides P76–95, P91–110, and P136–155 in comparison with P106–125 and P121–140 (**Figure 2D**). P136–155 is located in RBM, which is critical for ACE2 binding (8), while P76–95 and P91–110 are located in the core subdomain in the N-terminal of RBD.

Adsorption of immunized mouse sera with immunodominant linear peptides

To evaluate the neutralizing effect of antibodies against the immunodominant linear peptides identified by Pepscan, serum adsorption assays with P76–95, P91–110, and P136–155 were performed by SVNT. The immunized mouse sera showed ID50 of about 1:50 dilution in SVNT (**Figure 2E**). We found that sera adsorbed with P76–95, P91–110, or P136–155 peptide

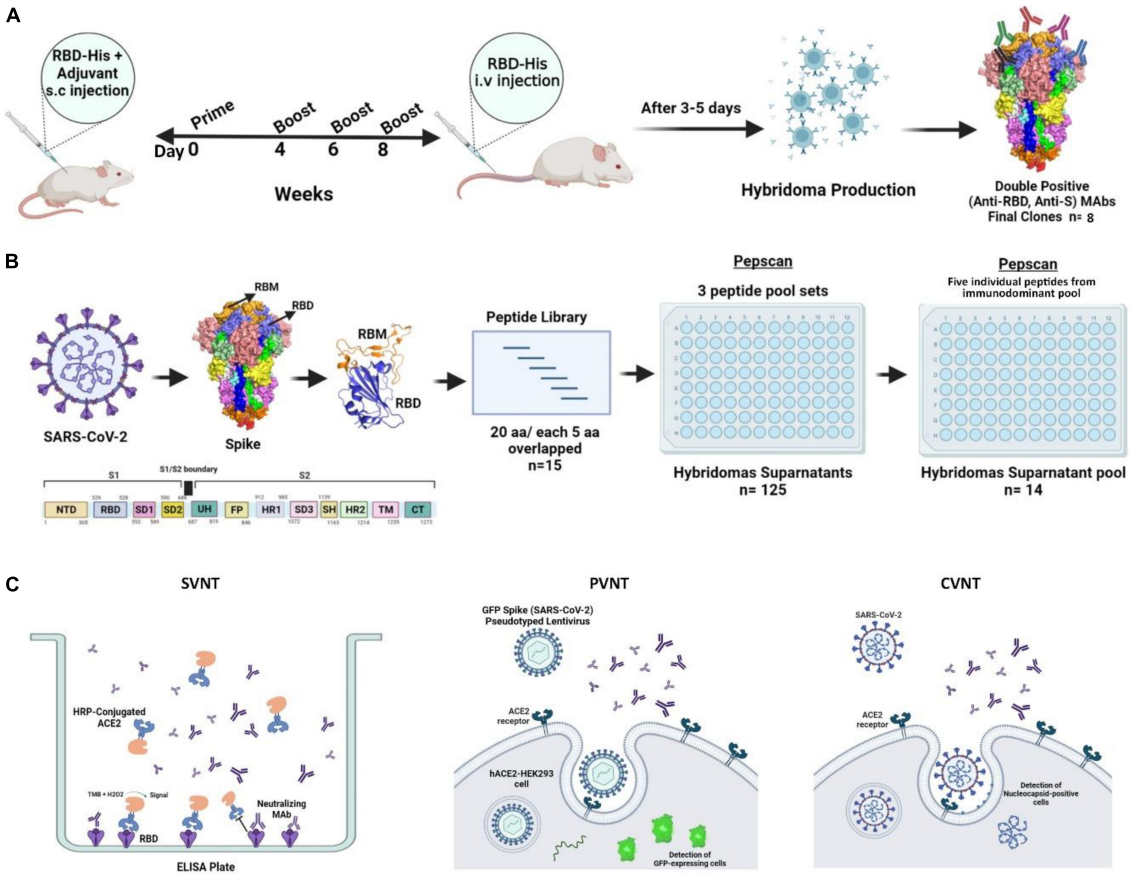


FIGURE 1
Schematic representation of mouse immunization and hybridoma production workflow to generate and characterize anti-SARS-CoV-2-RBD MAbs. **(A)** A total of 6–8-week old female BALB/c mice were immunized with RBD-His and complete Freund's adjuvant on day 0 followed by boosts with RBD-His and incomplete Freund's adjuvant at weeks 4, 6, and 8. Blood samples were collected for ELISA before each injection. A total of 3–5 days after intravenous injection of RBD, mice were sacrificed and splenocytes were fused with Sp2/0 cells. Finally, eight clones were generated with reactivity against both RBD and S antigens. **(B)** SARS-CoV-2 RBD Pepscan was performed using a panel of linear peptides spanning the entire sequence of SARS-CoV-2 RBD of spike containing 20 amino acid residues with 5 residues overlapping with the adjacent peptides for the characterization of hybridoma supernatants and purified MAbs. First, the reactivity of hybridoma supernatants was assessed against three peptide pool sets. Subsequently, the linear peptide-reacting hybridoma supernatants were pooled and their reactivity was assessed against individual peptides of the immunodominant peptide pool. **(C)** SVNT, PVNT, and CVNT were performed to characterize MAbs. aa, amino acid; ACE2, angiotensin-converting enzyme-2; CVNT, conventional virus-neutralizing test; ELISA, enzyme-linked immunosorbent assay; HRP, horse-radish peroxidase; i.v., intravenous; MAb, monoclonal antibody; PVNT, pseudovirus neutralizing test; RBD, receptor binding domain; RBM, receptor binding motif; S, spike; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; s.c., subcutaneous; SVNT, surrogate virus-neutralizing test.

could significantly reduce binding ability of ACE2 to RBD compared with the non-adsorbed serum controls ($p < 0.01$) (Figure 2F), implying that the antibodies targeting these three immunodominant linear epitopes contribute to the anti-RBD neutralizing response.

Isolation and characterization of receptor binding domain-specific monoclonal antibodies

Preliminary growing hybridomas were screened against RBD by ELISA. A total of 125 RBD-reactive hybridomas were initially identified and assessed by Pepscan. Subsequently, eight stable hybridomas producing double reactive (anti-RBD and anti-S) MAbs were selected, cloned, and characterized.

Epitope mapping of anti-receptor binding domain hybridoma supernatants

To determine epitope specificity of the 125 initial RBD-reactive hybridomas as well as the final eight selected clones, three peptide pool sets designated A, B, and C, each composed of five consecutive peptides, were employed and tested by ELISA

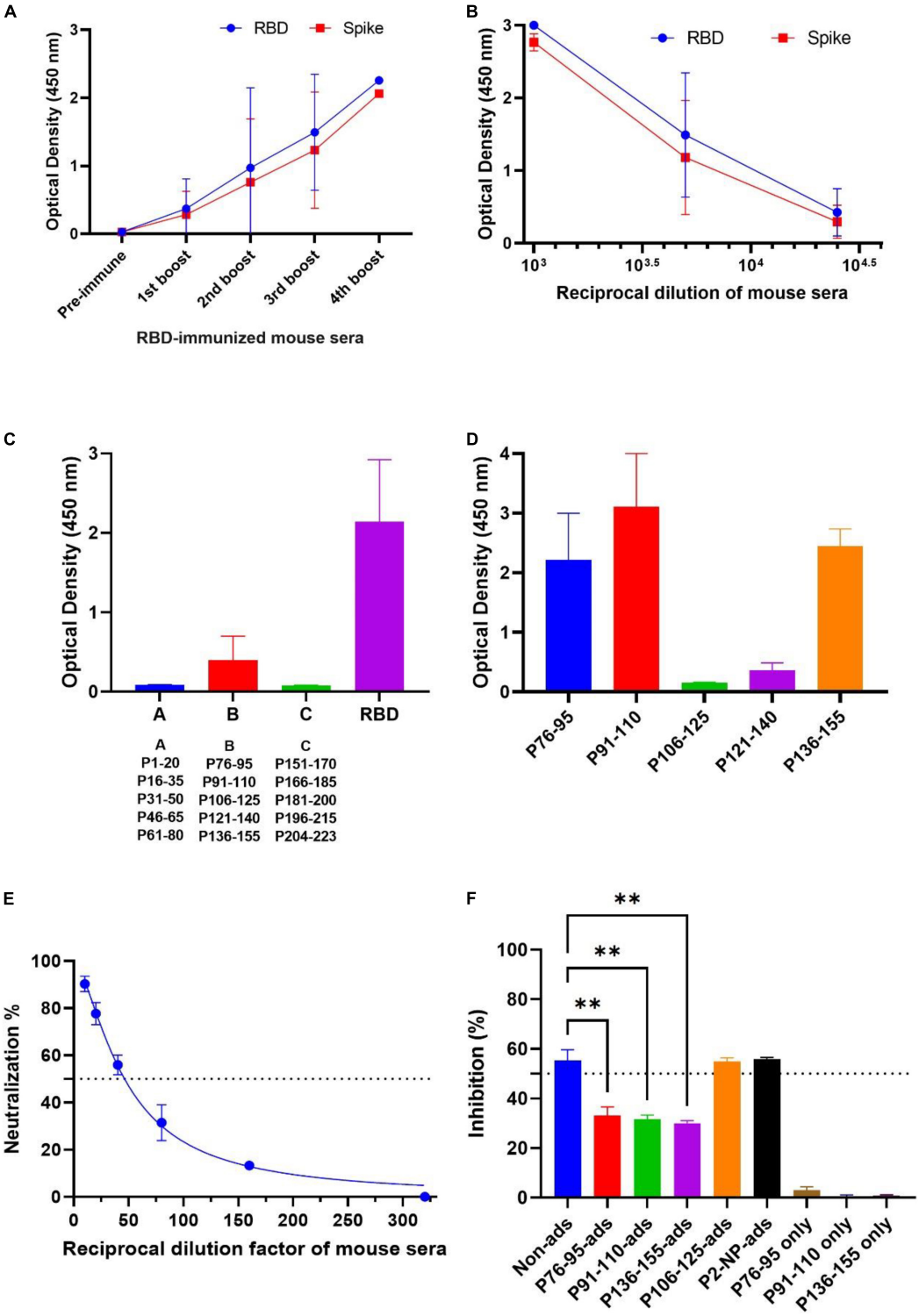


FIGURE 2
Antibody responses in the immunized mouse sera against spike, RBD, and linear peptides by ELISA and peptide-adsorption assay by SVNT. **(A)** Antibody responses against RBD and spike in the immunized mouse sera during the immunization schedule. Sera were tested by ELISA at 1:5,000 dilution. **(B)** Titration of sera collected from the sacrificed mice before the spleen harvest against RBD and S proteins. **(C)** The immunodominant epitopes of the SARS-CoV-2 RBD protein were mapped by Pepscan analysis against mouse immunized sera and the immunodominant peptide pool was identified. Mouse hyperimmune sera (1:200) were subjected to three peptide pools (A–C) of overlapping peptides and RBD antigen by ELISA. Each pool contained five 20-mer peptides spanning the entire RBD domain of the spike. **(D)** The

(Continued)

FIGURE 2 (Continued)
immunodominant epitopes among "pool B" peptides from the RBD protein was identified by ELISA. Individual peptides (P76–95, P91–110, P106–125, P121–140, and P136–155) were coated to screen the serum samples from hyperimmune mice to determine the most reactive peptides. **(E)** Titration curves of immunized mouse sera were presented in SVNT by testing serially diluted sera at a starting dilution of 1:10. The dotted lines on each graph indicate 50% inhibition. **(F)** Binding inhibition of ACE2 in SVNT by immunodominant peptides (P76–95, P91–110, and P136–155) or non-reactive and irrelevant peptides (P106–125 and P2–NP)-adsorbed RBD-immunized mouse sera. Peptides were separately used in the absence of mouse sera as peptide controls. Data analyzed by ordinary one-way ANOVA $^{**}p < 0.01$. The solid lines represent mean values ± SD. ACE2, angiotensin-converting enzyme-2; ELISA, enzyme-linked immunosorbent assay; NP, nucleocapsid; RBD, receptor binding domain; S, spike; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SVNT, surrogate virus-neutralizing test; ads, adsorbed sera.

(Figure 3). The results showed that most of the hybridoma supernatants (n = 70) react with peptide pool B encompassing amino acids 76–155 of the RBD sequence (Figures 3A,B). These findings are consistent with the results obtained from the RBD-immunized mouse sera. Subsequently, peptide pool B-reactive hybridoma supernatants were mixed in 14 pool sets each consisting of five supernatants to evaluate their reactivities against each individual peptide of pool B. Our data demonstrated that a large number of the hybridoma pools sets were reactive to P76–95 (12 out of 14 pool sets) and P136–155 (14 out of 14 pool sets) peptides, which are located in the core sub-domain and RBM of RBD, respectively (Figures 3C,D). Due to shortage of materials, we could not test all the positive hybridoma supernatants individually against the five peptides of the pool B. Finally, eight hybridomas reactive with both RBD and S proteins, designated 1D1, 1D10, 2C5, 2D9, 2F8, 2G3, 3B6, and 3G5 were stable and could be cloned and characterized. Based on Pepscan results, 2F8 MAb reacts with the peptide P136–155 (RLFRKSNLKPFERDISTEIY), which is located in RBM of RBD (8) and 3G5 MAb reacts with the peptide P76–95 (NVYADSFVIRGDEVRQIAPG), which is located in the N-terminal region of the RBD (Table 1). However, despite ELISA binding to the spike and RBD, the remaining 6 MAbs failed to react with any of the peptides.

Affinity and isotype

Representative binding affinity curves obtained for RBD-specific MAbs are presented in Figure 4A. The affinity constants were found to be in the range of 0.36–1.71 nM, the highest and lowest affinities belonged to 3G5 and 2F8 MAbs, respectively. The isotype was found to be IgG for all the MAbs (Table 2).

Cross-competition of monoclonal antibodies for binding to receptor binding domain

Three MAbs (1D1, 1D10, and 3G5) were randomly HRP-conjugated and their binding activities to RBD was measured in the presence or absence of the other unlabeled anti-RBD MAbs, as competitors. Accordingly, three distinct groups of MAbs were identified. Five MAbs, including 1D1, 1D10, 2C5, 2D9, and 2G3 were classified into one group, since they competed with two of the HRP-conjugated MAbs (1D1 and 1D10) with a similar pattern. The binding of 3G5 MAb to RBD, which recognizes a distinct linear epitope (P76–95), is partially inhibited by five other MAbs, including 1D1, 1D10, 2C5, 2D9, and 2G3, implying that 3G5 binds to an epitope in the vicinity of the epitope(s) recognized by these five MAbs. The other two MAbs (3B6 and 2F8) displayed different cross-competition profile (Figure 4B). These results demonstrate that our MAbs recognize multiple epitopes and are not limited to a few immunodominant epitopes in RBD.

Western blotting analysis of the monoclonal antibodies

The reactivities of the MAbs with non-reduced and 2ME-reduced RBD proteins were assessed by Western blotting (Figure 4C). Due to the overload of protein in the gel and the high intensity of the bands, we considered the pale bands as non-specific (pale bands in 50 kD of reduced RBD, corresponding to 2D9 and 3G5 MAbs, as well as pale bands in 25 kD of reduced RBD, corresponding to 3B6, 2G3, and 2C5 MAbs). Although 1D1, 1D10, and 2D9 were not reactive to any of the linear peptides in Pepscan, they were reactive against both non-reduced and 2ME-reduced RBD proteins, suggesting that their epitope is a disulfide bond-independent conformational epitope. It is possible that 1D1, 1D10, and 2D9 (but not 2C5 and 2G3) might be originated from a common clone based on their reactivity obtained from competition ELISA and Western blotting. However, 2G3, 2C5, and 3B6 did not react with 2ME-reduced RBD protein, which indicates that they recognize conformational disulfide bond-dependent epitopes. 3G5, as a linear peptide-reactive MAb (P76–95), showed reactivity against both non-reduced and 2ME-reduced RBD protein. The last MAb (2F8) which reacted with the linear peptide P136–155 in ELISA, recognized the 2ME-reduced, but not non-reduced form of RBD protein which suggests reactivity against a disulfide-bond dependent linear epitope. Two-band pattern of the non-reduced form of recombinant RBD with MW sizes of 25 KD and 50 KD, is observed in almost all developed lanes, which is most likely the result of RBD dimerization (Figure 4C).

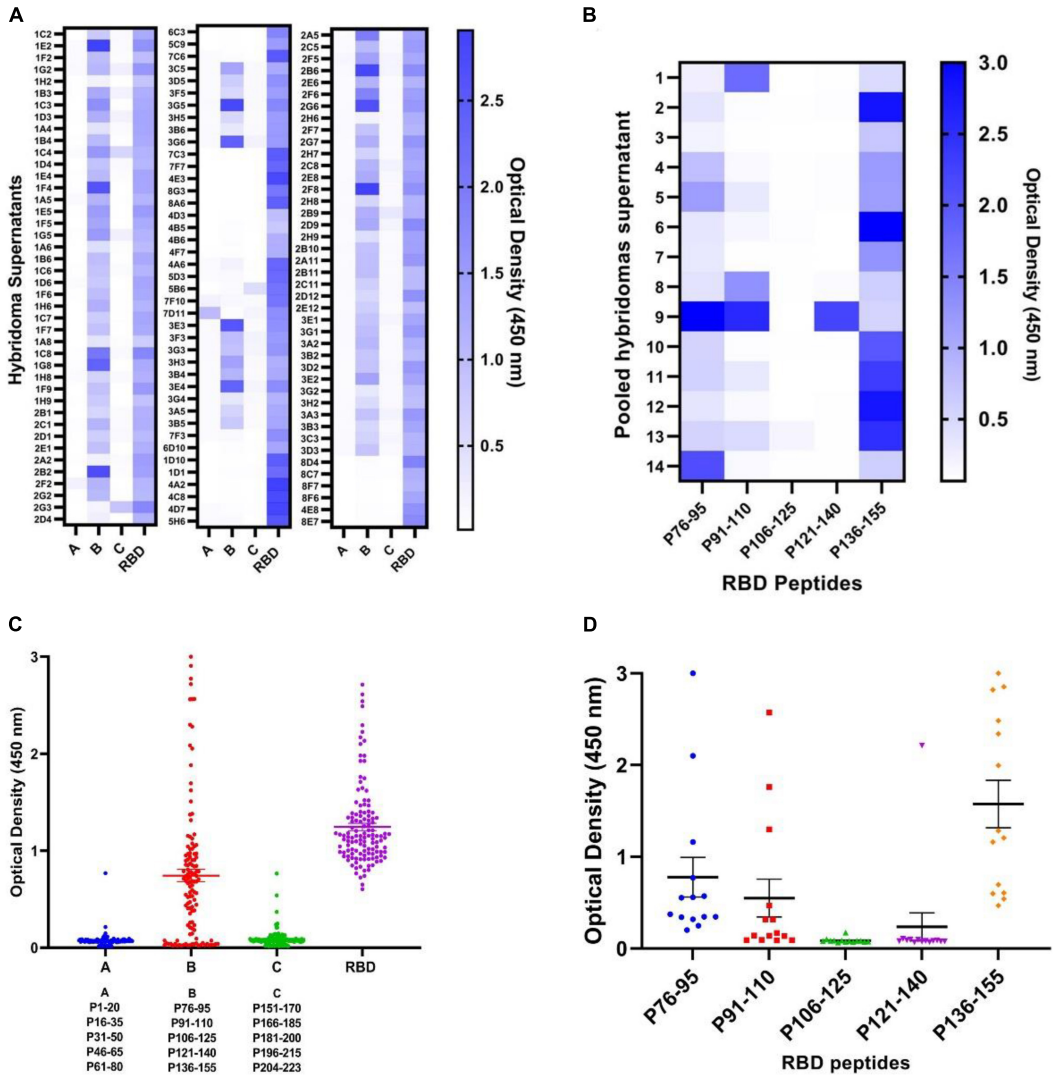


FIGURE 3
Pepscan analysis of hybridoma supernatants by ELISA. **(A,B)** Mapping the immunodominant epitopes of the SARS-CoV-2 RBD protein by Pepscan analysis against hybridoma supernatants and identification of the immunodominant peptide pools. A total of 125 anti-RBD hybridoma supernatants were subjected to three pools **(A–C)** of overlapping peptides and RBD antigen by ELISA. Each pool contains five 20-mer overlapping peptides covering the entire RBD. Reactivities are presented as heatmap and dot plot. **(C,D)** Identification of the immunodominant epitope among "pool B" peptides from the RBD protein. Individual peptides (P76–95, P91–110, P106–125, P121–140, and P136–155) were screened with hybridoma supernatants to determine the most reactive peptide. Reactivities are presented in the heatmap and dot plot. The solid lines represent mean ± SD. ELISA, enzyme-linked immunosorbent assay; RBD, receptor binding domain; S, spike; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Wild-type severe acute respiratory syndrome coronavirus 2 virus neutralization assessment

Surrogate viral neutralization test, PVNT, and CVNT assays were employed to assess the neutralizing potential of the eight selected RBD-specific MAbs generated in this study. In the first step, a single high concentration of all eight selected MAbs was used for the SVNT (40 μg/ml), PVNT (15 μg/ml), and CVNT (10 μg/ml) assays to assess the neutralization potency of the MAbs. As shown in Table 2 and based on the preliminary SVNT, PVNT, and CVNT results, four MAbs, including 1D1, 1D10, 2C5, and 2D9 belonging to a cross-reactive group, demonstrated neutralization capacity in all assays. Two MAbs, 3B6 and 3G5, showed weak neutralization activity (based on SVNT and PVNT results), while the remaining two MAbs (2F8 and 2G3) did not show any neutralization activity at the highest concentrations mentioned above. The six neutralizing MAbs, including 1D1, 1D10, 2C5, 2D9, 3B6, and 3G5, were further tested to determine their IC50 in SVNT, PVNT, and CVNT.

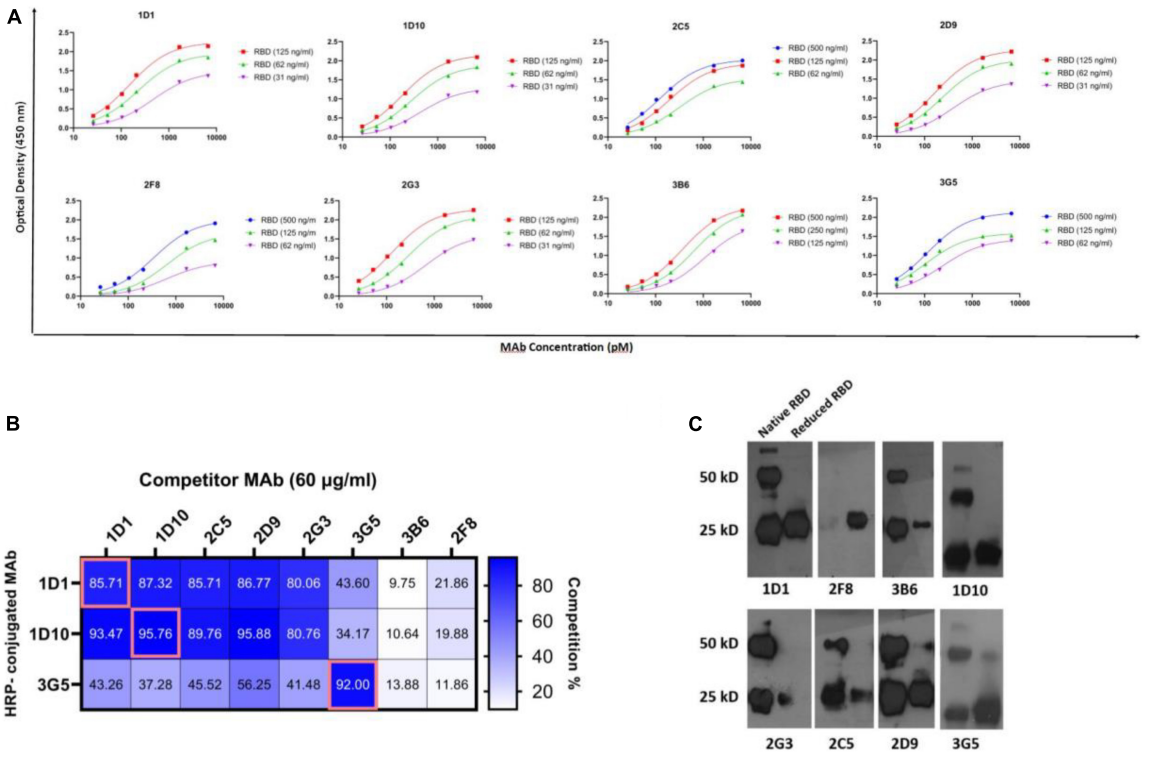


FIGURE 4
Binding affinity, competition profile, and structural characterization of anti-RBD MAbs. **(A)** The binding affinity of MAbs to RBD protein was measured by an ELISA-based method. MAbs concentration at OD 50% was used to calculate affinity constants (Ka) between each pair of sigmoidal curves in the graphs based on the equations referred to in section "Materials and methods." The final Ka was obtained by averaging three calculated Ka. **(B)** The competition profile of MAbs was determined by competitive ELISA. Results of competition ELISA are presented as the percentage of competition by the competitor MAbs compared with no-competitor groups. Competition for more than 50% was considered as positive. **(C)** Western blot representative of reactivity of selected double-reactive MAbs against 2ME-reduced and native RBD proteins. A total of 25 and 50 kD bands show monomeric and dimeric RBD protein. Ag, antigen; ELISA, enzyme-linked immunosorbent assay; HRP, horse-radish peroxidase; MAb, monoclonal antibody; nM, nanomolar; OD, optical density; pM, picomolar; RBD, receptor binding domain.

Using the SVNT assay, all these MAbs were found to dose-dependently block ACE2 binding to RBD, with IC50 values ranging from 6.3 to 48 µg/ml in SVNT (Figure 5A and Table 2). While the PVNT assay using SARS-CoV-2 pseudovirus (wild-type, D614G genotype) and human ACE2-overexpressing HEK293T cells gave IC50 values ranging from 0.9 to 38 µg/ml (Figures 5B,G and Table 2), the authentic CVNT assay using wild-type SARS-CoV-2 (wild-type, D614G genotype) and Vero 76 cell lines demonstrated IC50 values ranging from 0.1 to 0.6 µg/ml (Figure 5C and Table 2). Representative PVNT results are illustrated in Figure 5G. The IC50 results obtained from all three neutralization assays are depicted in Table 2. Among the six neutralizing MAbs, 1D1 and 1D10 were the most potent and 3G5 was the weakest in terms of neutralization potency (based on PVNT and CVNT results). Positive correlation for the IC50 values was observed between all three neutralization assays. While PVNT and CVNT demonstrated the highest correlation (p = 0.044), SVNT and CVNT (p = 0.105), and SVNT and PVNT (p = 0.058) displayed lower correlation (Figures 5D–F).

Neutralizing capacity of monoclonal antibodies on alpha and omicron variants

Severe acute respiratory syndrome coronavirus 2 VOCs are differentially mutated in different regions of the spike protein (Figure 6A). We initially evaluated binding reactivity of the selected neutralizing MAbs to the trimeric spike of SARS-CoV-2 VOCs (alpha, beta, gamma, and delta) compared with that of wild-type virus by ELISA. Our results showed no substantial increased or decreased reactivity of MAbs against alpha variant. Relative reactivities of 1D1, 1D10, 2C5, and 2D9 declined by more than 50% against beta and gamma variants. However, 3B6 and 3G5 revealed no substantial decline against beta and gamma variants (Figure 6B). We did not have access to the recombinant spike protein of omicron variant to perform the experiment and compare the results.

The neutralization efficiency of our MAbs against the omicron variant was assessed by CVNT assay in parallel with the alpha variant (Figure 7A). Consistent with our ELISA

10.3389/fmed.2022.973036

TABLE 2 Characteristics of anti-RBD MAbs.

No.	MAbs designation	Isotype	Affinity (nM)	Peptide reactivity	SVNT (inhibition %) 40 μ g/ml	PVNT (neutralization %) 15 μ g/ml	CVNT (neutralization %) 10 μ g/ml	SVNT (IC50 μ g/ml)	PVNT (IC50 μ g/ml)	CVNT* (IC50 μ g/ml)	IC50 fold change against omicron
1	**1D1**	IgG	0.49	–	100	100	100	8.5	1.8	0.1	64
2	**1D10**	IgG	0.57	–	100	100	100	9.2	0.9	0.1	62
3	**2C5**	IgG	0.51	–	87	56	100	14.5	12.4	0.3	12
4	**2D9**	IgG	0.53	–	100	100	100	6.3	1.6	0.2	22
5	**2F8**	IgG	1.71	P136–155	0	0	0	–	–	–	–
6	**2G3**	IgG	0.76	–	0	0	8	–	–	–	–
7	**3B6**	IgG	1.06	–	44	32	100	43.1	32	0.2	23.5
8	**3G5**	IgG	0.36	P76–95	8	28	100	48	38	0.6	18.7

*IC50 values were expressed as IC50 mean of three independent experiments. CNVT, conventional virus neutralization test; IgG, immunoglobulin G; MAb, monoclonal antibody; nM, nanomolar; PVNT, pseudovirus virus neutralization test; RBD, receptor binding domain; SVNT, surrogate virus neutralization test. The bold terms in the second column are the MAbs designations.

results, minor or no changes were found in IC50 values of MAbs against the alpha variant (Figures 7B,C). However, a substantial reduction of neutralization activity to the omicron variant was found for all MAbs. Specifically, the IC50 values of 1D1 and 1D10 increased 64- and 62-fold (IC50: 6.5 and 6.3 μg/ml, respectively), and the IC50 values of 2D9, 3B6, and 3G5 increased 22, 23, and 18-fold, respectively. Among the selected MAbs, 2C5 showed the smallest drop of neutralizing activity (IC50: 3.9 μg/ml) against omicron, with about 12-fold increase of IC50 compared to the WT virus (Figures 7B,C).

Discussion

Coronavirus disease 2019 pandemic has caused serious public health crisis during the last 2 years. Development of prophylactic and therapeutic MAbs may help to protect the patients at high risk of progression to severe COVID-19 against more pathogenic or transmissible SARS-CoV-2 variants. The S protein of SARS-CoV-2 is a key protein responsible for binding to ACE2. The RBD is the major immunodominant and immunoprotective region of the S protein which elicits potent virus neutralizing antibodies and has been used for the design and development of SARS-CoV-2 vaccines (30–32). Such neutralizing antibodies are widely detected in serum of COVID-19 patients (33–35). RBM is a part of RBD which directly binds to the ACE2 receptor and could be considered as a promising target for generating neutralizing antibodies (36).

In the present study, we generated and characterized a panel of murine neutralizing MAbs against SARS-CoV-2 using recombinant RBD protein as the immunogen (Figure 1). This approach has also been used in other studies (37–41). We investigated epitope specificity of the serum antibody as well as all preliminary growing hybridomas by Pepscan. Our results showed that the antibody response in BALB/c mice is largely directed against linear epitopes. These findings are different from our previous results obtained from COVID-19 convalescent sera which showed that the antibody response is mainly directed against conformational epitopes of RBD in human (42).

Our Pepscan results showed that P76–91, P91–110, and especially P136–155 peptides account for a fraction of the neutralizing antibody pool in immunized mouse sera based on our peptide adsorption assays (Figures 2C–F). This finding is consistent with a recent study in which epitope profiling was performed on a panel of sera from RBD-immunized mice. They demonstrated immunogenicity of R465 (overlapping with P136–155) and R405 (overlapping with P76–95) peptides (41). In another study, 33 predicted linear epitopes of spike were applied to immunize BALB/c mice of which two peptides (S406–420 and S455–469) overlapping with our P76–110 (aa 394–428 of spike) and P136–155 (aa 454–473 of spike) peptides, elicited robust antibody responses against S protein (43).

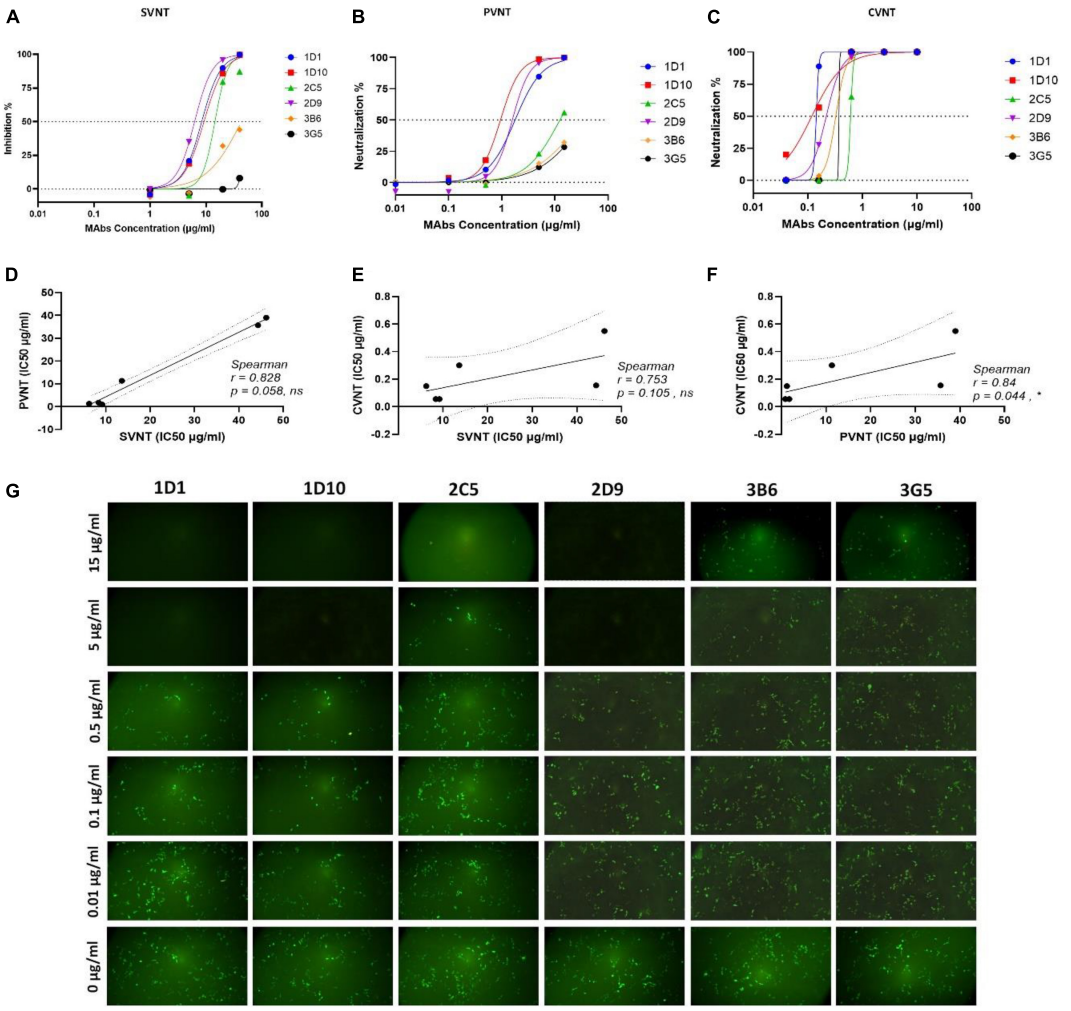


FIGURE 5
Neutralizing activity of MAbs was determined by PVNT, SVNT, and CVNT. **(A)** Selected RBD-specific MAbs were assessed for neutralization by SVNT. HRP-conjugated ACE2 protein was used to determine its binding to immobilized RBD in the presence of MAbs at the concentrations of 1, 5, 20, and 40 µg/ml. The percentage (%) of inhibition was calculated based on the OD of HRP-conjugated ACE2 with or without presence of RBD-specific MAbs. Dose-response curves are presented for each MAb. The IC50 value was calculated by non-linear regression (four-parameter), as represented in Table 2. **(B)** Dose-response curves of selected MAbs determined by PVNT. Five different concentrations of selected MAbs, 0.01, 0.1, 0.5, 5, and 15 µg/ml, were incubated with pseudovirus and added to HEK293T-hACE2 cells (14×10^3 cells/well). ImageJ software was used to calculate the fluorescence positive cells 60 h post-infection to calculate IC50 values and the inhibition ratios. The IC50 was calculated by non-linear regression (four-parameter), as represented in Table 2. **(C)** Representative dose-response curves of selected MAbs determined by CVNT. The selected MAbs were diluted to different concentrations, incubated with the WT D614G SARS-CoV-2 to reach the final concentrations of 0.04, 0.16, 0.63, 2.5, and 10 µg/ml, and added to Vero 76 cells. The percentage of infected cells was calculated by counting nucleocapsid positive cells versus total cells. The IC50 was calculated by non-linear regression (four-parameter), as represented in Table 2. The dotted lines on each graph indicate 0 and 50% neutralization. **(D–F)** The correlation of obtained IC50 values between SVNT, PVNT, and CVNT was calculated by Spearman analysis. **(G)** Representative fluorescence images of HEK293T cells expressing hACE2 after infection with eGFP-pseudotyped lentiviruses in the absence or presence of different concentrations of anti-RBD MAbs. ACE2, angiotensin-converting enzyme-2; CVNT, conventional virus-neutralizing test; HRP, horse-radish peroxidase; IC50, inhibition concentration 50%; MAb, monoclonal antibody; OD, optical density; PVNT, pseudovirus neutralization test; RBD, receptor binding domain; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SVNT, surrogate virus-neutralizing test.

Interestingly, the P136–155 (RLFRKSNLKPFERDISTEI) peptide, highly recognized by our serum samples and hybridoma supernatants (Figures 2, 3), is located in RBM according to structural analysis of the RBD-ACE2 complex and was shown to elicits neutralizing antibodies in COVID-19 patients (36). One of our final selected MAbs (2F8) reacted with this peptide. Although 2F8 binds to this peptide, it possesses a low affinity (1.71 nM), which may account for its lack of neutralization activity.

The other peptide-reactive MAb, 3G5, which recognizes P76–95 (NVYADSFVIRGDEVRQIAPG) peptide, located in the core subdomain of RBD (aa 394–413), displayed weak

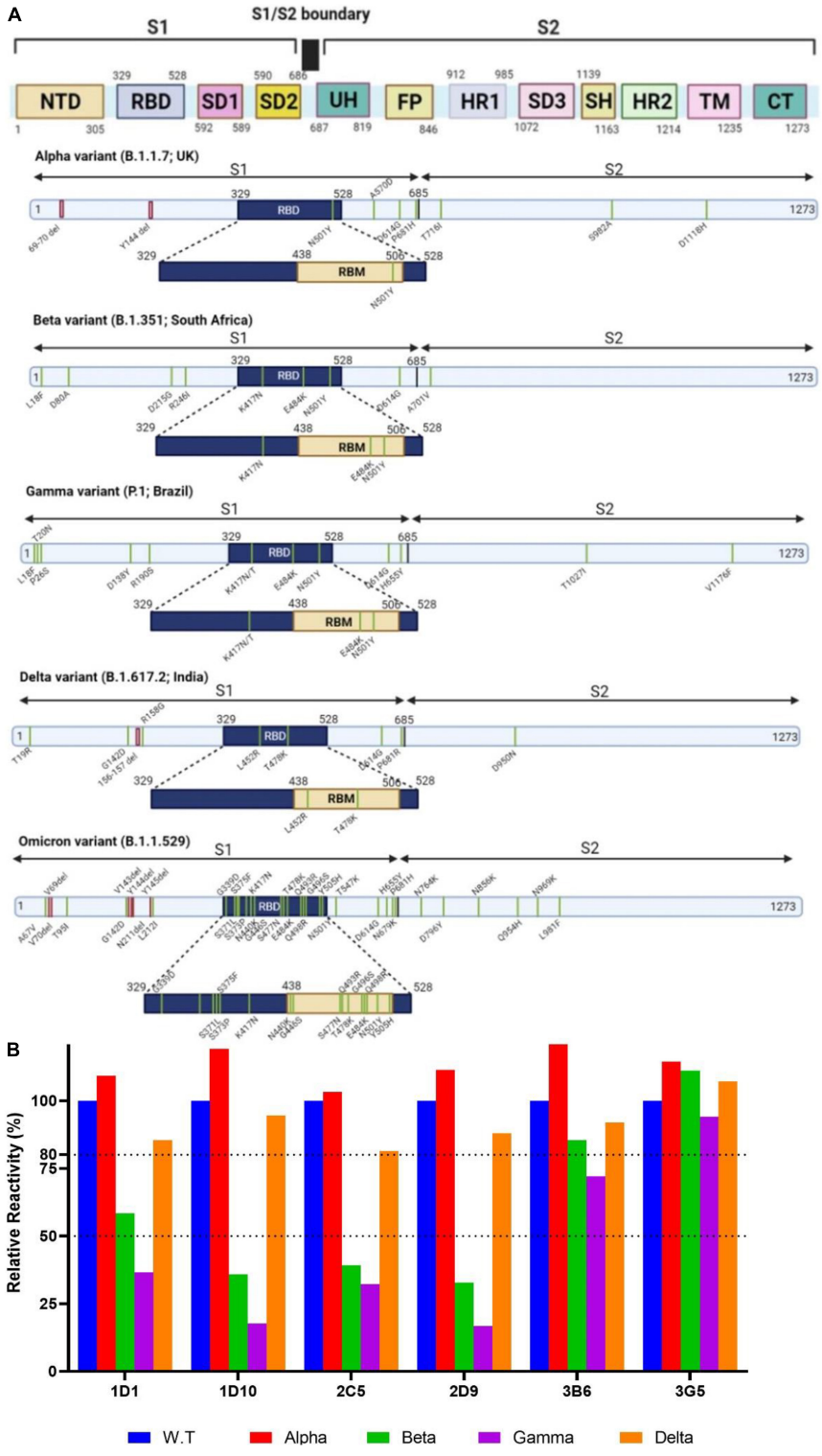


FIGURE 6
Schematic overview of the spike proteins from the SARS-CoV-2 VOCs and binding activity of MAbs against VOCs. **(A)** The primary structure of the spike protein and the location of the mutations in the context of the spike protein are shown in distinct panels. **(B)** Relative binding reactivity of selected MAbs was measured against the trimeric spike of SARS-CoV-2 alpha (red), beta (green), gamma (purple), and delta (brown) variants compared with that of wild-type (blue) by ELISA. ELISA, enzyme-linked immunosorbent assay; FP, fusion peptide; HR, heptad repeat; MAb, monoclonal antibody; NTD, N-terminal domain; RBD, receptor binding domain; RBM, receptor-binding motif; S, spike; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SD, subdomain; SH, stem helix; TM, transmembrane; UH, upstream helix; VOC, variant of concern.

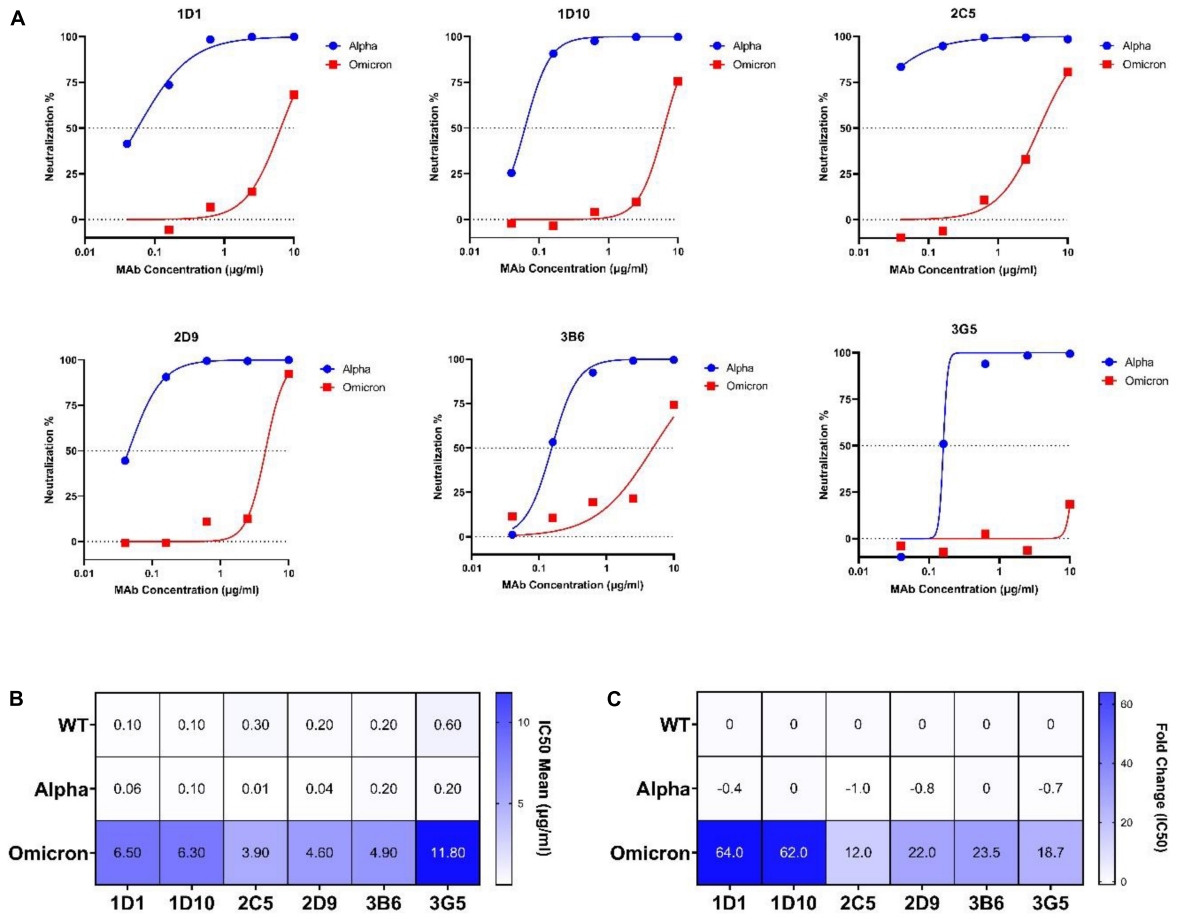


FIGURE 7
Neutralization potency of selected MAbs against alpha and omicron variants. **(A)** Dose-response curves and IC50 values of selected MAbs at different concentrations (0.04, 0.16, 0.63, 2.5, and 10 µg/ml) against alpha and omicron variants determined by CVNT. The dotted lines on each graph indicate 0 and 50% neutralization. The IC50 was calculated by non-linear regression (four-parameter), as represented in **(B)**. **(B)** Heatmap of IC50 mean values of selected MAbs against WT (wild-type), alpha, and omicron viruses. **(C)** Heatmap of calculated IC50 fold changes of selected MAbs against alpha and omicron variant obtained from CVNT based on IC50 values against WT virus. The increase or decrease in IC50 values relative to WT is shown as positive or negative fold change values, respectively. CVNT, conventional virus-neutralizing test; IC50, inhibition concentration 50%; MAb, monoclonal antibody; VOC, variant of concern.

neutralizing potency against WT SARS-CoV-2 pseudovirus. Recent studies have reported two MAbs, CB6 and B38, recognizing residues within P76–95 as well as RBM, which displayed neutralizing activity and were able to completely abolish ACE2/RBD binding (44, 45). These MAbs may probably elicit steric hindrance or allosteric effects for binding to ACE2.

Five of our MAbs, including 1D1, 1D10, 2C5, 2D9, and 2G3, displayed similar binding competition patterns, but different neutralizing activities (**Figures 4B**, **5**). These five MAbs showed different epitope specificity in immunoblotting assay (**Figure 4C**). While 1D1, 1D10, and 2D9 recognized both reduced and non-reduced RBD, 2C5 and 2G3 did not bind to the reduced form of RBD. We speculate that the orientation of binding of these antibodies to their target epitopes may influence their neutralizing capacity, a finding also reported by other investigators (46, 47).

The emergence of immune resistant variants harboring escape mutations in response to the immune pressure is an important issue that must be taken into consideration to control COVID-19. Emerging variants are categorized as either variants of interest (VOI) or VOC (48). Five variants are classified as VOC, including alpha variant (B.1.1.7) containing N501Y substitution in RBD, beta variant (B.1.351) containing three important mutations in RBD, including N501Y, E484K, and K417N, gamma variant (P.1) with biologically important mutations in the RBD region, including N501Y, E484K, and K417N/T, delta variant (B.1.617.2) harboring two substitutions in RBD, including L452R and T478K associated with its higher transmissibility (49–52), and omicron variant (B.1.1.529) harboring 34 mutations, 15 of which are in the RBD region, leading to fourfold increased infectivity compared with the WT SARS-CoV-2 (**Figure 6A**). Alpha and beta variants are

significantly more transmissible (43–82 and 50%, respectively), compared to Wuhan SARS-CoV-2 virus (53, 54), due to N501Y substitution that enhances the accessibility of RBD and binding affinity to ACE2 (53, 55–57). Although K417N/T substitutions found in beta and gamma variants decreased the binding affinity, N501Y and E484K mutations enhanced the binding affinity of their RBDs to ACE2 (58).

Recently, an extensive study was conducted on MAbs authorized for emergency use by the FDA-EU to assess their neutralizing activities against the current VOCs. The results showed slightly reduced neutralizing activity of sotrovimab against the alpha variant due to N501Y mutation. Accordingly, the neutralizing potency of bamlanivimab and casirivimab was completely or significantly lost against the beta variant because of E484K and K417N substitutions. Bamlanivimab lost the neutralization effect against beta variant carrying the E484K substitution (59). Another study reported that yeast expressing mutant RBD harboring the E484K and K417N/T substitutions escaped bamlanivimab and etesevimab, respectively (60). However, imdevimab maintained its neutralization activity against alpha and beta variants (61). Therefore, the emergence of mutations similar to alpha and beta variants is considered an important challenge for therapeutic MAbs. In accordance with these findings, our neutralizing MAbs displayed similar ELISA binding and neutralization against WT and alpha strains (**Figures 6B**, 7), indicating that our MAbs are insensitive to the N501Y mutation of alpha variant. Of note, the binding of 1D1, 1D10, 2C5, and 2D9 to the trimeric spike of beta and gamma variants showed a substantial decline compared with the WT trimeric spike (**Figure 6B**), implying contribution of K417 and E484 mutations in these two variants. Similarly, a panel of anti-RBD human MAbs showed decreased binding reactivity and neutralizing activity to spike and authentic virus of beta and gamma variants, as compared to the alpha variant (62). Analysis of a panel of MAbs, including COV2-2196, COV2-2130, sotrovimab, casirivimab, imdevimab, bamlanivimab, and etesevimab, demonstrated no significant changes in neutralizing activity against beta and delta variants, except for imdevimab and bamlanivimab, which displayed 10-fold decrease and complete loss of binding against the delta variant, respectively (63). Our ELISA results indicate that none of our MAbs lost substantial binding against the delta trimeric spike (**Figure 6B**), implying that in contrast to mutations in RBD of beta and gamma variants, the substitutions in delta variant, including L452R and T478K, might not be critical for reactivity of our MAbs. A chimeric MAb with low binding and neutralization potency against the spike proteins of alpha, beta, and gamma variants has recently been reported which binds and neutralizes the delta variant potently (64). Based on our knowledge from the literature, there is an association between binding ability of MAbs to the critical residues of spike protein of VOCs and their neutralizing potency (62, 64). However, we cannot attribute this association to our MAbs, unless we confirm it by PVNT or CVNT assays. Unfortunately, we could not check the neutralization potency of our MAbs against beta, gamma, and delta variants, because we had no access to these variants.

Assessment of the neutralization potency of MAbs against the latest VOC Omicron is key due to its extensive mutations within the RBD and S proteins. In our panel of MAbs, 1D1 and 1D10, which showed potent neutralization activity against the authentic WT virus, were still able to neutralize the omicron variant, although a more than 70-fold increase of IC50 was observed. Notably, 2C5 with a higher IC50 value against the WT virus was more resilient to the authentic omicron variant (16-fold increase of IC50) compared to 1D1 and 1D10 (**Figure 7**). Recent study has reported that sotrovimab revealed a threefold reduction, the combination of COV2-2130 and COV2-2196 showed a ∼200-fold decline, and casirivimab, imdevimab, bamlanivimab, etesevimab, and CT-P59 completely lost the neutralizing function against omicron (10). Interestingly, sotrovimab maintained neutralizing potency against the omicron variant (65).

Conclusion

We produced and characterized new SARS-CoV-2 RBD-specific murine MAbs exhibiting distinct epitope binding and neutralization potency against different VOCs of SARS-CoV-2. Our MAbs, therefore, represent new additional MAb candidates for SARS-CoV-2 inhibition. We are planning to chimerize a number of our potent neutralizing MAbs, to be able to compare their functional and structural properties with their mouse counterparts and to evaluate their therapeutic effects in a preclinical and clinical settings. Understanding the interactions between these MAbs and the RBD epitopes and identifying engaged residues is required for the design of new immunogens for the development of new generation of vaccines protective for a broad spectrum of SARS-CoV-2 variants. The anti-RBD MAbs could also potentially be used to design highly specific and sensitive immunoassays to detect viral particles of SARS-CoV-2 variants in patients samples as a diagnostic tool, in combination with MAbs against other structural proteins of SARS-CoV-2, such as nucleoprotein. However, due to the extensive mutations accumulated in the spike and RBD proteins progressively, as opposed to the anti-nucleoprotein MAbs, the anti-RBD MAbs lose their affinity and reactivity which limits their application for diagnostic purposes.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was reviewed and approved by the Research Ethics Committee of Tehran University of Medical Sciences (IR.TUMS.SPH.REC.1400.167) and the National Institute for Medical Research Development (NIMAD) of Iran (IR.NIMAD.REC.1399.194).

Author contributions

FM performed the assays and wrote the original draft. JK, MM, HA, SS, and GL contributed to the performing the assays. A-HZ, VS, GK, AG, and MJ-T contributed to the conceptualization and design of the study. MJ-T reviewed and edited the manuscript. MA, CM, and FS contributed to the project conceptualization, data validation, project administration, supervision, and reviewing and editing of the manuscript. All authors have read and agreed to the published version of the article.

Funding

This study was supported by a grant from Tehran University of Medical Science (TUMS) of Iran (Grant No. 1400-2-99-54802), the National Institute for Medical Research Development (NIMAD) of Iran (Grant No. 993421), INSERM U1109, and ANRS COVID Sud project (Grant No. ECTZ144757).

Acknowledgments

We thank Renap Therapeutics Co., Iran, for providing pseudotyped lentiviruses and the hACE2-HEK293T cell line.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. World Health Organization. *Coronavirus disease (COVID-19) pandemic*. Geneva: World Health Organization (2019).

2. World Health Organization. *Weekly epidemiological update on COVID-19 - 27 April 2022*. Geneva: World Health Organization (2022).

3. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol.* (2020) 28:1–16. doi: 10.1038/s41579-020-00468-6

4. Zheng JSARS-. CoV-2: An emerging coronavirus that causes a global threat. *Int J Biol Sci.* (2020) 16:1678–85. doi: 10.7150/ijbs.45053

5. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* (2020) 19:141–54. doi: 10.1038/s41579-020-00459-7

6. Samrat SK, Tharappel AM, Li Z, Li H. Prospect of SARS-CoV-2 spike protein: Potential role in vaccine and therapeutic development. *Virus Res.* (2020) 288:1–18. doi: 10.1016/j.virusres.2020.198141

7. Huang Y, Yang C, Xu X-f, Xu W, Liu S-w. Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol Sin.* (2020) 41:1141–9. doi: 10.1038/s41401-020-0485-4

8. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* (2020) 581:215–20. doi: 10.1038/s41586-020-2180-5

9. Yi C, Sun X, Ye J, Ding L, Liu M, Yang Z, et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol.* (2020) 17:621–30. doi: 10.1038/s41423-020-0458-z

10. Cameroni E, Bowen JE, Rosen LE, Saliba C, Zepeda SK, Culap K, et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature.* (2021) 602:664–70. doi: 10.1038/s41586-021-04386-2

11. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses.* (2020) 12:1–15. doi: 10.1101/2020.02.03.933226

12. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* (2020) 181:271–80. doi: 10.1016/j.cell.2020.02.052

13. Wang C, Li W, Drabek D, Okba N, van Haperen R, Osterhaus AD, et al. A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat Commun.* (2020) 11:1–6. doi: 10.1038/s41467-020-16256-y

14. Zeng F, Hon CC, Yip CW, Law KM, Yeung YS, Chan KH, et al. Quantitative comparison of the efficiency of antibodies against S1 and S2 subunit of SARS coronavirus spike protein in virus neutralization and blocking of receptor binding: Implications for the functional roles of S2 subunit. *FEBS Lett.* (2006) 580:5612–20. doi: 10.1016/j.febslet.2006.08.085

15. Tong P-B-V, Lin L-Y, Tran TH. Coronaviruses pandemics: Can neutralizing antibodies help? *Life Sci.* (2020) 255:1–11. doi: 10.1016/j.lfs.2020.117836

16. Lang AB, Cryz SJ Jr., Schurch U, Ganss MT, Bruderer U. Immunotherapy with human monoclonal antibodies. Fragment A specificity of polyclonal and monoclonal antibodies is crucial for full protection against tetanus toxin. *J Immunol.* (1993) 151:466–72.

17. Pelletier JPR, Mukhtar F. Passive monoclonal and polyclonal antibody therapies. In: Maitta RW editor. *Immunologic Concepts in Transfusion Medicine*. Amsterdam: Elsevier Health Sciences (2020). p. 251–348. doi: 10.1016/B978-0-323-67509-3.00016-0

18. Esmaeilzadeh A, Elahi R. Immunobiology and immunotherapy of COVID-19: A clinically updated overview. *J Cell Physiol.* (2020) 2020:1–25. doi: 10.1002/jcp.30076

19. AminJafari A, Ghasemi S. The possible of immunotherapy for COVID-19: A systematic review. *Int Immunopharmacol.* (2020) 83:1–5. doi: 10.1016/j.intimp.2020.106455

20. Ahangarzadeh S, Payandeh Z, Arezumand R, Shahzamani K, Yarian F, Alibakhshi A. An update on antiviral antibody-based biopharmaceuticals. *Int Immunopharmacol.* (2020) 86:1–8. doi: 10.1016/j.intimp.2020.106760

21. Salazar G, Zhang N, Fu T-M, An Z. Antibody therapies for the prevention and treatment of viral infections. *NPJ Vaccines.* (2017) 2:1–12. doi: 10.1038/s41541-017-0019-3

22. Fda. *Coronavirus (COVID-19) update: FDA authorizes additional monoclonal antibody for treatment of COVID-19*. Silver Spring, MA: FDA (2021).

23. Fda. *Coronavirus (COVID-19) update: FDA authorizes monoclonal antibodies for treatment of COVID-19*. Silver Spring, MA: FDA (2020).

24. Fda. *FDA authorizes bamlanivimab and etesevimab monoclonal antibody therapy for post-exposure prophylaxis (prevention) for COVID-19*. Silver Spring, MA: FDA (2021).

25. Fda. *Casirivimab and imdevimab EUA letter of authorization*. Silver Spring, MA: FDA (2020).

26. Fda. *Sotrovimab EUA letter of authorization*. Silver Spring, MA: FDA (2022).

27. Beatty JD, Beatty BG, Vlahos WG. Measurement of monoclonal antibody affinity by non-competitive enzyme immunoassay. *J Immunol Methods.* (1987) 100:173–9. doi: 10.1016/0022-1759(87)90187-6

28. Hajighasemi F, Sabour-Yaraghi A, Shokri F. Measurement of affinity constant of anti-human IgG monoclonal antibodies by an ELISA-based method. *Iran J Immunol.* (2004) 1:154–61.

29. Ferrara F, Temperton N. Pseudotype neutralization assays: From laboratory bench to data analysis. *Methods Protoc.* (2018) 1:8. doi: 10.3390/mps1010008

30. Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, et al. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature.* (2020) 586:572–7. doi: 10.1038/s41586-020-2599-8

31. Tai W, Zhang X, Drelich A, Shi J, Hsu JC, Luchsinger L, et al. A novel receptor-binding domain (RBD)-based mRNA vaccine against SARS-CoV-2. *Cell Res.* (2020) 30:932–5. doi: 10.1038/s41422-020-0387-5

32. Liu Z, Xu W, Xia S, Gu C, Wang X, Wang Q, et al. RBD-Fc-based COVID-19 vaccine candidate induces highly potent SARS-CoV-2 neutralizing antibody response. *Signal Transduct Target Ther.* (2020) 5:282. doi: 10.1038/s41392-020-00402-5

33. Li Y, Lai D-y, Zhang H-n, Jiang H-w, Tian X, Ma M-l, et al. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol.* (2020) 17:1095–7. doi: 10.1038/s41423-020-00523-5

34. Jiang S, Zhang X, Yang Y, Hotez PJ, Du L. Neutralizing antibodies for the treatment of COVID-19. *Nat Biomed Eng.* (2020) 4:1134–9. doi: 10.1038/s41551-020-00660-2

35. Brouwer PJM, Caniels TG, van der Straten K, Snitselaar JL, Aldon Y, Bangaru S, et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science.* (2020) 369:643. doi: 10.1126/science.abc5902

36. Wang H, Wu X, Zhang X, Hou X, Liang T, Wang D, et al. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *ACS Central Sci.* (2020) 6:2238–49. doi: 10.1021/acscentsci.0c00742

37. Xiong HL, Wu YT, Cao JL, Yang R, Liu YX, Ma J, et al. Robust neutralization assay based on SARS-CoV-2 S-protein-bearing vesicular stomatitis virus (VSV) pseudovirus and ACE2-overexpressing BHK21 cells. *Emerg Microbes Infect.* (2020) 9:2105–13. doi: 10.1080/22221751.2020.1815589

38. Chapman AP, Tang X, Lee JR, Chida A, Mercer K, Wharton RE, et al. Rapid development of neutralizing and diagnostic SARS-CoV-2 mouse monoclonal antibodies. *Sci Rep.* (2021) 11:9682. doi: 10.1038/s41598-021-88809-0

39. Amanat F, Strohmeier S, Lee WH, Bangaru S, Ward AB, Coughlan L, et al. Murine monoclonal antibodies against the receptor binding domain of SARS-CoV-2 neutralize authentic wild-type SARS-CoV-2 as well as B.1.1.7 and B.1.351 viruses and protect in vivo in a mouse model in a neutralization-dependent manner. *mBio.* (2021) 12:e0100221. doi: 10.1128/mBio.01002-21

40. Antipova NV, Larionova TD, Siniavin AE, Nikiforova MA, Gushchin VA, Babichenko II, et al. Establishment of murine hybridoma cells producing antibodies against spike protein of SARS-CoV-2. *Int J Mol Sci.* (2020) 21:9167. doi: 10.3390/ijms21239167

41. Jiang M, Zhang G, Liu H, Ding P, Liu Y, Tian Y, et al. Epitope profiling reveals the critical antigenic determinants in SARS-CoV-2 RBD-based antigen. *Front Immunol.* (2021) 12:707977. doi: 10.3389/fimmu.2021.707977

42. Maghsood F, Shokri MR, Jeddi-Tehrani M, Torabi Rahvar M, Ghaderi A, Salimi V, et al. Identification of immunodominant epitopes on nucleocapsid and spike proteins of the SARS-CoV-2 in Iranian COVID-19 patients. *Pathog Dis.* (2022) 80:ftac001. doi: 10.1093/femspd/ftac001

43. Lu S, Xie X-x, Zhao L, Wang B, Zhu J, Yang T-r, et al. The immunodominant and neutralization linear epitopes for SARS-CoV-2. *Cell Rep.* (2021) 34:108666. doi: 10.1016/j.celrep.2020.108666

44. Shi R, Shan C, Duan X, Chen Z, Liu P, Song J, et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature.* (2020) 584:120–4. doi: 10.1038/s41586-020-2381-y

45. Wu Y, Wang F, Shen C, Peng W, Li D, Zhao C, et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science.* (2020) 368:1274–8. doi: 10.1126/science.abc2241

46. Barnes CO, Jette CA, Abernathy ME, Dam K-MA, Esswein SR, Gristick HB, et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature.* (2020) 588:682–7. doi: 10.1038/s41586-020-2852-1

47. Finkelstein MT, Mermelstein AG, Parker Miller E, Seth PC, Stancofski, Fera D. Structural analysis of neutralizing epitopes of the SARS-CoV-2 spike to guide therapy and vaccine design strategies. *Viruses.* (2021) 13:134. doi: 10.3390/v13010134

48. World Health Orgnization. *Tracking SARS-CoV-2 variants*. Geneva: World Health Orgnization (2022).

49. Adam D. The rush to study fast spreading coronavirus variants. *Nature.* (2021) 594:19–20. doi: 10.1038/d41586-021-01390-4

50. Mahase E. Delta variant: What is happening with transmission, hospital admissions, and restrictions? *Br Med J Publ Group.* (2021) 373:n1513. doi: 10.1136/bmj.n1513

51. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance.* (2021) 26:2100509. doi: 10.2807/1560-7917.ES.2021.26.24.2100509

52. Motozono C, Toyoda M, Zahradnik J, Saito A, Nasser H, Tan TS, et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe.* (2021) 29:1124.e–36.e. doi: 10.1016/j.chom.2021.06.006

53. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv* [Preprint] (2020). doi: 10.1101/2020.12.21.20248640

54. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science.* (2021) 372:eabg3055. doi: 10.1126/science.abg3055

55. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7. *Nature.* (2021) 593:270–4. doi: 10.1038/s41586-021-03426-1

56. Team EE. Updated rapid risk assessment from ECDC on the risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA–first update. *Eurosurveillance.* (2021) 26:2101211. doi: 10.2807/1560-7917.ES.2021.26.3.2101211

57. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* (2020) 182:1295.–1310. doi: 10.1016/j.cell.2020.08.012

58. Barton MI, MacGowan S, Kutuzov M, Dushek O, Barton GJ, van der Merwe PA. Effects of common mutations in the SARS-CoV-2 Spike RBD domain and its ligand the human ACE2 receptor on binding affinity and kinetics. *Elife* (2021) 10:e70658. doi: 10.7554/eLife.70658

59. Widera M, Wilhelm A, Hoehl S, Pallas C, Kohmer N, Wolf T, et al. Bamlanivimab does not neutralize two SARS-CoV-2 variants carrying E484K in vitro. *medRxiv* [Preprint] (2021). doi: 10.1101/2021.02.24.21252372

60. Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep Med.* (2021) 2:100255. doi: 10.1016/j.xcrm.2021.100255

61. Tada T, Dcosta BM, Zhou H, Vaill A, Kazmierski W, Landau NR. Decreased neutralization of SARS-CoV-2 global variants by therapeutic anti-spike protein monoclonal antibodies. *bioRxiv* [Preprint] (2021). doi: 10.1101/2021.02.18.431897

62. Wheatley AK, Pymm P, Esterbauer R, Dietrich MH, Lee WS, Drew D, et al. Landscape of human antibody recognition of the SARS-CoV-2 receptor binding domain. *Cell Rep.* (2021) 37:109822. doi: 10.1016/j.celrep.2021.109822

63. Chen RE, Winkler ES, Case JB, Aziati ID, Bricker TL, Joshi A, et al. In vivo monoclonal antibody efficacy against SARS-CoV-2 variant strains. *Nature.* (2021) 596:103–8.

64. Liang KH, Chiang PY, Ko SH, Chou YC, Lu RM, Lin HT, et al. Antibody cocktail effective against variants of SARS-CoV-2. *J Biomed Sci.* (2021) 28:80.

65. Gruell H, Vanshylla K, Tober-Lau P, Hillus D, Schommers P, Lehmann C, et al. mRNA booster immunization elicits potent neutralizing serum activity against the SARS-CoV-2 Omicron variant. *Nat Med.* (2022) 28:477–80.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fimmu.2022.1004804

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Shaza Shantier,
University of Khartoum, Sudan
Abdelrahman Hamza Abdelmoneim,
Al-Neelain University, Sudan

*CORRESPONDENCE
Muhammad Waqas
mwaqas@unizwa.edu.om
Amjad Ali
amjad.genetics@hu.edu.pk

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Vaccines and Molecular Therapeutics, a section of the journal Frontiers in Immunology

RECEIVED 27 July 2022
ACCEPTED 27 September 2022
PUBLISHED 13 October 2022

CITATION
Aziz S, Almajhdi FN, Waqas M, Ullah I, Salim MA, Khan NA and Ali A (2022) Contriving multi-epitope vaccine ensemble for monkeypox disease using an immunoinformatics approach. *Front. Immunol.* 13:1004804. doi: 10.3389/fimmu.2022.1004804

Contriving multi-epitope vaccine ensemble for monkeypox disease using an immunoinformatics approach

Shahkaar Aziz[1†], Fahad Nasser Almajhdi[2†], Muhammad Waqas[3,4*], Inam Ullah[3], Muhammad Adil Salim[5,6], Nasir Ali Khan[4] and Amjad Ali[3*]

[1]Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, Pakistan, [2]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia, [3]Department of Biotechnology and genetic Engineering, Hazara University, Mansehra, Pakistan, [4]Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, Oman, [5]Microbiology Graduate Group, University of California, Davis, Davis, CA, United States, [6]Genome Center, University of California, Davis, Davis, CA, United States

The current global outbreak of monkeypox (MPX) disease, caused by Monkeypox virus (MPXV), has resulted in 16 thousand infection cases, five deaths, and has been declared a global health emergency of international concern by the World Health Organization. Given current challenges in the safety of existing vaccines, a vaccine to prevent MPX infection and/or onset of symptoms would significantly advance disease management. In this context, a multi-epitope-based vaccine could be a well-suited approach. Herein, we searched a publicly accessible database (Virus Pathogen Database and Analysis Resource) for MPXV immune epitopes from various antigens. We prioritized a group of epitopes (10 CD8+ T cells and four B-cell epitopes) using a computer-aided technique based on desirable immunological and physicochemical properties, sequence conservation criteria, and non-human homology. Three multi-epitope vaccines were constructed (MPXV-1–3) by fusing finalized epitopes with the aid of appropriate linkers and adjuvant (beta-defensin 3, 50S ribosomal protein L7/L12, and Heparin-binding hemagglutinin). Codon optimization and *in silico* cloning in the pET28a (+) expression vector ensure the optimal expression of each construct in the *Escherichia Coli* system. Two and three-dimensional structures of the constructed vaccines were predicted and refined. The optimal binding mode of the construct with immune receptors [Toll-like receptors (TLR2, TLR3, and TLR4)] was explored by molecular docking, which revealed high docking energies of MPXV-1–TLR3 (–99.09 kcal/mol), MPXV-2–TLR3 (–98.68 kcal/mol), and MPXV-3–TLR2 (–85.22 kcal/mol). Conformational stability and energetically favourable binding of the vaccine-TLR2/3 complexes were assessed by performing molecular dynamics simulations and free energy calculations (Molecular Mechanics/Generalized Born Surface Area method). *In silico* immune simulation suggested that innate, adaptive, and humoral responses will be elicited upon administration of such potent multi-epitope vaccine constructs. The vaccine

constructs are antigenic, non-allergen, non-toxic, soluble, topographically exposed, and possess favourable physicochemical characteristics. These results may help experimental vaccinologists design a potent MPX vaccine.

KEYWORDS

monkeypox, monkeypox virus, multi-epitope vaccine, *in silico* vaccine, immunoinformatics, B-cell epitopes, CTL epitopes

Introduction

Monkeypox (MPX) is an emerging zoonotic disease caused by monkeypox virus (MPXV), a member of the *Orthopoxvirus* genus in the family *Poxviridae* (1). The same genus also contains variola (etiological agent of smallpox), vaccinia, and cowpox virus (2). The MPXV is a double-stranded DNA virus with a genome of ~19.7 kb long that codes for 190 open reading frames, which make up most of the material required for viral replication in the cytoplasm of cells (3). MPXV strains are divided into two clades with ~ 0.5% difference in their genomic sequence and circulate in different African regions. The viral isolates of the Central African (Congo Basin) clade are more virulent than the West African clade in humans (4).

The virus was first discovered in monkeys in a Danish laboratory in 1958; thus, it was named 'monkeypox'. The first human instance was reported in 1970 in the Democratic Republic of the Congo (DRC). Since then, MPX has become endemic in the DRC and has spread to other African nations (5). As of 23rd July 2022, the recent outbreak of MPX has resulted in 16,000 confirmed infection cases and five deaths, as reported to the World Health Organization (WHO) from 75 countries in five of the six WHO regions since 1 January. The global outbreak of MPX now represents a public health emergency of international concern as per the WHO declaration (6). Recent shotgun metagenomics revealed that the MPXV outbreak variant segregates in a significantly divergent phylogenetic branch (~50SNPs), probably signifying a recent evolutionary shift (7).

A diverse group of animal species has been found susceptible to MPXV infection; however, the native host remains unknown. The transmission routes of infection are thought to be saliva/respiratory excretions and contact with lesion exudate or crust material. Another route of viral can be viral shedding *via* faeces (1). Clinical manifestations of MPX are quite like smallpox, but the early lymph node enlargement, often at the onset of fever, distinguishes monkeypox from smallpox. A rash occurs 1–3 days following the onset of fever and lymphadenopathy, with lesions emerging simultaneously and developing at the same rate. Their distribution is mostly peripheral, but it may spread throughout the whole body during a severe illness. Until the lesion desquamates, the infection might linger up to 4 weeks (8). Secondary bacterial infections, respiratory distress, bronchopneumonia, gastrointestinal involvement, dehydration, sepsis, encephalitis, and corneal infection with resulting loss of vision are among the consequences that may occur in patients with MPX (1).

Smallpox vaccination with vaccinia virus (another *orthopoxvirus*) was shown to be nearly 85% protective against MPX in the past (9). Routine smallpox vaccination was terminated after the elimination of smallpox in 1980 (10), and no *orthopoxvirus* vaccination program has been undertaken in almost 40 years. A smallpox vaccine, Dryvax, has also been shown to protect against monkeypox. Nevertheless, the vaccinee and others who come into contact with the vaccinee can suffer from diverse side effects (11). Two vaccines licensed by the U.S. Food and Drug Administration (FDA) for smallpox are available for preventing MPX infection: ACAM2000 and JYNNEOS (also known as Imvamune or Imvanex) (12). The former is a live, replication-competent vaccinia virus (13). Because of this, there is a possibility for adverse side effects (such as progressive vaccinia and eczema vaccinatum) and myopericarditis (estimated incidence of 5.7 per 1,000 primary vaccinees based on clinical trial data); human to human transmission of vaccinia virus can also occur (13, 14). The latter is a replication-deficient modified vaccinia Ankara vaccine recommended for use in preexposure prophylaxis in persons at occupational risk for exposure to orthopoxviruses (13, 15). However, since the mechanism for myopericarditis caused by ACAM2000 is assumed to be immune-mediated, it is unknown if the antigen or antigens that precipitate autoantibodies are also present in JYNNEOS (14). There is no evidence on the efficacy of these vaccinations in the present MPXV outbreak (12).

Abbreviations: CTL, Cytotoxic T lymphocytes; FEL, free energy landscape; MHC, major histocompatibility complex; MM-GBSA, Molecular mechanics/generalized born surface area; MPX, monkeypox; MPXV, Monkeypox virus; Rg, radius of gyration; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SASA, solvent accessible surface area; TLR, toll-like immune receptors.

Heraud et al. revealed protection against MPX by recombinant subunit vaccine (DNA/protein immunization) that provoked helper T-cell (HTL) response and induced the production of limited B-cell epitopes/peptides in the sera of the immunized macaques (11). Franceschi et al. (16) tested a novel vaccination technique on STAT1 mutant mice to examine whether it protects against a fatal MPXV challenge using three recombinant Bovine Herpesvirus 4 (BoHV-4) vectors, each of which expressed a different MPXV glycoprotein, A29L, M1R, or B6R. Their research showed that BoHV-4-based vectors are effective and may be used as a platform for vector-vaccine development. There are, however, no FDA-approved treatments for MPX, and existing vaccinations are restricted owing to safety concerns, and the patients are provided with supportive care and indicative treatment. As a result, new investigations on pathogenesis, prevention, and treatments continue to spark the scientific community's attention (16). Developing specific therapeutics, including a vaccine against MPX, could help prevent and treat the disease.

Immune responses are crucial to combat viral infection, and a fundamental unit that generates either a cellular or a humoral immune response is called an antigenic epitope. Thus, a multi-epitope vaccine comprising a collection of peptides is a promising strategy to prevent and treat viral infection (17). Compared to experiment-based techniques, developing a multi-epitope-based vaccination is more convenient, cost-effective, and time-saving (18, 19). Additionally, multi-epitope vaccines have the advantage of simultaneously inducing diverse immune responses such as humoral, innate, and cellular responses than monovalent vaccines (20). Multi-epitope-based vaccine models against many viruses, including Lassa, Cytomegalovirus, Ebola, Zika, Chikungunya, Dengue, and Hepatitis C, have been developed utilizing vaccinomics (21). An *in-vitro* study against *Mycobacterium Tuberculosis* verified the viability of the multi-epitope vaccine (21, 22). Moreover, a vaccine developed using a similar technique has been proven to have *in-vivo* prophylactic properties, and many of these vaccines have advanced to the clinical trial stage (23). Therefore, a multi-epitope subunit vaccine may become a viable treatment option for MPXV infection. For that aim, a multi-epitope subunit vaccine for MPXV was designed employing various immunoinformatic methods.

Materials and methods

Acquisition and assessment of epitopes

Collection of epitopes

The experimentally determined epitopes sequences of the MPXV were obtained from NIAID Virus Pathogen Database and Analysis Resource [(ViPR) (http://www.viprbrc.org); Immune epitopes > Family: *Poxviridae* > Subfamily: *Chordopoxvirinae* > Genus: *Orthopoxvirus* > Species: Monkeypox virus]. Total 139 epitopes sequences, comprising 133 major histocompatibility complex (MHC) class I, one MHC class II, and five B-cell epitopes, were downloaded from ViPR in Excel format. The overall workflow to design a multi-epitope vaccine for MPXV is shown in Figure 1.

Allergenicity, antigenicity, toxicity and solubility analysis of epitopes

AllerTOP v.2.0 (24) and the Non-Allergen prediction tool of VaxELAN pipeline (25) were utilized to evaluate the allergenicity of the collected epitope. Only epitopes labelled as non-allergen were selected for further analysis. Next, the VexiJen v2.0 server (26) was implemented to test the antigenicity of epitopes with a threshold of 0.5 and the model set as "virus". Epitope toxicity was checked through ToxinPred server (27), applying the default parameters. Similarly, the virulence potential of the chosen non-allergen, antigenic, and non-toxic epitopes was assessed by using Virulentpred (Cascased SVM module with a threshold of 0.0) (28). Epitopes labelled as "virulent" were selected for the next evaluation. Subsequently, the Innovagen website (http://www.innovagen.com/proteomics-tools) was used to filter non-soluble epitopes.

Conservancy, non-homology, and physicochemical property analysis of epitopes

We computed the degree of epitopes conservancy within the MPXV proteins sequences through Immune Epitope Database and Analysis Resource (IEDB) epitope conservancy tool (http://tools.iedb.org/conservancy/). Proteins sequences of Monkeypox virus (NCBI Reference Sequence: NC_063383.1) and Monkeypox virus Zaire-96-I-16 (NCBI Reference Sequence: NC_003310.1) were downloaded in FASTA format from NCBI Reference Sequence Database (https://www.ncbi.nlm.nih.gov/refseq/). The selected epitopes sequences were aligned to their source protein sequences using the IEDB conservancy tool with an identity level set at 100. Epitope sequences were screened against *Homo sapiens* proteome (NCBI taxid: 9606) using the BLASTp tool (29) to discard epitopes that could potentially trigger an autoimmune reaction. Epitopes having percent identity ≤ 70% were retained. Besides, we predicted the similarity of epitopes sequences with the proteome of human gut microbiota with the aid of Non-Bacterial Pathogen prediction tool of VaxELAN pipeline. Physicochemical properties of shortlisted epitopes were estimated *via* the Expasy ProtParam program (https://web.expasy.org/protparam/) as well as Stability and Molecular Weight (kDa) prediction tools of VaxELAN pipeline.

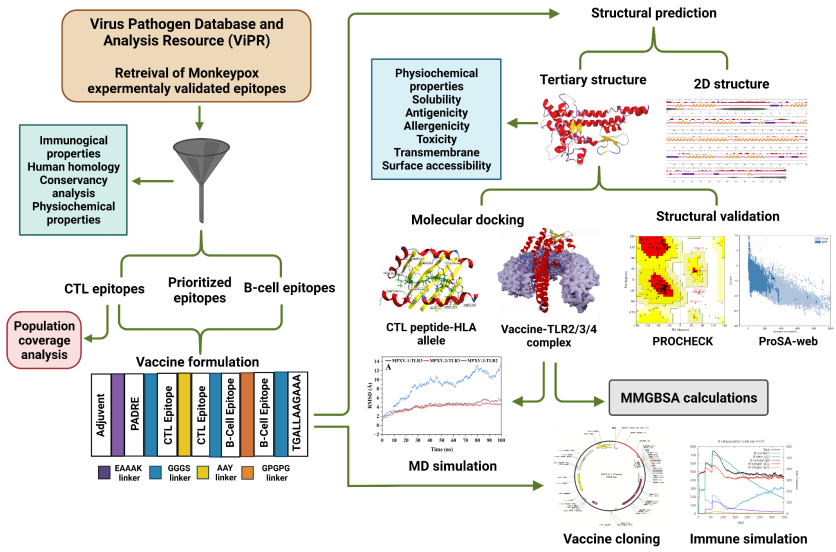


FIGURE 1
Overall workflow to design vaccine candidates for monkeypox in this study.

Proteome screening of monkeypox to predict best-ranking extracellular proteins

Proteins sequences (191) of the reference Monkeypox virus Zaire-96-I-16 (NCBI Reference Sequence: NC_003310.1) were retrieved from NCBI Reference Sequence Database. A state-of-the-art pipeline called VaxELAN was utilized to select the best-ranked outer cellular membrane proteins of MPXV. FASTA sequence of the retrieved proteins was submitted to VaxELAN. Literature-based cut-offs and predefined strategy 1B (integrated with VaxELAN) were applied to find out the eligible protein candidates. Next, the shortlisted proteins' molecular weight (MW) was computed with the VaxELAN pipeline tool to filter those with MW >110 kDa. To select the final protein candidates, antigenicity and transmembrane (TM) helices of the proteins were checked through VexiJen v2.0 (cut-off score for antigen ≥0.4) and DeepTMHMM (cut-off number of helices ≤ 1) (30), respectively.

Formulation and evaluation of vaccine constructs

Formulation of multi-epitope vaccines

The final epitopes were fused with separate linkers and adjuvant to formulate three multivalent vaccine constructs for MPXV. A potent immunostimulatory adjuvant was added to each multi-epitope vaccine to enhance immunogenicity and trigger long-established innate and adaptive immunity (31). Three different adjuvants were used in the study, including human beta-defensin 3, 50S ribosomal protein L7/L12, and Heparin-binding hemagglutinin [(HBHA) both from *Mycobacterium* sp.],. First, an adjuvant was added to the N-terminal of the vaccine using the EAAAK linker, which is a rigid linker. Following this linker, the pan-HLA DR binding epitopes (PADRE epitope 13aa) sequence was inserted to further increase the vaccine immunogenicity (32). Linkers are essential for achieving extended conformation (flexibility), protein folding, and functional domain separation, all of which contribute to a more stable protein structure (33). The cytotoxic T-lymphocytes [CTLs or major histocompatibility complex (MHC) I binders] epitopes were linked with the AAY linker, and B-cell epitopes were linked with the GPGPG linker. Epitope class was separated with the aid of a GGGS linker (32). Finally, a TAT sequence (11 aa) was adjoined to the vaccine's C-terminal (34).

Population coverage evaluation

The IEDB's population coverage tool (http://tools.iedb.org/population/) was used to estimate the selected epitopes coverage in the target population. We examined the HLA binding alleles of selected CTL epitopes – predicted by the IEDB SMM method (http://tools.immuneepitope.org/mhci/) with an IC_{50} threshold of 500 – for the population coverage analysis.

Prediction of antigenicity, allergenicity, and toxicity of multi-epitope vaccines

Two servers were employed to predict the allergenicity of the modeled vaccines: AllerTOP v.2.0 and AllergenFP v.1.0 (35). Using the VexiJen v2.0 server, the antigenicity of the designed vaccines was evaluated with a threshold value of 0.5. The

antigenicity of vaccine constructs was also evaluated using ANTIGENpro, which can be found at http://scratch.proteomics.ics.uci.edu/. ToxinPred server was employed to predict whether the constructed vaccines are non-toxic.

Solubility and physicochemical profile evaluation of vaccine constructs

Employing the Protein-Sol (36) and SOLpro (37) servers, the solubility of vaccine constructs was predicted. Protein-Sol server exploits the protein solubility data from an *Escherichia coli* expression system to predict the protein sequence solubility (38). SOLpro implements an SVM-based technique to predict the solubility of protein sequence with an accuracy of 75% (39). Several physicochemical characteristics of vaccine constructs, such as molecular weight, Grand average of hydropathicity (GRAVY), theoretical isoelectric point (pI), instability index, aliphatic index, and half-life, were estimated using the Expasy ProtParam program (https://web.expasy.org/protparam/). Using the DeepTMHMM (30), number of TM helices in the proposed vaccine constructs were predicted. To avoid autoimmunity, the vaccine constructs should not be similar to the human proteins. Therefore, BLASTp tool was used to check the sequence similarity between designed vaccine constructs and human proteins (NCBI taxid: 9606).

Prediction of secondary structure, solvent accessibility, and disorder regions

With the aid of NetSurfP-3.0 server (https://services.healthtech.dtu.dk/service.php?NetSurfP-3.0), the secondary structure features (α-helixes, β-sheets, and random coils) of vaccine constructs were predicted. This server implements the ESM-1b language model to produce sequence embedding, which is then processed by a deep neural network (40). The solvent accessibility and disorder regions of designed vaccine constructs were also predicted exploiting the NetSurfP-3.0 server.

Modeling, refinement, and quality validation of tertiary structure

I-TASSER (41) and RoseTTAFold (42) servers were used to predict the tertiary structures of the designed vaccines. Following the primary three-dimensional (3D) modeling, the modeled structures were further improved by employing GalaxyRefine Server (http://galaxy.seoklab.org/). To improve loop or terminal regions in the primary 3D model, this server applies ab initio modeling. Quality validation of the predicted tertiary structure was carried out using the ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) and ERRAT (https://saves.mbi.ucla.edu/) server. The Z-Score provided by ProSA-web could be used to assess the overall and local model quality. Besides, the ERRAT quality factor and PROCHECK Ramachandran plot (which showed favourable backbone dihedral angles in relation to amino acid residues in tertiary structure) were further used for model validation (43–45).

Prediction of confirmational B-cell epitopes

Using the ElliPro web server (http://tools.iedb.org/ellipro/), conformational B-cell epitopes within vaccine constructs were predicted. The default parameters (minimum score of 0.5 and maximum score of 0.6) of this server was kept for epitopes predictions.

Molecular docking study

Peptide modeling and docking with MHC-I molecules

The final CTL epitopes were modeled by PEP-FOLD 3.0 server (46) with 2000 simulations and sOPEP sorting scheme. Molecular docking was performed between CTL epitopes and major histocompatibility class I molecules [HLA-A*0201 (PDB: 4U6X) and HLA-B*15:02 (PDB: 6VB2)] applying the protein-protein (PP) docking protocol of MOE2020 software (47). Before docking, retrieved PDB structures were prepared by removing ligand and water molecules, and energy was minimized with the MOE QuickPrep module. The binding pocket and active site residues of the attached peptide with the PDB structure of selected HLA alleles were selected as the docking site for epitopes. The predominant binding mode (out of five poses) was retained based on strong interaction and high docking score (DS) for each complex. The protein-protein interaction panel of MOE2020 was used for peptide–receptor contact analysis.

Molecular docking of vaccine and TLRs

In this study, the binding affinity of vaccine constructs and immune receptors was tested through molecular docking. PP docking protocol of MOE2020 software was applied to perform docking between constructs and several immune receptors, including Toll-like receptors [TLR2 (PDB: 6NIG), TLR3 (PDB: 2A0Z), and TLR4 (PDB: 3FXI)]. All PDB structures were prepared, and energy minimized *via* the MOE QuickPrep module. Attached ligands, heteroatoms, and chains B, C, and D (in the case of TLR2 and TLR4) were removed before docking. Employing a rigid body refinement approach of MOE2020, the final 30 poses were retained during the docking. The PDBsum server [24] was utilized for vaccine–receptor contacts analysis, while structural illustrations were generated using the Blender software (48). Based on a high DS (high negative score), the top-

docked vaccine–TLR complex for each construct was further evaluated.

In silico cloning and RNA secondary structure prediction of constructed vaccines

The cDNA of constructed vaccines was obtained by submitting their amino acid sequence into EMBOSS Backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). To optimize the codon usage by *Escherichia coli* (*E. coli*, prokaryotic organism), codon optimization was done *via* Java Codon Adaptation Tool (JCat) (49). The *E. coli* strain K12 was selected to increase vaccine expression efficiency. Additional options, such as avoiding rho-independent transcription termination, prokaryote ribosome attachment site, and the cleavage site of restriction enzymes, were checked for the correct translation of the vaccine gene. Codon optimization index (CAI) and percent GC-content of the output sequence was checked to examine protein's expression potency. At the N and C-terminal of the vaccine codon sequence, the cleavage site for XhoI and SacI was appended. Using the SnapGene software (https://www.snapgene.com/), the optimized sequence was cloned between XhoI and SacI site in the pET28a (+) expression vector. In addition, minimum free energy mRNA secondary structure from each construct optimized cDNA sequence was predicted using the RNAfold program (50).

Immune simulation

Computation immune simulation was performed exploiting the C-ImmSim server to ascertain the immune response of the multi-epitope vaccine to the host (51). To describe cellular and humoral profile of mammalian immune system against the formulated vaccine, this server implements the Celada-Seiden model. Following the literature (21), three vaccine doses were injected at suggested 28 days intervals. Total simulation steps were 1050, and time steps of injection were 1, 84, and 170. The rest of the simulation's parameters were retained as default.

Molecular dynamics simulations

The vaccine-receptor complex was subjected to molecular dynamics (MD) simulations applying the Particle Mesh Ewald Molecular Dynamics (PMEMD) (52) engine of AMBER22 (53) software. To generate the coordinate and topology file of the system for protein amino acids, the AMBER22 Leap module (53) was employed with ff19SB (54) force field. Monovalent optimal point charge (OPC) ions (55) (Na^+ and Cl^-, ~0.1 M) were introduced to neutralize the system. Solvation of each system was carried out in a truncated octahedral box with an OPC water model and 10% of a buffer. Employing the PMEMD engine (52) on GPUs, parallel scaling in long-range electrostatics was optimized. A two-step initial energy minimization was applied: steepest descent minimization (20000 steps) and conjugate gradient minimization (10000 steps) (56). A Langevin thermostat and an NVE ensemble (microcanonical) were implemented to attain continual heating in 400ps time from 0.1 to 300K before the minimization phase. Besides, a Langevin thermostat (57) was used to adapt the kinetic energy of harmonic oscillators for dynamic propagation with $2.0ps^{-1}$ collision frequency. The density was adjusted in the 400ps run using the same method as before. Using an NVE ensemble for 400ps, system equilibration was performed at 300K for 2000ps with no restraint and a pressure relaxation time of 2ps. During the equilibration phase, the pressure was upheld by applying the isotropic position scaling method and relaxed for 1ps. SHAKE (58), and particle-mesh Ewald (59) method was used to restrict all hydrogen bonds and estimate long-range electrostatics (threshold of 8Å), respectively. Finally, 100ns production molecular dynamics (MD) was run for each vaccine–TLR complex with same protocol as equilibration phase. For investigating the trajectory, frames were gathered every 10ps.

Analysis of simulation trajectories

The resulting trajectory obtained from the MD simulation was investigated using the CPPTRAJ package of AMBER22. Based on cartesian coordinates of C-alpha atoms, overall deviation from the initial structure is measured through the root mean square deviation (RMSD) plot and the resultant root mean square fluctuations (RMSF) (60). The compactness of structure over the simulation duration was examined by the radius of gyration (Rg) and applying the equation mentioned elsewhere (60). Moreover, the protein's surface properties were analyzed with solvent-accessible surface area (SASA) calculation. Since hydrogen bonds (H-bond) are crucial factors that affect protein stability, we estimated vaccine-receptor interface H-bonds. The H-bond donor and acceptor atom's distance threshold of 3.5 Å and angel cut-off of 120° were applied in H-bond analysis in this study.

Gibbs free energy distribution

The conformational free energy values associated with stable and transient states of the complex were measured hereby. Using the CPPTRAJ package of AMBER22 (60), the system's Free Energy Landscape (FEL) was investigated. The trajectories data were distributed into 100 bins by principal components, PC1 and PC2, with the highest variations. An artificial barrier was set for bins (having no population) with a population size of 0.5 throughout free energy estimations. Free energy estimations were carried out at 300°C, and all units were measured in kcal/mol.

Binding free energy calculations

The binding free energy of the vaccine-TLR complex was computed by MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) method (61). The free energy binding of the complex was computed by applying the below formula (62).

$$\Delta Gbind = \Delta GR + L - (\Delta GR + \Delta GL)$$

GR + L, GR, and GL are protein-ligand (vaccine) complex, free protein, and ligand energy, respectively. In addition, applying the equations (63), the ΔG term (free energy) is computed.

Results

Collection of monkeypox experimentally validated epitopes from ViPR

A total of 139 experimentally validated epitopes of monkeypox, including 133 MHC class I, one MHC class II, and five B-cell epitopes, were collected from ViPR. These epitopes were then analyzed for several parameters (described in the method section) to select final candidates for vaccine construction.

Prioritization of monkeypox experimentally validated epitopes

Evaluation of immunological properties and solubility of epitopes

Out of 139 epitopes, 46 allergens, 29 non-antigenic, 23 non-virulent epitopes, and 27 poorly soluble predicted by the respective servers were discarded. All epitopes were predicted non-toxic; therefore, no epitope was eliminated after the toxicity analysis. Antigenicity scores of the shortlisted 14 epitopes range from 0.58 to 1.57. While the virulence scores of shortlisted epitopes range from 1.02 to 1.12 (Table 1).

Evaluation of conservancy and human homology of epitopes

All shortlisted 14 epitopes indicated 100% sequence conservation with their source proteins sequences from monkeypox and monkeypox strain Zaire 77-0666. According to the BLASTp homology evaluation, none of the epitopes showed an exact match with *Homo Sapiens* proteins (% identify ≤70%), hinting that the shortlisted epitopes are not expected to trigger autoimmunity.

Evaluation of physicochemical properties

The majority of shortlisted peptides have instability index value<40 (stable) and high aliphatic index value (thermostable). The molecular weight of final peptides ranges from 998.14 to 2475.81 Da, and the predicted half-life of the peptides in the mammalian cell is between 1 h to 20 hrs (Table 1). The antigen source (or protein name) and entry details of the finalized 14 peptides are provided in Table S1.

Reverse vaccinology pipeline to select exposed proteins

Out of 191 proteins of MPXV, the integrated pipeline finalized eight proteins (A21L, A30L, A43R, B8R, B9R, B20R, C22L, and J2R) as potential vaccine candidates (PVCs) (Figure S1). NCBI IDs, various properties, locus, and proteins names of final vaccine candidates are the given Table S2.

Population coverage analysis

The worldwide population coverage was estimated using the IEDB analysis tool to predict the population coverage of the shortlisted MHC-I epitopes. These epitopes interacted with the assembly of reference alleles (Table S3). The worldwide population coverage of MHC-I epitopes was 82.6%, with a pc90 average of 0.57 and an average hit per HLA antigen of 5.63 (Figure 2). The combined population coverage of the selected epitope class was highest in Europe (89.6%), followed by North America (85.63%), West Africa (84.84%), West Indies (83.08%), North Africa (81.83%), and East Africa (80.47%). The selected epitopes had >70% population coverage in East Asia, South Africa, and Central Africa. In addition, with over 60% combined value, final epitopes indicated better coverage in Northeast Asia, Southeast Asia, Southwest Asia, and South Asia populations. In South American and Oceanian populations, population coverage of final epitopes was 52.45% and 50.7%, respectively. On the other hand, in Central America, the population coverage of the shortlisted epitopes class was estimated to be the lowest (3.55%). The average hit per HLA antigen of MHC-I epitopes was 4.14, whereas the pc90 value was calculated to be 0.44. Individual coverage of MHC-I epitopes in the world population is presented in Table 2. Among all epitopes, KLSSYHVVSV was predicted to have the highest coverage of 67.23%, followed by ITVGMLIYSM (62.19%), NVDSTDELM (59.24%), LQKFSFKIA (55.78%), and AIIDIEPDL (55.25%). The analysis revealed that selected epitopes could cover a wide range of the human population; thus, they can be promising candidates for formulating polypeptide vaccine.

10.3389/fimmu.2022.1004804

TABLE 1 Sequence conservation, human homology, immunological and physicochemical properties of finalized experimentally validated epitopes of MPXV.

Epitopes	AT	Allergenicity	Solubility	Toxicity	Virulence	Conser-vation	Human-homology	II	AI	Half-life (Mammals, hrs)	MW
AIIDIEPDL	1.36	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	70.73	184.44	4.4	998.14
IHLEWLLGF	1.57	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	19.61	173.33	20	1127.35
LQKFSFKIA	0.69	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	18.98	97.78	5.5	1081.32
RTVIHLEWL	1.51	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	19.61	162.22	1	1166.39
SLKDVLVSV	0.89	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	– 0.54	183.33	1.9	959.15
WKVLSIMAF	0.92	NA	soluble	NT	Virulent (1.09)	100% (2/2)	Non-H	13.17	130	2.8	1094.38
ITVGMLIYSM	0.94	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	0.51	146	20	1127.42
KLSSYHVVSV	1.00	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	32.11	126	12.6	1118.30
NVDSTDELM	0.58	NA	soluble	NT	Virulent (1.05)	100% (2/2)	Non-H	13.59	75.56	1.4	1023.08
WAIIPLSASV	1.14	NA	soluble	NT	Virulent (1.06)	100% (2/2)	Non-H	-8.91	166	2.8	1056.27
*LSAATETYSGLT PEQKAYVPAMF	0.74	NA	soluble	NT	Virulent (1.12)	100% (2/2)	Non-H	50.62	63.91	5.5	2475.81
*DSGYHSLDPNAV CETD	0.87	NA	soluble	NT	Virulent (1.02)	100% (2/2)	Non-H	23.82	48.75	1.1	1722.76
*YGAPGSPTNLEFI NTGSSK	1.09	NA	soluble	NT	Virulent (1.07)	100% (2/2)	Non-H	27.09	46.32	2.8	1940.11
*CVRSNEEFDPVDD GPDDETDLSKLSKD	0.61	NA	soluble	NT	Virulent (1.03)	100% (2/2)	Non-H	35.39	50.37	1.2	3026.15

*B-cell epitopes; AT, Antigenicity; NA, Non-allergen; NT, Non-toxic; Non-H, Non-human homologue; II, Instability index; AI, Aliphatic index; MW, Molecular weight.

Vaccine construction with different adjuvants

The multi-epitope vaccine was constructed by fusing the finalized 10 MHC-I and four B-cell epitopes with adjuvant and linkers. An AAY linker was used to connect MHC-I binding epitopes, which may help create an appropriate site for antigen epitopes to bind to TAP transporters and improve epitope presentation. To join B-cell epitopes, GPGPG linkers were used. This linker stimulates HTL responses while preserving confirmational dependent immunogenicity of T-helpers and antibody epitopes (64). An adjuvant was added to the N-terminal of the construct *via* the EAAAK linker to improve the immunogenicity and immune response against the antigen. A PADRE sequence (13 aa) was also inserted to further increase the vaccine's immunogenicity. The pan-HLA DR binding epitopes in the vaccine construct enable binding to a wide range of mice and human MHC-II alleles to generate CD4+ helper cell responses (33). A TAT sequence (TGALLAAGAAA 11aa) was added to the C-terminal to allow intracellular delivery of the modeled vaccine (65). Three potential multi-epitope vaccines were constructed. Monkeypox vaccine-1 (MPXV-1) has 306 amino acids (aa) containing beta-defensin 3 adjuvant. MPXV-2 (388 aa) and MPXV-3 (417 aa) have 50S ribosomal protein L7/L12 and HBHA as adjuvants, respectively. The whole aa sequences of modeled vaccines are shown in Table 3. The sequence homology of the final vaccine protein to the human protein sequence showed no significant alignments.

Assessment of antigenicity, allergenicity, and toxicity of constructed vaccines

VexiJen v2.0 server predicted the constructed vaccines as potential antigens with antigenicity scores of 0.63 (MPXV-1), 0.55 (MPXV-2), and 0.59 (MPXV-3). ANTIGENpro server also confirmed the probable antigen character of the proteins with a prediction score >0.7. AllerTOP v.2.0 and AllergenFP v.1.0 labelled

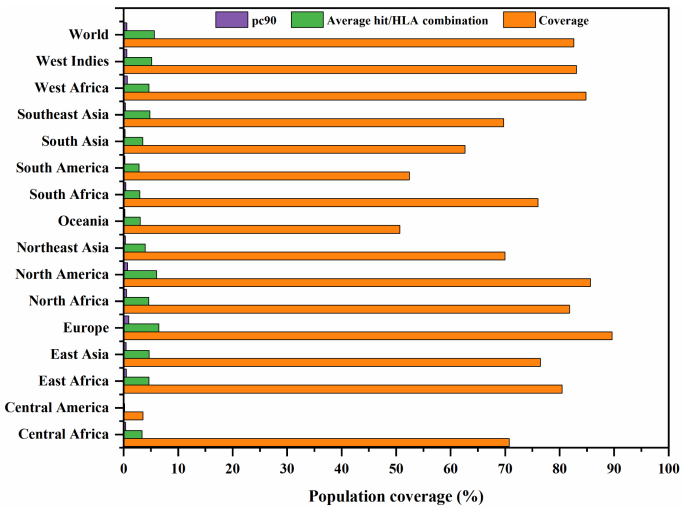


FIGURE 2
Global population coverage of selected MHC class I epitopes. pc90, the minimum number of epitope hits/HLA combinations recognized by 90% of the population; Coverage, projected population coverage; Average hit/HLA combination, an average number of epitope hits/HLA combinations recognized by the population.

the potential vaccine constructs as non-allergen. Similarly, the designed constructs were predicted non-toxic by the ToxinPred server (Table 4).

Analysis of solubility and physicochemical properties of the designed vaccine constructs

Protein-Sol and SOLpro servers predicted the constructed vaccines as soluble. MPXV-1–3 vaccine has MW of 31.98kDa, 39.93kDa, and 44.17kDa and theoretical pI of 7.66 (alkaline), 4.68 (acidic), and 4.70 (acidic), respectively. The estimated MW of constructs is suitable for vaccine application, and they could be purified easily (>110kDa). The computed instability index of MPXV-1 (23.82), MPXV-2 (19.39), and MPXV-3 (27.43) indicated that the designed vaccine constructs are stable proteins (value >40 indicates a stable vaccine). The aliphatic index of the proteins ranges from 84.02 to 92.99, classifying the formulated vaccine constructs as highly thermostable. The estimated half-life of all constructs is 30 hrs *in-vitro*. The estimated half-life of all constructs in yeast and *E. coli* (*in-vivo*) is more than 20 hrs and 10 hrs, respectively. The predicted TM helices (0–1) also suggest the suitability of vaccines for application (Table 4).

TABLE 2 Individual world population coverage of selected MHC-I epitopes.

Epitope	Coverage Class I
AIIDIEPDL	55.25%
IHLEWLLGF	19.33%
LQKFSFKIA	55.78%
SLKDVLVSV	49.91%
RTVIHLEWL	49.85%
ITVGMLIYSM	62.19%
NVDSTDELM	59.24%
KLSSYHVVSV	67.23%
WAIIPLSASV	53.43%
WKVLSIMAF	19.33%
Epitope set	82.60%

Prediction of two-dimensional structure, solvent accessibility, and disorder regions

In the result of the MPXV-1 secondary structure, there were 40.19% β-sheets, 11.03% α-helix, and 48.37% random coil. The secondary structure of MPXV-2 and 3 revealed 44.84% and 61.63% β-sheets, 13.14% and 5.27% α-helix, and 43.29% and 33.33% of the random coil, respectively. The solvent accessibility of three vaccine constructs showed sufficient exposure of selected epitopes to solvent; however, few MHC-I epitopes (AIIPLSASV, IHLEWLLGF, and WAIIPLSASV) were moderately exposed in all designed vaccines. In the results of disorder regions prediction,

TABLE 3 The amino acid sequence of MPXV-1–3 ensemble. An adjuvant is inserted at the N-terminal of the vaccine ensemble. EAAAK linker connecting the PADRE sequence to the adjuvant is represented in bold letters.

Vaccine Name	Adjuvant	Length	Vaccine Ensemble
Monkeypox vaccine- 1 (MPXV-1)	beta-defensin 3	306	GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK**EAAAK**AKFVAAWTLKAAA**GGGS**AIIPLSASV **AAY**IHLEWLLGF**AAY**LQKFSFKIA**AAY**RTVIHLEWL**AAY**SLKDVLVSV**AAY**WKVLSIMAF**AAY**ITVGMLIYSM**AAY** KLSSYHVVSV**AAY**NVDSTDELM**AAY**WAIIPLSASV**GGGS***LSAATETYSGLTPEQKAYVPAMF***GPGPG***DSGYHSLD* *PNAVCETD***GPGPG***YGAPGSPTNLEFINTGSSK***GPGPG***CVRSNEEFDPVDDGPDDETDLSKLSKD***GGGS***TGALLAAGAAA
Monkeypox vaccine- 2 (MPXV-2)	50S ribosomal protein L7/L12	388	MAKLSTDELLKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKV VREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVK**EAAAK**AKFVAAWTLKAAA**GGGS**AIIPL SASV**AAY**IHLEWLLGF**AAY**LQKFSFKIA**AAY**RTVIHLEWL**AAY**SLKDVLVSV**AAY**WKVLSIMAF**AAY**ITVGMLIYSM **AAY**KLSSYHVVSV**AAY**NVDSTDELM**AAY**WAIIPLSASV**GGGS***LSAATETYSVGLTPEQKAYVPAMF***GPGPG***DSGYHSLD* *PNAVCETD***GPGPG***YGAPGSPTNLEFINTGSSK***GPGPG***CVRSNEEFDPVDDGPDDETDLSKLSKD***GGGS***TGALLAAGAAA
Monkeypox vaccine- 3 (MPXV-3)	Heparin-binding hemagglutinin	417	MAENPNIDDLPLAALGAADLALATVNDLIANLRERAEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFTTEELRKA AEGYLEAATNRYNELVERGEAALQRLRSQTAFEDASARAEGYVDQAVELTQEALGTVASQTRAVGERAAKLVGIEL**EAA** **AK**AKFVAAWTLKAAA**GGGS**AIIPLSASV**AAY**IHLEWLLGF**AAY**LQKFSFKIA**AAY**RTVIHLEWL**AAY**SLKDVLVSV **AAY**WKVLSIMAF**AAY**ITVGMLIYSM**AAY**KLSSYHVVSV**AAY**NVDSTDELM**AAY**WAIIPLSASV**GGGS***LSAATET* *YSGLTPEQKAYVPAMF***GPGPG***DSGYHSLDPNAVCETD***GPGPG***YGAPGSPTNLEFINTGSSK***GPGPG***CVRSNEEFDP* *VDDGPDDETDLSKLSKD***GGGS***TGALLAAGAAA

Different linkers (GGGS, AAY, and GPGPG) are also represented in bold letters. B-cell epitopes are italicized. A TAT sequence (TGALLAAGAAA) is inserted at the N-terminal of the vaccine ensemble.

starting and ending residues of the three vaccine constructs were predicted disordered. There were 91.51% ordered and 8.49% disordered residues in the MPXV-1, 86.35% ordered and 13.65% disordered residues in MPXV-2, and 94.73% ordered and 5.27% disordered residues in the MPXV-3 structure. Secondary structures, solvent accessibility, and disorder regions predictions of the three vaccine constructs are illustrated in Figures 3–5.

Prediction, optimization, and quality assurance of tertiary structure

The tertiary structure of MPXV-1–3 provided by I-TASSER was predicted to have better structural quality and was refined by the GlaxayRefine server, which yielded five optimized models. The details of five generated models for each construct are given in Table S4. Model 1 of MPXV-1 and model 5 of MPXV-2 and 3

TABLE 4 Evaluation of Estimated solubility, immunological, and physicochemical properties of designed vaccine constructs by Expasy ProtParam server.

Property	MPXV-1	MPXV-2	MPXV-3
SOLpro	0.76	0.96	0.93
Protein-Sol	0.46	0.65	0.52
Molecular weight	31984.51 Da	39930.45 Da	44170.60 Da
Formula	$C_{1435}H_{2228}N_{374}O_{428}S_{13}$	$C_{1804}H_{2823}N_{445}O_{556}S_9$	$C_{1964}H_{3071}N_{523}O_{619}S_8$
Theoretical pI	7.66	4.68	4.70
Ext. coefficient	53, 330 M^{-1} cm^{-1}	49, 975 M^{-1} cm^{-1}	54, 445 M^{-1} cm^{-1}
Instability index	23.82 (stable)	19.39 (stable)	27.43 (stable)
Aliphatic index	84.02 (thermostable)	92.99 (thermostable)	87.96 (thermostable)
Half-Life	30 hrs (mammalian reticulocytes, *in-vitro*). >20 hrs (yeast, *in-vivo* >10 hrs (*E. coli*, *in-vivo*).	30 hrs (mammalian reticulocytes, *in-vitro*). >20 hrs (yeast, *in-vivo*). >10 hrs (*E. coli*, *in-vivo*).	30 hrs (Mammalian reticulocytes, *in-vitro*). >20 hrs (yeast, *in-vivo*). >10 hrs (*E. coli*, *in-vivo*).
Allergenicity	AllerTOP v.2.0 (Non-allergen), AllergenFP v.1.0 (Non-allergen)	AllerTOP v.2.0 (Non-allergen), AllergenFP v.1.0 (Non-allergen)	AllerTOP v.2.0 (Non-allergen), AllergenFP v.1.0 (Non-allergen)
Antigenicity	ANTIGENpro (0.73), VexiJen v2.0 (0.63)	ANTIGENpro (0.75), VexiJen v2.0 (0.55)	ANTIGENpro (0.72), VexiJen v2.0 (0.59)
Toxicity	Non-toxic	Non-toxic	Non-toxic
TM helices	None	1	1

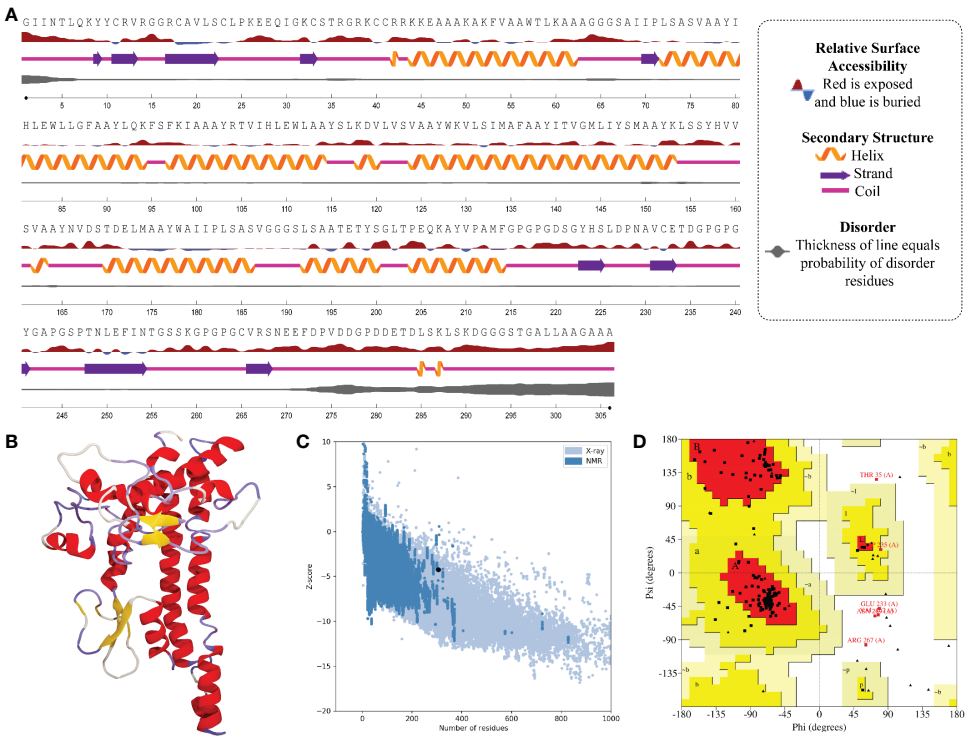


FIGURE 3
Structural analysis and validation of MPXV-1 **(A)** The predicted secondary structure features, relative surface accessibility, and disorder regions using NetsrurfP-3.0 server **(B)** Final 3D modeled structure of the constructed vaccine **(C)** Z-score graph *via* ProSA-web showing the modeled 3D structure corresponds to X-Ray crystallographic determined structure for the protein of similar sizes **(D)** Ramachandran plot details of the final modeled structure of designed vaccine.

had top Ramachandran favored, thus selected for docking study. A model with more residues located in the favored region of the Ramachandran plot and fewer residues in the additionally allowed, generously allowed, and disallowed region was considered ideal. The initial model (I-TASSER) and refined model (GalaxyRefine) were compared for Ramachandran plot details *via* PROCHECK.

The initial model of MPXV-1 had 89.5% residues in the most favoured regions of the Ramachandran plot, 7.4% in additional allowed regions, 1.2% in generously allowed regions, and 2.0% in the disallowed region. The refined model of MPXV-1 had 93.0% residues in most favoured regions, 4.7% in additional allowed regions, 1.2% in generously allowed regions, and 2.0% in the disallowed regions. Similarly, the initial MPXV-2 structural model had 88.2%, 6.9%, 2.1%, and 2.7% residues in the Ramachandran favoured, additionally allowed, generously allowed, and disallowed regions, respectively. On the other hand, the improved model had 89.4%, 6.0%, 1.5%, and 3.0% residues in the Ramachandran favoured, additionally allowed, generously allowed, and disallowed regions, respectively. For the MPXV-3 refined model, the Ramachandran plot distribution of residues was as follows: 94.8% in most favored regions (initial model 92%), 2.8% in additional allowed regions (initial model 5.8%), 1.1% in generously allowed regions (initial model 1.4%), 1.1% in disallowed regions (same as the initial model).

ProSA-web and ERRAT were employed to verify the quality and potential errors in the three-dimensional (3D) model of the final vaccine. The model with a lower Z-score was considered a high-quality one. The Z-score of the refined model of MPXV-1–3 was –4.25, –5.62, and –5.87, respectively. ProSA-web Z-score and Ramachandran plot details of MPXV-1–3 are provided in Figures 3–5 and Table 5. A higher value of the overall quality factor provided by ERRAT indicates a high-resolution structure for the given protein. The refined vaccine models showed a quality factor score >92%, suggesting the final models were of higher quality (Figures 3–5).

Prediction of discontinuous B-cell epitopes

Tertiary structure and folding of the new protein (vaccine construct 3D model) can result in a new discontinuous B-cell epitope, prompting further predictions. Therefore, such epitopes

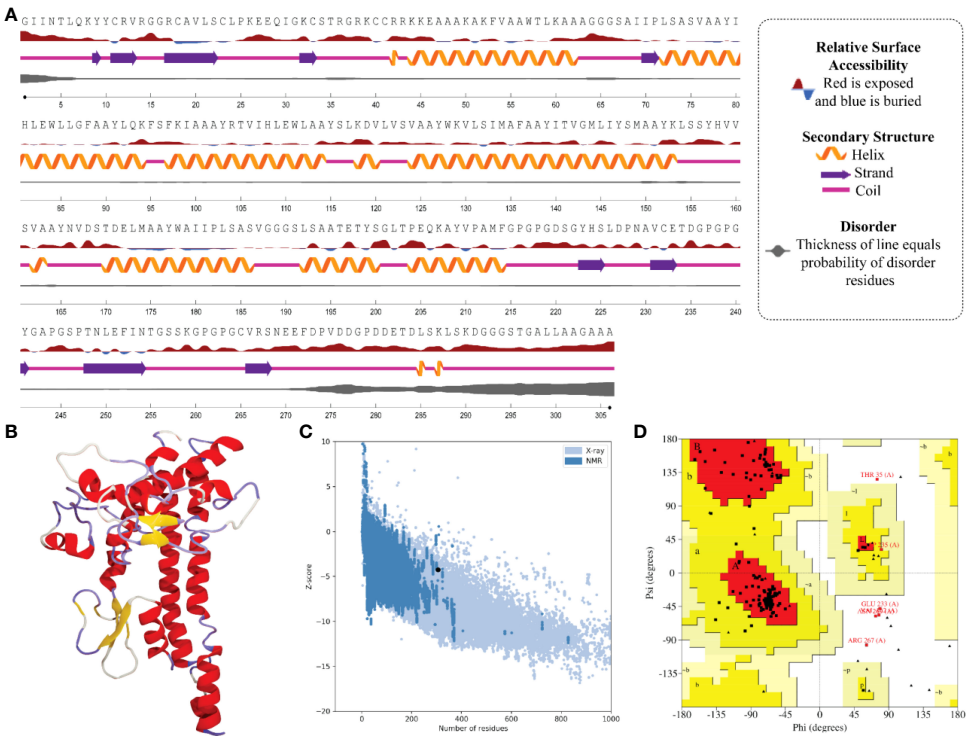


FIGURE 4
Structural analysis and validation of MPXV-2 **(A)** The predicted secondary structure features, relative surface accessibility, and disorder regions using NetsrurfP-3.0 server **(B)** Final 3D modeled structure of the constructed vaccine **(C)** Z-score graph *via* ProSA-web showing the modeled 3D structure corresponds to X-Ray crystallographic determined structure for the protein of similar sizes **(D)** Ramachandran plot details of the final modeled structure of designed vaccine.

in the refined tertiary structure of each construct were predicted and enlisted in Figure S2-S4. Briefly, seven discontinuous B-cell epitopes were predicted for MPXV-1 involving a total of 171 residues with a score ranging from 0.55 to 0.76. For MPXV-2, four discontinuous B-cell epitopes were predicted, which involved a total of 207 residues with a score in the range of 0.62 to 0.86. In addition, predicted discontinuous B-cell epitopes for MPXV-3 (score ranged from 0.52 to 0.90) covered 221 residues.

Molecular docking analysis

Finalized CTL peptides were modeled and subjected to molecular docking with their common MHC-I allele (HLA-A*0201 and HLA-B*15:02). All 10 peptides showed stable and high-affinity binding with the HLA-A*0201 molecule, as shown by their high DS ranging from –9.43 to –11.13 kcal/mol. With HLA-A*0201 molecule, RTVIHLEWL peptide showed the highest DS (–11.13 kcal/mol) and several H-bonds, followed by ITVGMLIYSM (–10.7761 kcal/mol) and IHLEWLLGF (–10.6489 kcal/mol). Likewise, the selected CTL peptides indicated a substantial affinity for the HLA-B*15:02 molecule (DS in the range of –9.53 to –12.55 kcal/mol). Peptide WKVLSIMAF also exhibited the highest DS of –12.55 kcal/mol with HLA-B*15:02 molecule, followed by LQKFSFKIA and RTVIHLEWL (DS: –11.84 kcal/mol for each epitope). Peptide–MHC-I molecule docking calculations, interface interactions, and predominant binding mode with the respective HLA receptor are provided in the Tables S5, S6 and Figures S5, S6.

Antigen molecule-immune receptor interactions are essential for transporting antigenic molecules and prompting the immune response pathway. To study the strength of binding energy and molecular interaction between various immune receptors (TLR2, TLR3 and TLR4) and modeled vaccines (MPXV-1–3), molecular docking was performed. The optimal-docked complex was selected out of 30 poses, indicating vaccine constructs' efficient binding at the receptor-binding domain. Amongst all TLR receptors, TLR3 exhibited high-affinity binding with MPXV-1 (–99.09 kcal/mol) and MPXV2 (–98.68 kcal/mol). In contrast, TLR2 showed the best docking energy with MPXV-3 (–85.22 kcal/mol) among all TLR receptors. Thus, molecular docking analysis implies successful binding and substantial affinity between immune receptors and proposed

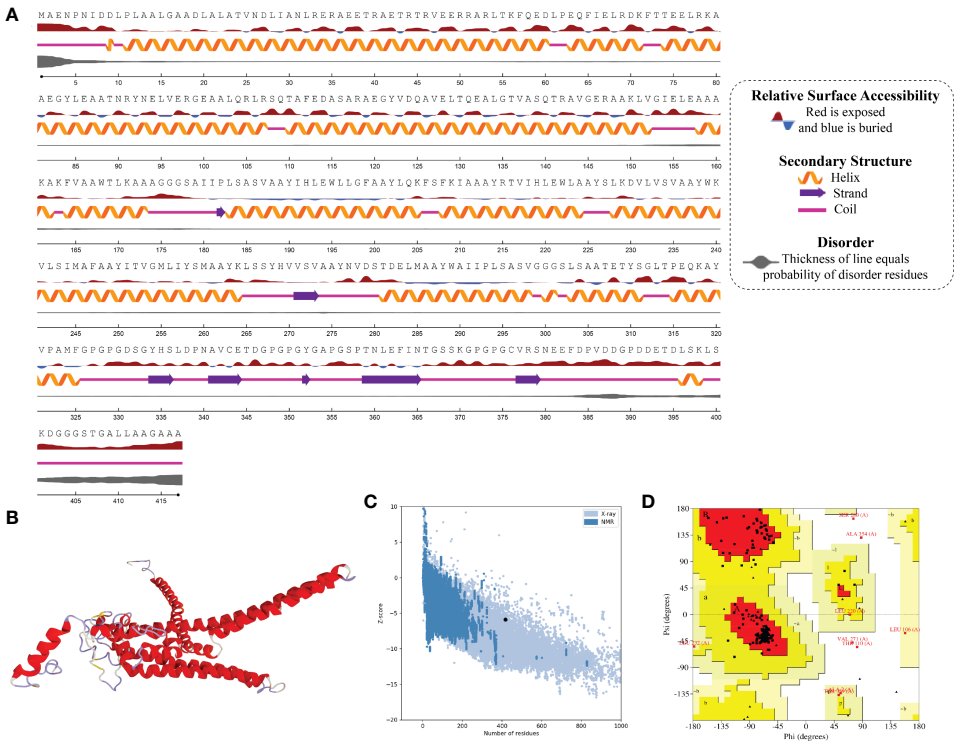


FIGURE 5
Structural analysis and validation of MPXV-3 **(A)** The predicted secondary structure features, relative surface accessibility, and disorder regions using NetsrurfP-3.0 server **(B)** Final 3D modeled structure of the constructed vaccine **(C)** Z-score graph *via* ProSA-web showing the modeled 3D structure corresponds to X-Ray crystallographic determined structure for the protein of similar sizes **(D)** Ramachandran plot details of the final modeled structure of designed vaccine.

vaccine constructs. Vaccine-TLRs docking estimation results are supplied in the Table 6.

Interface contact analysis of MPXV-1 and 2 with TLR3 receptor and MPXV-3 with TLR2 receptor is discussed herein due to their substantial docking energy. Strong molecular interactions were noticed for these complexes, including hydrogen bonds (H-bonds), salt bridges, and non-bonded contacts. With a bond distance< 2.75Å, MPXV-1 residues that made H-bond with TLR3 are: Met149, Lys153, Tyr157, Ser161, Ala164, Tyr165, Asp220, Ser246, Asp276, Asp277, Asp280, Asp291. MPXV-1 residues that mediated H-bond with a bond distance< 3Å included Gly66, Leu72 (2x), Leu145, Lys153, Tyr165(2x), Asn166, Gly240, Tyr241, Pro244, Gly245, Asn269, Glu270, and Asp291(2x). Moreover, His158 and Ala164 of MPXV-1 formed H-bond with TLR3 with a bond distance >3Å. Furthermore, salt bridge interaction was shown by MPXV-1 residues [Asp220, Arg267, Glu270, Asp280, and Asp280 (2x)] within a 3Å bond distance with TLR3. The detailed atom-atom contacts between MPXV-1 and TLR3 interface are given in the Figure 6 and Table S7.

In the case of MPXV-2, the following residues of the modeled vaccine mediated H-bond with TLR3 having a bond distance< 2.75Å: Lys3(2x), Glu30, Glu279, Glu287, Gln288,

TABLE 5 Validation of the modeled 3D structures of constructed vaccines.

Vaccine	ProSA-web		Ramachandran Plot details								ERRAT score	
	Z-score		Most favoured regions		Additional allowed regions		Generously allowed regions		Disallowed regions			
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
MPXV-1	−4.05	−4.25	89.5%	93.0%	7.4%	4.7	1.2%	0.4%	2.0%	2.0%	96.64	94.17
MPXV-2	−5.35	−5.62	88.2%	89.4%	6.9%	6.0%	2.1%	1.5%	2.7%	3.0%	97.34	96.81
MPXV-3	−5.72	−5.87	92.0%	94.8%	5.8%	2.8%	1.1%	1.4%	1.1%	1.1%	98.28	92.38

TABLE 6 Molecular docking results of MPXV-1–3 with selected TLRs.

Vaccine	Receptor PDB ID	Name	DS (kcal/mol)	E-Confirm	E-Refine
MPXV-1	6NIG	TLR2	−77.56	−3576.82	−77.56
MPXV-1	2A0Z	TLR3	**−99.09**	−3610.45	−99.09
MPXV-1	4G8A	TLR4	−78.11	−3598.26	−92.27
MPXV-2	6NIG	TLR2	−91.67	−8074.44	−98.68
MPXV-2	2A0Z	TLR3	**−98.68**	−8081.54	−47.13
MPXV-2	4G8A	TLR4	−94.23	−8070.04	−94.23
MPXV-3	6NIG	TLR2	**−85.22**	−12495.07	−85.22
MPXV-3	2A0Z	TLR3	−80.24	−12496.43	−80.24
MPXV-3	4G8A	TLR4	−80.91	−12508.11	−80.91

E-Conf, Energy of Confirmation: E-Refine, Energy of refinement; DS, Docking score.

Asp317 (2x), Ser340 (2x), Glu352, Asp362, and Ser371. Similarly, MPXV-2 residues, such as Ala33, Ala145, Gly147, Glu315 (2x), Asp317, Asn336, Lys341, Glu352, Asp362, and Lys372 formed H-bond with TLR3 with bond distance in a range of 2.76 to 3Å. Residues, Ala33, Gly283, Lys341, and Ser340 also made H-bond with TLR3 with a bond distance >3Å. In addition, several residues of the construct showed salt bridge interactions [Glu279, Glu287 (2x), Asp317 (2x), and Lys341], with TLR3 having a bond distance ranging from 2.72Å to 3.66Å. The detailed atom-atom contacts between MPXV-2 and TLR3 interface are given in Figure 6 and Table S8.

In the case of MPXV-3, residues Tyr190, Glu194, and Glu221 (2x) mediated H-bond with TLR2 with >2.75Å bond distance. Likewise, the construct residues that showed hydrogen bonding with TLR2 with a bond distance in a range of 2.76Å to 2.97Å are Thr55, Gln58, Leu193, Phe199, Glu194, Val217, His219, Leu220(2x), Glu221 (2x), Ala225, Lys229 (2x), Val231, Ser243(2x), and Ala246. Moreover, Glu194 and Glu221(3x) residues of MPXV-3 interacted through salt bridges with TLR3 within a 2.81Å bond distance. The detailed atom-atom contacts between MPXV-3 and TLR2 interface are given in Figure 6 and Table S9.

Molecular dynamics simulation

The predominant binding mode of the vaccine-TLRs complex was evaluated for structural stability by 100ns MD simulation. RMSD versus time plot was used to characterize the conformational stability of the complexes (Figure 7A). The mean RMSD of 4.49Å, 4.37Å, and 9.33Å was obtained for MPXV-1–TLR3, MPXV-2–TLR3, and MPXV-3–TLR2 complex, respectively. The first two complexes showed stable behaviour throughout the simulation. However, in the case of the MPXV-3–TLR2 complex, a sharp increase in RMSD (from 0.55Å to 9.34Å) was noticed till 40ns. Then the system underwent fluctuations in RMSD and remained unstable till the end. The residue flexibility of the vaccine-TLRs complex was evaluated through RMSF to understand the mobility of contact residues and their influence on the binding of the construct with the receptor (Figure 7B). The average RMSF of 1.15Å, 1.74Å, and 2.63Å was detected for the MPXV-1, MPXV-2, and MPXV-3 complex, respectively. In the case of the MPXV-1–TLR3 complex, a maximum RMSF of 11.57Å was noticed for Thr35 (MPXV-1; adjuvant) and a minimum of 0.57Å for Cys335 (TLR3). Besides, interacting residues of the MPXV-1 construct (H-bonded residues stated in docking results) with the receptor showed stable RSMF, estimated< 3Å. For the MPXV-2–TLR3 complex, the highly mobile residue was Ala196 (RSMF =8.63Å, MPXV-2 epitope) and the lowest mobile residue was Asn231 (0.60Å, TLR3). In addition, most of the MPXV-2 interacting residues (H-bonded residues stated in docking results) had RMSF ≤ 4.6Å. The highest and least fluctuating residue in the case of MPXV-3–TLR2 complex was Gly405 (RMSF = 20.82Å, MPXV-3 linker) and Ser342 (RMSF=1.3381, TLR3), respectively. The estimated RMSF for interacting residues of MPXV-3 (H-bonded residues stated in docking results) with receptor was ≤ 3.8Å. In the MPXV1/3 system, adjuvant residues of the construct fluctuated significantly. In contrast, in the MPXV-2 system, residues comprising a few MHC-I epitopes and AAY linkers revealed higher flexibility. These motion variations over the simulation run seem important for the vaccine construct to attain a stable confirmation for appropriate exposure to the host immune system.

The Rg profile of vaccine-TLRs complexes was plotted, as shown in Figure 7C, to examine the structural compactness and regular packing of secondary structural elements in the 3D structure of the complex. Less stable folding of secondary structural elements results in loose packing of the structure, which is indicated by a high Rg descriptor. The mean Rg value of 32.12Å and 34.45Å and 37.94Å was estimated for MPXV-1–3 bound complexes, respectively. Consistent with the RMSD results, the Rg plot of these complexes stayed stable during the simulation (although the MPXV-3 system experienced fluctuations), which points to fine compactness of the tertiary structure. Alteration in the complex volume was studied by the

10.3389/fimmu.2022.1004804

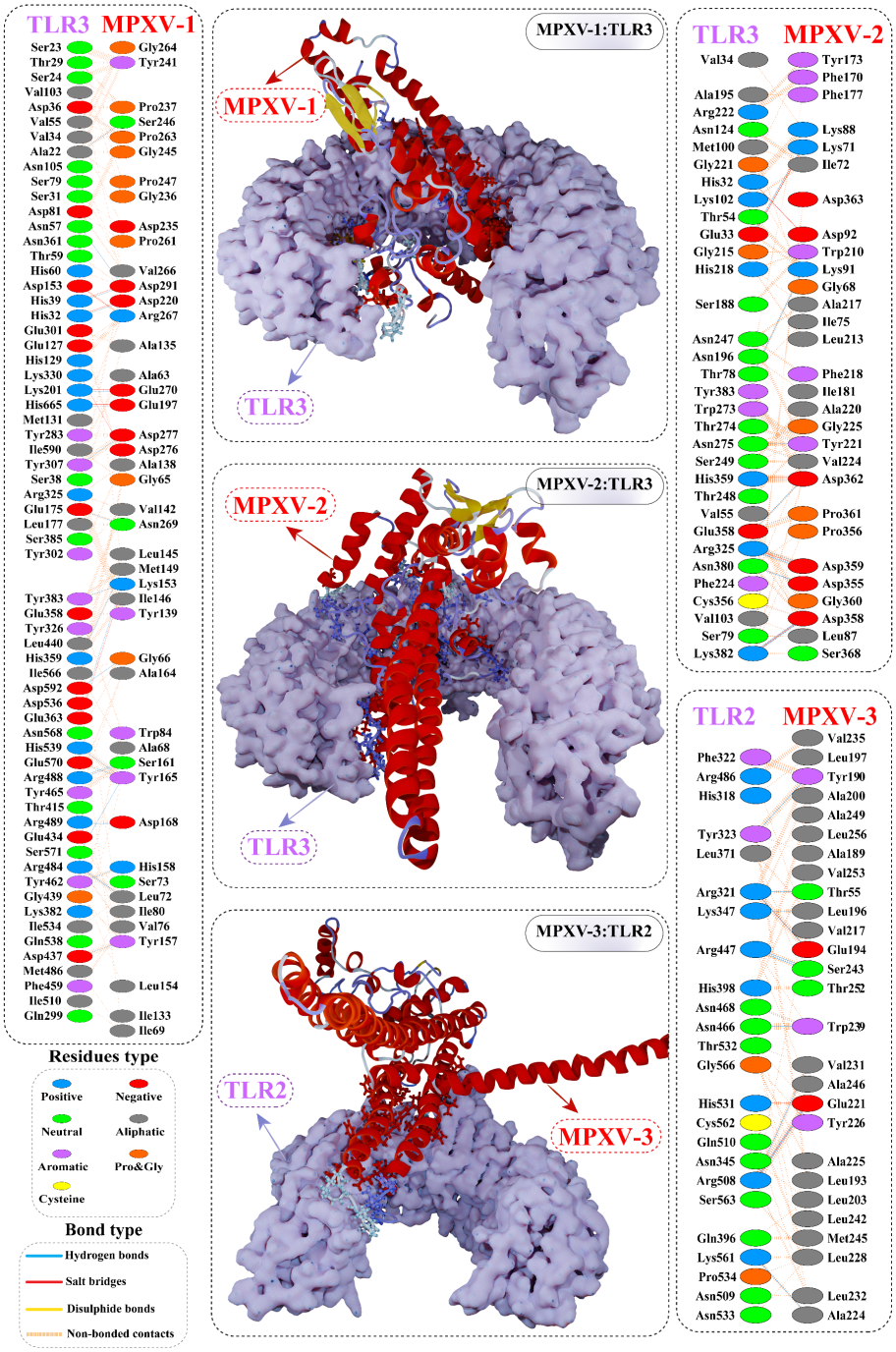


FIGURE 6
Optimal binding mode of MPXV vaccine construct with human Toll-like receptor-3 (TLR3) and TLR2 receptor. The docked complexes are illustrated in the middle. Protein-protein contacts of the vaccine construct with TLR3/2 receptor are shown on either side. Hydrogen bonds are depicted in blue lines. Residue type is shown with a distinctive colour.

SASA analysis (Figure 7D), which yielded a mean SASA of 45382.27Å, 50046.83 Å, and 55170.52Å for MPXV-1–3 systems, respectively. The SASA values of each complex gradually increased till 30ns and then sustained around the mean SASA value till the end of the simulation, indicating the expansion in the protein complex volume during the initial phase.

The transition phase of each MPXV–TLR complex was explored through the free energy landscape (FEL). To

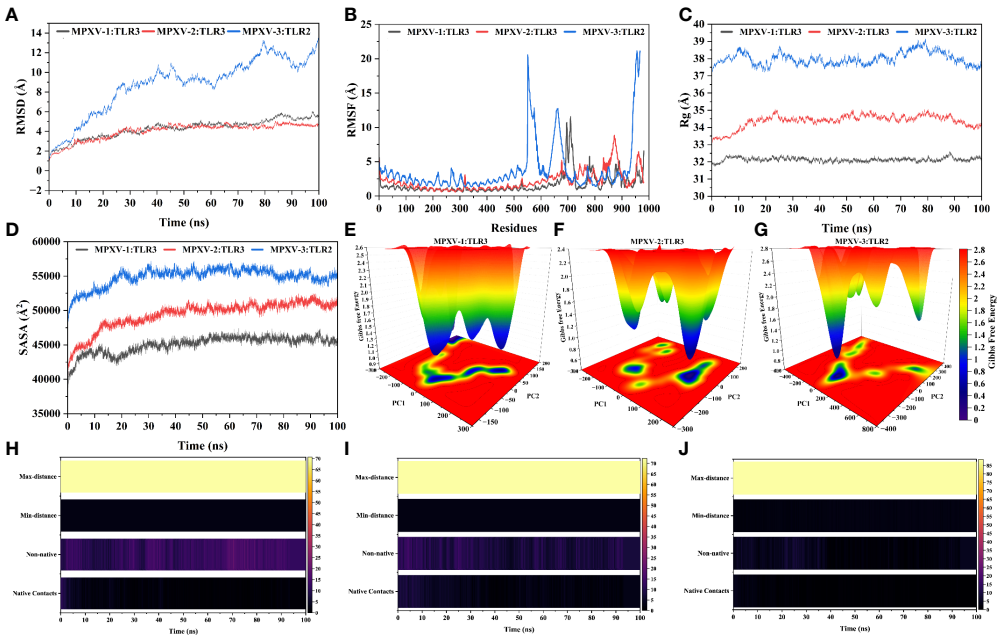


FIGURE 7
Molecular dynamics simulation of MPXV vaccines and TLR3/2 complex at 100ns **(A)** Root means square deviation (RMSD) plot of the complex, demonstrating slight fluctuations **(B)** Root means square fluctuation (RMSF) plot of the docked complex. TLR3 receptor starts at residue number 1 and goes to residue number 675. MPXV-1 starts at residue number 676 and goes to residue number 982. MPXV-2 starts at residue number 676 and goes to residue number 1064. TLR2 receptor starts at residue number 1 and goes to residue number 550. MPXV-3 start at residue number 551 and residue number 968 **(C)** Radius of Gyration plot of the docked complex **(D)** Variations in Solvent-Accessible Surface Area (SASA) profile of the docked complex during the simulation. Free energy landscapes (FELs) of the **(E)** MPXV-1–TLR3 **(F)** MPXV-2–TLR3 **(G)** MPXV-3–TLR2 docked complex. In the graph, red, yellow/green, and light-to-dark blue depicts high, intermediate, and low/stable energy states, respectively. Native and non-native contacts between **(H)** MPXV-1–TLR3 **(I)** MPXV-2–TLR3 **(J)** MPXV-3–TLR2 interface during 100ns.

investigate the progression from initial to metastable state, the first two eigenvectors from the resulting trajectory were plotted to generate FEL (Figures 7E-G). In addition, the low energy state of each complex system was displayed to analyze structural alteration after the attachment of the construct with the receptor. Across the simulation, high translation confirmation was seen within each system. Allocation of various metastable states (high energy level red, intermediate energy level yellow, stable energy state blue) reveal structural adjustment during the simulation.

The patterns of H-bonds for the three complexes were estimated in each frame within 3Å to probe the strength of intermolecular associations across the simulation period. For the MPXV-1–TLR3 complex, H-bond between $Leu145_{MPXV-1}$ and $His338_{TLR3}$, Tyr165 and Glu549, and Glu270 and Lys180 had a significant retention time of 52%, 33%, and 26% during the simulation, indicating these interactions are crucial for the specificity and strength of intermolecular association. Besides, Lys153, Asp280, and Tyr157 of the vaccine mediated H-bond with the receptor's Glu342, Arg310, and Asp416, respectively, with over 10% occupancy. All H-bond occupancy details between the MPXV-1–TLR3 interface are provided in Table S10. In the case of the MPXV-2–TLR3 complex, a high-frequency H-bond was detected between $Gly315_{MPXV-2}$ and $Lys180_{TLR3}$ (75%), Lys341 and Asp259 (56%), and Gly317 and Arg230 (~54%), Thr31 and Arg201 (42%), and Asp317 and Lys179 (36%) throughout the simulation. Additionally, Thr28, Ser343, Gly147, and Gln288 residues of the vaccine construct maintained H-bond with the receptor's Arg230, Glu285, Glu413, and Ser17 residues for the considerable period of simulation (>10% occupancy). The detailed H-bond interactions with percent occupancy between MPXV-2–TLR3 interface are enlisted in Table S11. Similarly, H-bond between $Glu221_{MPXV-3}$ and $Arg482_{TLR2}$ [(2x), 35% occupancy], $Ser243_{MPXV-3}$ and $Arg421_{TLR2}$(23%), and $Val217_{MPXV-3}$ and $Arg\ 482_{TLR2}$ (18%) contributed significantly towards the vaccine-receptor binding strength. The detailed H-bonds for MPXV-3–TLR2 complex are given in Table S12. Furthermore, changes in native contacts between vaccine-TLRs complex with simulation time can be seen in Figures 7H-J, whereby non-native contacts interactions between vaccine and TLRs receptors were improved after simulation, suggesting the complexes' stability.

Codon optimization

After reverse translation, adaptability, and preference analysis of codon, the DNA sequence of MPXV-1–3 constructs were 915, 1164, and 1248 nucleotides in length. Furthermore, the codon adaptation index (CAI) of MPXV-1 (1), MPXV-2 (1), and MPXV-3 (0.98) were greater than 0.8, meaning that the optimized codon has high adaptability for *E. coli* usage. In addition, the GC-content of the optimized codon was 52.45%, 51.37%, and 53.60% for MPXV-1–3, respectively, implying the stable and high expression potential of the designed constructs in the *E. coli* (strain K12) system.

In silico cloning and prediction of RNA secondary structure

Before cloning, cutting sites of XhoI (CTCGAG at the N-terminal) and SacI (GAGCTC at the C-terminal) were added flanking the constructs' DNA sequence. The new DNA sequence of MPXV-1–3 was 927, 1176, and 1260 nucleotides (Table S13 and Figure S7). Using the SnapGene software, the nucleotide sequence of each construct was then inserted between XhoI (site 158) and SacI (site 190) site in the pET-28a (+) vector. The cloned plasmid with MPXV-1–3 construct has a total length of 6262bp, 6511bp, and 6595bp, respectively, as shown in the Figure 8. The mRNA secondary structure prediction using the RNAfold program in the case of MPXV-1 yielded optimal secondary structure and centroid secondary structure with a minimum free energy of –298.90 kcal/mol and –213.80 kcal/mol, respectively. In the case of MPXV-2, the minimum free energy of obtained optimal secondary structure was –550.40 kcal/mol, and centroid secondary structure was –443.30 kcal/mol. Besides, the optimal secondary structure and centroid secondary structure of MPXV-3 had minimum free energy of –415.20 kcal/mol and –300.20 kcal/mol, respectively. The graphical output of predicted mRNA secondary structures of MPXV-1–3 is presented in Figure S8.

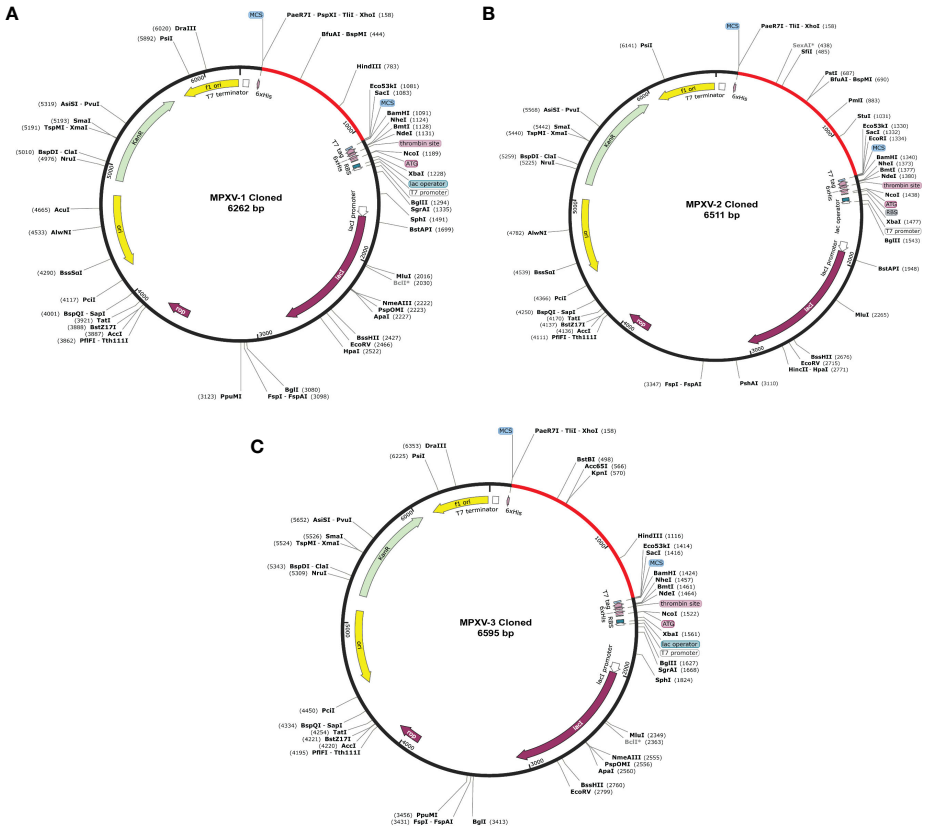


FIGURE 8
In silico cloning of the designed vaccine constructs **(A)** MPXV-1 **(B)** MPXV-2 **(C)** MPXV-3. The optimized DNA sequence of the proposed vaccine construct(s) (shown in red) was cloned between the XhoI and SacI enzyme loci in the expression vector pET-28a (+) (shown in black).

Computational immune assay

The immune simulation conducted by C-ImmSim server showed similar immune response data for the MPXV-1–3 constructs (Figures 9, S9). The antigen count declined as the antibody level in the immunological response increased, which was primarily attributed to the generation of total B-cells and T cells population. After every injection, there was an increase in the antigen level (>600,000 on 1st and ~500,000 on 2nd and 3rd doses), that subsequently neutralized at the fifth day of injection. This was followed by a rise in the production of secondary immunological responses (IgM + IgG: large titer scale of >200000 per ml) after exposing the host immune system to every vaccine injection. The highest concentration of >200,000 antibody titer per ml was achieved between 54–64 days. As measured by IgM, the primary immune response was similarly elevated to a high degree. Secondary immune responses were followed by tertiary immunological responses, resulting in a large B-cell population, and increasing levels of IgM + IgG, IgM, IgG1 + IgG2, IgG1, and IgG2. The IgM and IgG1 isotype levels rose to 540–560 cells per mm^3 and 200–220 cells per mm^3, respectively, and remained there for a long time, concurrently evoking memory cells. Similarly, sustained growth of active CTLs and pre-activation of CTLs response were seen after each immunization. This indicated the long-established immune reaction triggered by designed vaccines. With a peak population of 11000–12000 cells per mm^3, a substantial number of HTLs stayed active during the simulation, reflecting the generation of an adaptive immune response against the antigen. Additionally, the production of different cytokines, particularly IFN-γ (maximum concentration >400000 ng/ml), was noticed. These findings indicate that the proposed multi-epitope vaccines have the potential to elicit a robust immune response against MPX infection. Nonetheless, *in-vitro*, and *in-vivo* evaluation of the proposed vaccine constructs is recommended to clarify their ability to elicit adaptive immunity against MPXV infection.

Binding free energy estimation

In thermodynamics calculation, the Gibbs free energy or Delta G (ΔG) (binding affinity of the complex) is vital to estimate the

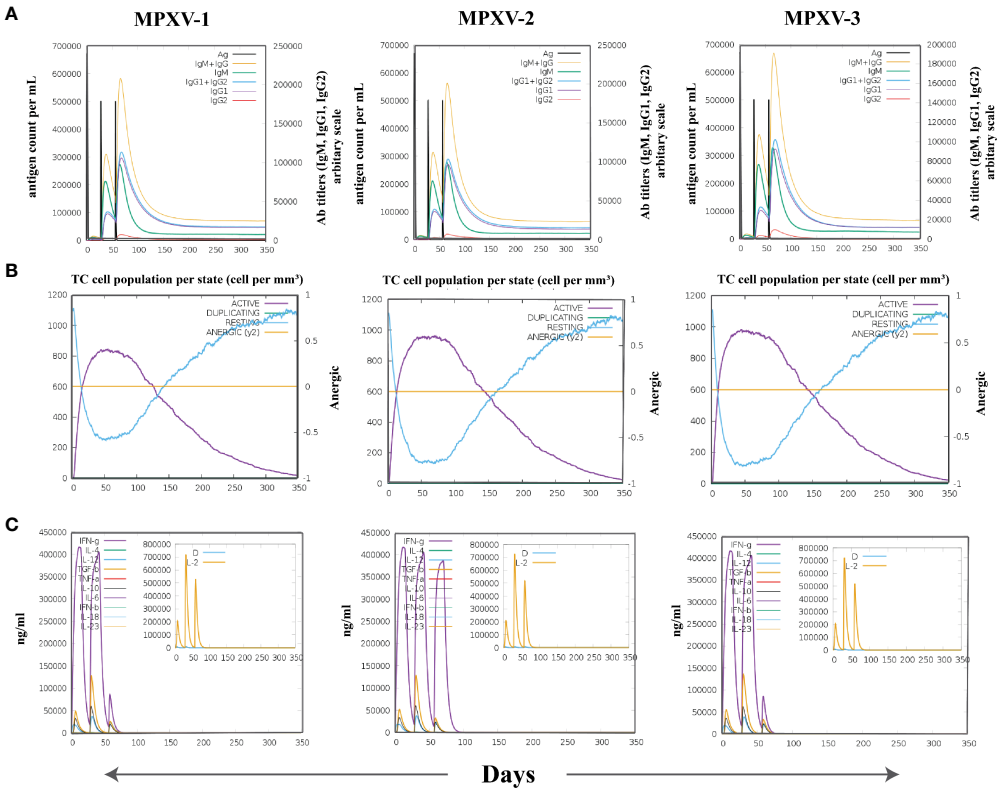


FIGURE 9
The spectrum of immune stimulated response. **(A)** Antibodies response against antigenic vaccine. Different colours reflected the formation of antibodies that signified the development of the immune response following vaccination and the diverse subtypes of immunoglobulins. **(B)** Cytotoxic T cell population in different states. **(C)** Cell population of cytokine and interleukin in the active and resting states with smaller graph depicting the Simpson index.

occurrence of a reaction under a cellular environment. To estimate the binding affinity of the MPXV-TLRs complex, MM/GBSA calculations were performed (Table 7). The total binding energy (ΔG TOTAL) of MPXV-1/2–TLR3 complex and MPXV-3–TLR2 complex was –103.77, –42.22, and –49.06 kcal/mol, respectively. The high negative ΔG value indicates each complex's stability and energetically favourable binding. Binding free energy terms, including gas-phase energy (ΔG_{gas} –793.61 kcal/mol) and electrostatic energy (ΔE_{elec} –521.40 kcal/mol), and Van der walls energy (ΔE_{vdW} –272.21 kcal/mol) contributed highest toward total binding energy of MPXV-1–TLR3 complex. In contrast, dominant energy terms in the MPXV-1/2–TLR3 complexes were solvation energy ($\Delta G_{solvation}$) and ΔE_{vdW}. ΔE_{elec} of MPXV-1 was negative compared to other constructs due to interactions of charge-charge residues between the vaccine and receptor. This can be attributed to structural adjustments within MPXV-1 exposing the charged residues that possibly established ionic interactions with TLR3. Furthermore, the MPXV-2/3 complex system had high solvation energy than the MPXV-1 system, indicating the structural alterations that likely made the buried regions accessible to solvent.

Discussion

The choice of appropriate antigens and their immunodominant epitopes is one of the current challenges in the field of multi-epitope vaccine design (17). ViPR comprises B-cell and T-cell epitopes that have undergone experimental validation and were produced from the IEDB for infectious illnesses, allergens, autoimmune disorders, transplant/alloantigens examined in mice, non-human primates, humans, and other animal species (66, 67). This experimental data repository can be exploited to investigate the known epitopes, their immunogenic responses and to formulate a new vaccine (66). In this study, a multi-epitope vaccine was developed for MPXV using the immune epitopes available at ViPR for this virus. Experimentally validated epitopes of MPXV were filtered down to candidate epitopes for vaccine designing by applying several *in silico* tools. Previously, Ismail et al. (68) evaluated the experimentally validated epitopes of hantavirus to shortlist the candidates epitopes for multi-epitope-based vaccine conception.

The antigens of origin of half of the selected epitopes appeared annotated as viral late transcription factor 1, which is an integral membrane component that regulates the DNA-templated transcription. Another antigen is essential IMV membrane protein (A15L), a major component of mature virion membrane necessary for membrane biogenesis. MIR [Monkeypox virus (strain Zaire-96-I-16)] and B6R (negative regulation of complement activation) are also integral membrane components of the virion. A single epitope antigen is a putative enzyme, DNA-directed RNA polymerase 7kDa subunit, with an intracellular localization that assists the synthesis of an RNA transcript from a DNA template. The host's immune system is more likely to interact with molecules found on the surface of microorganisms, such as plasma membrane proteins, which also play crucial roles in pathogen homeostasis (69). The membrane proteins of the vaccinia virus have been assessed as a part of subunit vaccines and target of neutralizing antibodies *in-vivo* and *in-vitro* (70).

The sequence's antigenicity prediction is essential because it reveals whether the epitope sequence is likely to be recognized by immunogenic cells in the human body. Another significant barrier to vaccine development is the possibility of allergenicity, which occurs when numerous vaccines drive the immune system into an allergic response (71). Thus, herein, we chose epitopes predicted to be highly antigenic and did not show allergenic potential for the host immune system. On practical grounds, the findings that virulence factors are often the targets of effective immune response could drive the designing of vaccines. Although this notion works well for bacterial pathogens, it can also be suitable for viruses (72). Therefore, the virulence potential of epitopes was tested, and those predicted as virulent were retained. Besides, the non-toxicity of selected epitopes indicates their unharmful effects on the human body (73).

A critical factor in prioritizing the epitope was avoiding any potential cross-reactivity with the host. In general, the least the

TABLE 7 MM/GBSA free energy calculations and individual free energy components of vaccine-TLR complex.

Energy Component	MM/GBSA (kcal/mol)					
	MPXV-1–TLR3 complex		MPXV-2–TLR3 complex		MPXV-3–TLR2 complex	
	Average	Std. Err. Mean	Average	Std. Err. Mean	Average	Std. Err. Mean
ΔE_{vdW}	–272.21	0.33	–129.54	0.41	–120.97	0.33
ΔE_{elec}	–521.40	2.28	322.23	2.77	346.34	2.00
ΔE_{GB}	725.66	2.22	–217.95	2.93	–259.64	1.97
ΔE_{SURF}	–35.82	0.04	–16.95	0.04	–14.79	0.04
ΔG_{gas}	–793.61	2.25	192.68	2.97	225.37	2.02
$\Delta G_{solvation}$	689.84	2.21	–234.90	2.90	–274.43	1.96
ΔG_{TOTAL}	–103.77	0.29	–42.22	0.25	–49.063	0.20

sequence identity, the higher the likelihood of preventing any possible cross-reactivity. Nonetheless, no research has been done to determine the sequence similarity percentage that may prevent cross-reactivity between the recognition of self-peptides and epitopes (74). When it comes to lengthy epitopes like CD4+ T cell and B-cell epitopes, a 70% cut-off has been defined as conservative by a strategy intended to reduce epitope redundancy (75). In terms of nonameric CD8+ T cell epitopes, it has been noted that residue positions 2 and 9 are crucial for determining epitope binding to MHC-I. In contrast, residues 3-6 and 8 are involved in TCR motif recognition (76). Following Michel-Todó *et al.* (77), a threshold of 70% could limit the selection of any MPXV epitopes similar to human peptides in more than six (out of nine) residues. Thus, at least one of those crucial places will always be different in any of the CD8+ T cell epitopes from any peptide sequence of that length found in humans. However, as previously stated, cross-recognition may still occur with little similarity; therefore, any possible cross-reactivity must be managed appropriately (77). As structural understanding of the epitope recognition mechanism improves, we think computational methods will soon be able to predict this more accurately. Besides non-homology, the prioritized epitopes sequences showed 100% conservancy with their source proteins sequences from reference MPXV Zaire 77-0666 (NCBI Reference Sequence: NC_003310.1) and MPXV (NCBI Reference Sequence: NC_063383.1). The idea is that a single vaccine would cover the virus' vast phylogenetic and geographic range. In addition, favourable physicochemical properties of selected epitopes reveal their suitability for application in wet-lab investigations. When choosing the antigens or epitopes that will be included in the vaccine, computational tools can help predict vaccine coverage. Since experimental HLA-restriction alleles of each CTL epitope are reported, we first evaluated whether the IEDB SMM method predicts the reported alleles as high-affinity binders of the respective epitopes. The employed method not only correctly estimated the reported HLA alleles as high-affinity CTL epitopes binders but also revealed additional potent epitopes binders (HLA alleles). Hence, we considered all reported and predicted HLA alleles of CTL epitopes for the population coverage analysis *via* the IEDB tool. The prioritized epitopes were predicted to cover a broad range of the world population (>80% coverage).

Recently, Rawal et al. (25) introduced a new computational system (Vax-ELAN pipeline) to evaluate novel vaccine targets that could act as potential candidates for multi-epitope vaccine construction. They screened the genomic and proteomic datasets of *Vibrio cholerae*, *Plasmodium falciparum*, and *Trypanosoma cruzi* to extract the PVC. Amongst these pathogens, *T. cruzi* proteome was subjected to thorough analysis to shortlist eight proteins as PVC using several properties, for example surface exposed/secretory proteins, virulence, least number of TM helices, antigenicity, non-allergenicity, gene essentiality, non-homology with human proteins and gut microbiota. Further, the researcher subjected the shortlisted proteins to B-cell and T-cell epitopes mapping in order to eventually design a putative multiple epitope vaccine candidate for *T. cruzi*. In the current study, we used Vax-ELAN pipeline together with other immunoinformatics tools to evaluate and shortlist eight MPXV proteins considering the exposed nature, non-allergen, antigen, least number of TM helices (≤1), non-homology, and suitable physiochemical profile. The finalized MPXV proteins (PVC) can be used to map novel B-cell and T-cell epitopes and design a multi-peptide-based vaccine against this virus.

Subunit vaccines need to be administered with an adjuvant because, although safer, they are often less immunogenic and efficacious. Adjuvants are thus crucial for augmenting and guiding the adaptive immune response to vaccine antigens (78). Based on literature mining (68, 79, 80), we used three types of adjuvants to construct MPXV vaccines, including beta-defensin 3, ribosomal protein, and HBHA. Human beta-defensin 3 is an antimicrobial peptide that can play an important role as an immunomodulator to activate human antigen-presenting cells (APCs) (81). 50S ribosomal protein L7/L12 has been shown to induce the maturation of dendritic cells (DCs), CD4+, CD8+, and IFN-producing cells following the stimulation of naïve T cells (21). HBHA is a novel TLR4 agonist with no systemic toxicity. It has a strong immunostimulatory potential and the ability to stimulate the maturation of DCs in a TLR4-dependent manner (82). Following Robert et al. (18), CTL and B-cell epitopes were linked together by inserting AAY and GPGPG linker. An EAAAK and GGGS linker (provides flexible distribution) were added to link the PADRE sequence and an adjuvant and separate the epitope class, respectively, as previously reported (32). The TAT sequence can transport the macromolecular substances and makes the cell membrane penetration easier to stimulate the phagocytosis of vaccines by APCs. Therefore, the TAT sequence was incorporated at the N-termini of constructed vaccines (83).

The proposed vaccines in this study had reasonable molecular weight, i.e., less than 110kDa, which hinted that it is suitable for application (84). The estimated half-life in *E. coli* (10 hrs), instability index (19.39 – 27.43), and aliphatic index (84.02 – 92.99) of the vaccine constructs point towards the usage of this organism for heterologous expression. That is why we optimized the codon of constructs based on *E. coli* (strain K12) and carried out computational cloning in pET28a (+), a common expression vector. Furthermore, the predicted pI indicated that MPXV-1 (pI=7.66) would assume a positive charge under neutral pH conditions, while MPXV-2/3 (pI= 4.68 and 4.70, respectively) would be negatively charged. The negative net charge leans towards assisting in (Ni^{2+}) affinity purification procedure, like those carried in Nickel immobilized columns (85, 86). Further, considerable structural alterations in protein are unlikely to happen in such pH conditions. The successful manufacturing process is verified by solubility predictions of all designed vaccines. The secondary structure analysis showed that the

vaccine proteins have 33% to 48% random coil. Since the spatial structure of random coil is loose and it is easy to form epitopes; thus, the modeled vaccines have fine structural basis because they contain large amount of this structure. Hashemzadeh et al. (87), Rekik et al. (88), and Droppa-Almeida et al. (89) predicted tertiary structure of constructed vaccine with a Z-score of −2.11 kcal/mol, −9.5 kcal/mol, and −5.26 kcal/mol, respectively. They concluded that the predicted structures were reliable and of good quality. In the present study, the predicted tertiary structures of proposed vaccines (Z-score of MPXV-1–3 = −4.25, −5.62, and −5.87) resemble the structural properties of similar-sized proteins determined by the X-Ray crystallography.

When microorganisms break through the mucosal barrier, TLR can recognize them and initiate the adaptive immune response (84). Because the TLR3 molecule plays a vital role in innate immunity against the virus, the binding of antigen and TLR3 can help APCs in antigen presentation and release of local cytokines. Data also suggest that TLR3 can detect certain DNA viruses (90). We conducted protein-protein docking to prove that the designed vaccines had a strong affinity with the TLR3 receptor (79). The results showed the lowest energy value (more stable) for vaccine-TLR3 complexes. Further, ionic and hydrogen bonding also indirectly supported the stable binding between vaccine constructs and TLR3 (91). MPXV-1/2 had high docking energy with TLR3 and therefore was subjected to additional analysis. We also performed docking of designed vaccines with TLR2 and TLR4 receptors, which were previously studied in the context of their role in viral recognition and induction of inflammatory cytokines (92). All vaccine proteins demonstrated reasonable docking energy with these immune receptors (with several intermolecular interactions), implying their potential to stimulate the downstream immune responses. Interestingly, MPXV-3 indicated higher docking energy with TLR2 than other receptors; hence was recommended for further investigations.

Coherently, RMSD measures obtained by molecular dynamic simulation have indicated stable interaction between the ligand and the receptor. The agility and compactness of the vaccine-TLRs complex structure were also supported by the RMSF and Rg profile of the complexes. Moreover, structural adjustments in vaccine protein structure upon binding TLRs were revealed by FEL and SASA descriptors, which seem crucial to gathering stability for exposure to the immune system. H-bond analysis showed persistent H-bonds between vaccine and TLRs across the simulation, which is critical for complex stability (93). Consistent with the docking analysis, MM/GBSA estimations validated a high binding affinity (low negative ΔG values) of vaccine-TLR2/3 complexes.

A relevant antibodies generation after vaccination was predicted by the C-ImmSim server, which is also regarded as crucial for preventing infection. Memory B-cells can present antigens on the cell surface, activate HTLs, and serve as the subsequent target of effector HTLs in secondary antibody responses (94). Immune simulation prediction in this study also revealed significant memory B-cells production, which has appeared to engage HTLs as evident by their high numbers. The generation of HTLs is also attributed to GPGPG linkers added to vaccine constructs. Besides, the secretion of IFN-γ, long-lasting cellular and CTL response was detected during the immune response. *In silico* immune simulation results of the present study are comparable to previous investigations that used B-cell, CTL, and HTL epitopes in vaccine construction (21, 32, 95). The data highlights the potential of designed vaccine constructs to induce a robust immune response and provide protection against MPXV infection. To confirm the results of the current study, however, experimental validation of the constructed vaccines is warranted.

Limitations

This study highlighted an alternative vaccine design approach based on a multi-epitope ensemble of the antigenic proteins of the MPX genome to handle the antigenic complexity. The constructed vaccines were recommended based on immunoinformatics-guided evaluation and are believed to be immunogenic; however, the extent of protection from MPX infection is unknown. Concerning T cell components, a population coverage estimation was based on a binding prediction of peptide-MHC molecules. Although the reliability of peptide-MHC affinity predictions has been extensively proved, this feature will have to be tested further. A suitable antigen processing is a vital feature for the immunogenicity of epitopes, and this would need thorough evaluation in the context of epitope delivery in the form of a genetic construct. The ordering and spacing of the CD8+ T cell and B-cell epitopes herein were provided following the literature. Still, this would require more proof to determine the optimal ordering and spacing to acquire the best immunogenicity with CTL and B-cell epitopes. Omics database repository, including ViPR, and several immunoinformatics tools described in the manuscript, are extremely useful for carrying out *in silico* studies that can guide the web-lab experiments, contributing to time and money-saving. However, *in-vitro* immunological studies will be the next step to ascertain the designed vaccines' immunogenicity and design challenge-protection preclinical experiments to ultimately validate the strategy.

Conclusion

Multi-epitope vaccines have gained significant attention, with many showing protective efficacy *in-vitro* and progressed to clinical trials. The goal of the present study was to propose an alternative approach to prioritize the experimentally determined epitopes of MPXV for inclusion in designing a multi-epitope-based vaccine through an immunoinformatics-driven approach. The physicochemical and antigenic properties of constructed vaccines were analyzed computationally. The stability profile and molecular

interactions between the proposed vaccines and immunological receptors were examined using molecular docking and molecular dynamics simulations. Binding free energy calculations confirmed the stable and energetically favourable binding of MPXV-1–TLR3 (-103.7705 kcal/mol), MPXV-2–TLR3 (-42.2205 kcal/mol), and MPXV-3–TLR2 (-49.0633 kcal/mol). The proposed ensembles' potential to evoke cellular and humoral immunity against the virus was demonstrated by computational immune simulation. Using various immunoinformatics and *in silico*-based methods, protective vaccines against viral infection were developed and evaluated herein; nevertheless, experimental testing is required to ascertain the vaccine efficacy. Additionally, we encourage further research on the production and biological activities of the developed multi-epitope vaccines.

Data availability statement

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

Author contributions

SA, FA and AA conceived and designed the study. SA, MW and FA performed experiments. MA, IU and NK analyzed the data. SA, MW, MA and AA wrote the manuscript with inputs and comments from all co-authors. All authors contributed to the article and approved the submitted version.

Funding

The authors thank the Researchers Supporting Project number (RSP-2021/198), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1004804/full#supplementary-material

References

1. Sklenovska N, Van Ranst M. Emergence of monkeypox as the most important orthopoxvirus infection in humans. *Front Public Health* (2018) 6:241. doi: 10.3389/fpubh.2018.00241

2. Marennikova S, Moyer R. Classification of poxviruses and brief characterization of the genus orthopoxvirus. In: *Orthopoxviruses pathogenic for humans* (2005) (New York, NY:Springer) .

3. Shchelkunov SN, Totmenin AV, Safronov PF, Mikheev MV, Gutorov VV, Ryazankina OI, et al. Analysis of the monkeypox virus genome. *Virology* (2002) 297 (2):172–94. doi: 10.1006/viro.2002.1446

4. Xiang Y, White A. Monkeypox virus emerges from the shadow of its more infamous cousin: family biology matters. *Emerg Microbes Infect* (2022) 11(1):1768–77. doi: 10.1080/22221751.2022.2095309

5. Bunge EM, Hoet B, Chen L, Lienert F, Weidenthaler H, Baer LR, et al. The changing epidemiology of human monkeypox-a potential threat? a systematic review. *PloS Negl Trop Dis* (2022) 16(2):e0010141. doi: 10.1371/journal.pntd.0010141

6. World Health Organization. WHO director-general's statement at the press conference following IHR emergency committee regarding the multi-country outbreak of monkeypox. (2022) Available at: https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-the-press-conference-following-IHR-emergency-committee-regarding-the-multi–country-outbreak-of-monkeypox–23-july-2022 [Accessed July 27 2022).

7. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med* (2022) 28(8):1569–72. doi: 10.1038/s41591-022-01907-y

8. Fenner F, Henderson D, Arita I, Jezek Z, Ladnyi I. Human monkeypox and other poxvirus infections of man. In: *Smallpox and its eradication*. (1988) (Geneva, Switzerland:World health organization):1287–319.

9. Fine PE, Jezek Z, Grab B, Dixon H. The transmission potential of monkeypox virus in human populations. *Int J Epidemiol* (1988) 17(3):643–50. doi: 10.1093/ije/17.3.643

10. Jezek Z, Khodakevich LN, Wickett JF. Smallpox and its post-eradication surveillance. In: *Bulletin of the world health organization* (1987) (Geneva, Switzerland:World Health Organization) 65(4):425–34.

11. Heraud JM, Edghill-Smith Y, Ayala V, Kalisz I, Parrino J, Kalyanaraman VS, et al. Subunit recombinant vaccine protects against monkeypox. *J Immunol* (2006) 177(4):2552–64. doi: 10.4049/jimmunol.177.4.2552

12. Centers for Disease Control and Prevention. Interim Clinical Considerations for Use of JYNNEOS and ACAM2000 Vaccines during the 2022 U.S. Monkeypox Outbreak. (2022) Available at: https://www.cdc.gov/poxvirus/monkeypox/health-departments/vaccine-considerations.html [Accessed August 17 2022].

13. Petersen E, Zumla A, Hui DS, Blumberg L, Valdoleiros SR, Amao L, et al. Vaccination for monkeypox prevention in persons with high-risk sexual behaviours to control on-going outbreak of monkeypox virus clade 3. *Int J Infect Dis* (2022) 122:569–71. doi: 10.1016/j.ijid.2022.06.047

14. Rao AK, Petersen BW, Whitehill F, Razeq JH, Isaacs SN, Merchlinsky MJ, et al. Use of JYNNEOS (Smallpox and monkeypox vaccine, live, nonreplicating) for preexposure vaccination of persons at risk for occupational exposure to orthopoxviruses: Recommendations of the advisory committee on immunization practices - united states, 2022. *MMWR Morb Mortal Wkly Rep* (2022) 71(22):734–42. doi: 10.15585/mmwr.mm7122e1

15. Centers for Disease Control and Prevention. Monkeypox and Smallpox Vaccine Guidance. (2022). Available at: https://www.cdc.gov/poxvirus/monkeypox/clinicians/smallpox-vaccine.html [Accessed August 17 2022].

16. Franceschi V, Parker S, Jacca S, Crump RW, Doronin K, Hembrador E, et al. BoHV-4-based vector single heterologous antigen delivery protects STAT1 (-/-) mice from monkeypoxvirus lethal challenge. *PloS Negl Trop Dis* (2015) 9(6): e0003850. doi: 10.1371/journal.pntd.0003850

17. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* (2018) 15(2):182–4. doi: 10.1038/cmi.2017.92

18. Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, et al. *In-silico* design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci Rep* (2019) 9(1):4409. doi: 10.1038/s41598-019-40833-x

19. Ali A, Khan A, Kaushik AC, Wang Y, Ali SS, Junaid M, et al. Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein-Barr virus (EBV). *Sci Rep* (2019) 9(1):720. doi: 10.1038/s41598-018-37070-z

20. Amanna IJ, Slifka MK. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. *Virology* (2011) 411(2):206–15. doi: 10.1016/j.virol.2010.12.016

21. Sami SA, Marma KKS, Mahmud S, Khan MAN, Albogami S, El-Shehawi AM, et al. Designing of a multi-epitope vaccine against the structural proteins of marburg virus exploiting the immunoinformatics approach. *ACS Omega* (2021) 6 (47):32043–71. doi: 10.1021/acsomega.1c04817

22. Aziz S, Waqas M, Halim SA, Ali A, Iqbal A, Iqbal M, et al. Exploring whole proteome to contrive multi-epitope-based vaccine for NeoCoV: An immunoinformtics and *in-silico* approach. *Front Immunol* (2022) 13:956776. doi: 10.3389/fimmu.2022.956776

23. Mahmud S, Rafi MO, Paul GK, Promi MM, Shimu MSS, Biswas S, et al. Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Sci Rep* (2021) 11(1):15431. doi: 10.1038/s41598-021-92176-1

24. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2–a server for *in silico* prediction of allergens. *J Mol Model* (2014) 20(6):2278. doi: 10.1007/s00894-014-2278-5

25. Rawal K, Sinha R, Abbasi BA, Chaudhary A, Nath SK, Kumari P, et al. Identification of vaccine targets in pathogens and design of a vaccine using computational approaches. *Sci Rep* (2021) 11(1):17626. doi: 10.1038/s41598-021-96863-x

26. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* (2007) 8(1):4. doi: 10.1186/1471-2105-8-4

27. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery C, et al. *In silico* approach for predicting toxicity of peptides and proteins. *PloS One* (2013) 8(9):e73957. doi: 10.1371/journal.pone.0073957

28. Garg A, Gupta D. VirulentPred: A SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinf* (2008) 9(1):62. doi: 10.1186/1471-2105-9-62

29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* (1990) 215(3):403–10. doi: 10.1016/s0022-2836(05)80360-2

30. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* (2022) 2022:4.08.487609. doi: 10.1101/2022.04.08.487609

31. Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide vaccine: Progress and challenges. *Vaccines (Basel)* (2014) 2(3):515–36. doi: 10.3390/vaccines2030515

32. Fadaka AO, Sibuyi NRS, Martin DR, Goboza M, Klein A, Madiehe AM, et al. Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. *Sci Rep* (2021) 11(1):19707. doi: 10.1038/s41598-021-99227-7

33. Nezafat N, Ghasemi Y, Javadi G, Khoshnoud MJ, Omidinia E. A novel multi-epitope peptide vaccine against cancer: an *in silico* approach. *J Theor Biol* (2014) 349:121–34. doi: 10.1016/j.jtbi.2014.01.018

34. Dong R, Chu Z, Yu F, Zha Y. Contriving multi-epitope subunit of vaccine for COVID-19: Immunoinformatics approaches. *Front Immunol* (2020) 11:1784. doi: 10.3389/fimmu.2020.01784

35. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* (2014) 30(6):846–51. doi: 10.1093/bioinformatics/btt619

36. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-sol: A web tool for predicting protein solubility from sequence. *Bioinformatics* (2017) 33(19):3098–100. doi: 10.1093/bioinformatics/btx345

37. Magnan CN, Randall A, Baldi P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* (2009) 25(17):2200–7. doi: 10.1093/bioinformatics/btp386

38. Fadaka AO, Sibuyi NRS, Madiehe AM, Meyer M. Computational insight of dexamethasone against potential targets of SARS-CoV-2. *J Biomol Struct Dyn* (2022) 40(2):875–85. doi: 10.1080/07391102.2020.1819880

39. Abedi Karjiban R, Abdul Rahman MB, Basri M, Salleh AB, Jacobs D, Abdul Wahab H. Molecular dynamics study of the structure, flexibility and dynamics of thermostable l1 lipase at high temperatures. *Protein J* (2009) 28(1):14–23. doi: 10.1007/s10930-008-9159-7

40. Hoie MH, Kiehl EN, Petersen B, Nielsen M, Winther O, Nielsen H, et al. NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res* (2022) 50(W1):W510–W515. doi: 10.1093/nar/gkac439

41. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf* (2008) 9(1):40. doi: 10.1186/1471-2105-9-40

42. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (2021) 373(6557):871–6. doi: 10.1126/science.abj8754

43. Laskowski R, MacArthur M, Thornton J. PROCHECK: validation of protein-structure coordinates. In: *International tables for crystallography Volume F Chapter 21.4* (2006) :684–7 Willey Online Library. doi: 10.1107/97809553602060000882

44. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *J Bioinformatics* (2011) 27(3):343–350. doi: 10.1093/bioinformatics/btq662

45. Colovos C, Yeates T. Verification of protein structures: patterns of nonbonded atomic interactions. In: *Protein Sc* (1993) 2(9):1511–1519. doi: 10.1002/pro.5560020916

46. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* (2016) 44(W1):W449–54. doi: 10.1093/nar/gkw329

47. C. C. G. ULC. *Molecular operating environment. 2022:02*. Sherbrooke St. West, Suite #910, Montreal, QC, Canada: Chemical Computing Group ULC (2022).

48. Hess R. Blender foundations. In: *The essential guide to learning blender 2.5*. (New York:Routledge) (2013).

49. Grote A, Hiller K, Scheer M, Munch R, Nortemann 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(Web Server issue):W526–31. doi: 10.1093/nar/gki376

50. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res* (2008) 36(Web Server issue):W70–4. doi: 10.1093/nar/gkn188

51. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PloS One* (2010) 5(4):e9862. doi: 10.1371/journal.pone.0009862

52. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh ewald. *J Chem Theory Comput* (2013) 9(9):3878–88. doi: 10.1021/ct400314y

53. Case D, Belfon K, Ben-Shalom I, Brozell S, Cerutti D, Cheatham T, et al. *AMBER 2022: University of California*. University of California, San Francisco (2022).

54. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, et al. ff19SB: Amino-Acid-Specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* (2020) 16(1):528–52. doi: 10.1021/acs.jctc.9b00591

55. Sengupta A, Li Z, Song LF, Li P, Merz KMJr. Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J Chem Inf Model* (2021) 61(2):869–80. doi: 10.1021/acs.jcim.0c01390

56. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical recipes in C (2nd ed.): The art of scientific computing*. (40 W. 20 St. New York, NY United States: Cambridge University Press) (1992). doi: 10.5555/148286

57. Sindhikara DJ, Kim S, Voter AF, Roitberg AE. Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. *J Chem Theory Comput* (2009) 5(6):1624–31. doi: 10.1021/ct800573m

58. Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* (2001) 22(5):501–8. doi: 10.1002/1096-987X(20010415)22:5<501:AID-JCC1021>3.0.CO;2-V

59. Darden T, York D, Pedersen L. Particle mesh ewald: An n· log (N) method for ewald sums in large systems. *J Chem Phys* (1993) 98(12):10089–92. doi: 10.1063/1.464397

60. Roe DR, Cheatham TE3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* (2013) 9(7):3084–95. doi: 10.1021/ct400341p

61. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* (2011) 51(1):69–82. doi: 10.1021/ci100275a

62. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* (2004) 55(2):383–94. doi: 10.1002/prot.20033

63. Rahman MU, Liu H, Wadood A, Chen HF. Allosteric mechanism of cyclopropylindolobenzazepine inhibitors for HCV NS5B RdRp *via* dynamic correlation network analysis. *Mol Biosyst* (2016) 12(11):3280–93. doi: 10.1039/c6mb00521g

64. Dong R, Chu Z, Yu F, Zha Y. Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front Immunol* (2020) 11:1784. doi: 10.3389/fimmu.2020.01784

65. Frankel AD, Pabo CO. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* (1988) 55(6):1189–93. doi: 10.1016/0092-8674(88)90263-2

66. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* (2017) 8:278. doi: 10.3389/fimmu.2017.00278

67. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, et al. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res* (2012) 40(Database issue):D593–8. doi: 10.1093/nar/gkr859

68. Ismail S, Abbasi SW, Yousaf M, Ahmad S, Muhammad K, Waheed Y. Design of a multi-epitopes vaccine against hantaviruses: An immunoinformatics and molecular modelling approach. *Vaccines (Basel)* (2022) 10(3):378. doi: 10.3390/vaccines10030378

69. Gadelha C, Zhang W, Chamberlain JW, Chait BT, Wickstead B, Field MC. Architecture of a host–parasite interface: Complex targeting mechanisms revealed through proteomics. *Mol Cell Proteomics* (2015) 14(7):1911–26. doi: 10.1074/mcp.M114.047647

70. Lawrence SJ, Lottenbach KR, Newman FK, Buller RM, Bellone CJ, Chen JJ, et al. Antibody responses to vaccinia membrane proteins after smallpox vaccination. *J Infect Dis* (2007) 196(2):220–9. doi: 10.1086/518793

71. 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(1):21. doi: 10.1186/s13099-022-00495-z

72. Casadevall A, Pirofski L-a. Virulence factors and their mechanisms of action: The view from a damage–response framework. *J Water Health* (2009) 7(S1):S2–S18. doi: 10.2166/wh.2009.036

73. Murphy K, Janeway CA, Travers P, Walport M, Mowat A, Weaver CT. Janeway's immunobiology. *Garland Sci* (2012) 87(3) (London and New York: Garland Science (Taylor & Francis Group). doi: 10.1086/666778

74. Petrova G, Ferrante A, Gorski J. Cross-reactivity of T cells and its role in the immune system. *Crit Rev Immunol* (2012) 32(4):349–72. doi: 10.1615/critrevimmunol.v32.i4.50

75. Caoili SEC. Expressing redundancy among linear-epitope sequence data based on residue-level physicochemical similarity in the context of antigenic cross-reaction. *Adv Bioinf* (2016) 2016:1276594. doi: 10.1155/2016/1276594

76. Frankild S, de Boer RJ, Lund O, Nielsen M, Kesmir C. Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. *PloS One* (2008) 3(3):e1831. doi: 10.1371/journal.pone.0001831

77. Michel-Todo L, Reche PA, Bigey P, Pinazo MJ, Gascon J, Alonso-Padilla J. *In silico* design of an epitope-based vaccine ensemble for chagas disease. *Front Immunol* (2019) 10:2698. doi: 10.3389/fimmu.2019.02698

78. Apostolico J, Lunardelli VAS, Coirada FC, Boscardin SB, Rosa DS. Adjuvants: Classification, modus operandi, and licensing. *J Immunol Res* (2016) 2016:1459394. doi: 10.1155/2016/1459394

79. Fadaka AO, Sibuyi NRS, Martin DR, Goboza M, Klein A, Madiehe AM, et al. Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. *Sci Rep* (2021) 11(1):1–22. doi: 10.1038/s41598-021-99227-7

80. Rahman N, Ali F, Basharat Z, Shehroz M, Khan MK, Jeandet P, et al. Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: An immunoinformatics approach. *Vaccines (Basel)* (2020) 8(3):423. doi: 10.3390/vaccines8030423

81. Funderburg N, Lederman MM, Feng Z, Drage MG, Jadlowsky J, Harding CV, et al. Human β-defensin-3 activates professional antigen-presenting cells *via* toll-like receptors 1 and 2. *Proc Natl Acad Sci* (2007) 104(47):18631–5. doi: 10.1073/pnas.0702130104

82. Lei Y, Shao J, Ma F, Lei C, Chang H, Zhang Y. Enhanced efficacy of a multi-epitope vaccine for type a and O footand-mouth disease virus by fusing multiple epitopes with mycobacterium tuberculosis heparin-binding hemagglutinin (HBHA), a novel TLR4 agonist. *Mol Immunol* (2020) 121:118–26. doi: 10.1016/j.molimm.2020.02.018

83. Yu M, Zhu Y, Li Y, Chen Z, Li Z, Wang J, et al. Design of a recombinant multivalent epitope vaccine based on SARS-CoV-2 and its variants in immunoinformatics approaches. *Front Immunol* (2022) 13:884433. doi: 10.3389/fimmu.2022.884433

84. Barh D, Barve N, Gupta K, Chandra S, Jain N, Tiwari S, et al. Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in vibrio cholerae targeted by piper betel derived compounds. *PloS One* (2013) 8(1):e52773. doi: 10.1371/journal.pone.0052773

85. Cheung RC, Wong JH, Ng TB. Immobilized metal ion affinity chromatography: A review on its applications. *Appl Microbiol Biotechnol* (2012) 96(6):1411–20. doi: 10.1007/s00253-012-4507-0

86. Nelson DL, Lehninger AL, Cox MM. *Lehninger principles of biochemistry*. Freeman WH, editor. (New York:W.H. Freeman & Co) (2008).

87. Hashemzadeh P, Ghorbanzadeh V, Lashgarian HE, Kheirandish F, Dariushnejad H. Harnessing bioinformatic approaches to design novel multi-epitope subunit vaccine against leishmania infantum. *Int J Pept Res Ther* (2020) 26(3):1417–28. doi: 10.1007/s10989-019-09949-6

88. Rekik I, Chaabene Z, Grubb CD, Drira N, Cheour F, Elleuch A. *In silico* characterization and molecular modeling of double-strand break repair protein MRE11 from phoenix dactylifera v deglet nour. *Theor Biol Med Model* (2015) 12(1):23. doi: 10.1186/s12976-015-0013-2

89. Droppa-Almeida D, Franceschi E, Padilha FF. Immune-informatic analysis and design of peptide vaccine from multi-epitopes against corynebacterium pseudotuberculosis. *Bioinform Biol Insights* (2018) 12:1177932218755337. doi: 10.1177/1177932218755337

90. Sun Y, Reddy P. Intracellular sensors of immunity and allogeneic hematopoietic stem cell transplantation. In: *Immune biology of allogeneic hematopoietic stem cell transplantation*. (Amsterdam:Elsevier) (2013).

91. Jia S, Li C, Pan H, Wang M, Wang X, Lin Q. Preparation and pore-forming mechanism of hydrogen bond and ionic bond double-driven chitosan-based mesoporous carbon. *Int J Biol Macromol* (2021) 179:519–31. doi: 10.1016/j.ijbiomac.2021.03.024

92. Compton T, Kurt-Jones EA, Boehme KW, Belko J, Latz E, Golenbock DT, et al. Human cytomegalovirus activates inflammatory cytokine responses *via* CD14 and toll-like receptor 2. *J Virol* (2003) 77(8):4588–96. doi: 10.1128/jvi.77.8.4588-4596.2003

93. Lippert T, Rarey M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J Cheminform* (2009) 1(1):13. doi: 10.1186/1758-2946-1-13

94. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Helper T cells and lymphocyte activation. In: *Molecular biology of the cell. 4th edition*. (New York: Garland Science) (2002).

95. Sarkar B, Ullah MA, Araf Y, Islam NN, Zohora US. Immunoinformatics-guided designing and *in silico* analysis of epitope-based polyvalent vaccines against multiple strains of human coronavirus (HCoV). *Expert Rev Vaccines* (2021), 1–21. doi: 10.1080/14760584.2021.1874925

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fmed.2022.945938

OPEN ACCESS

EDITED BY
Adriana Harbuzariu,
Emory University, United States

REVIEWED BY
Ahmed A. Al-Karmalawy,
Horus University, Egypt
Muhammad Alaa Eldeen,
Zagazig University, Egypt
Hamidreza Majidiani,
Neyshabur University of Medical Sciences, Iran

*CORRESPONDENCE
Noorah Alsowayeh
n.alsowayeh@mu.edu.sa
Aqel Albutti
as.albutti@qu.edu.sa

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

RECEIVED 17 May 2022
ACCEPTED 20 September 2022
PUBLISHED 18 October 2022

CITATION
Alsowayeh N and Albutti A (2022) Designing a novel chimeric multi-epitope vaccine against *Burkholderia pseudomallei,* a causative agent of melioidosis. *Front. Med.* 9:945938. doi: 10.3389/fmed.2022.945938

Designing a novel chimeric multi-epitope vaccine against *Burkholderia pseudomallei,* a causative agent of melioidosis

Noorah Alsowayeh[1*†] and Aqel Albutti[2*†]

[1]Department of Biology, College of Education (Majmaah), Majmaah University, Al Majmaah, Saudi Arabia, [2]Department of Medical Biotechnology, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia

Burkholderia pseudomallei, a gram-negative soil-dwelling bacterium, is primarily considered a causative agent of melioidosis infection in both animals and humans. Despite the severity of the disease, there is currently no licensed vaccine on the market. The development of an effective vaccine against *B. pseudomallei* could help prevent the spread of infection. The purpose of this study was to develop a multi-epitope-based vaccine against *B. pseudomallei* using advanced bacterial pan-genome analysis. A total of four proteins were prioritized for epitope prediction by using multiple subtractive proteomics filters. Following that, a multi-epitopes based chimeric vaccine construct was modeled and joined with an adjuvant to improve the potency of the designed vaccine construct. The structure of the construct was predicted and analyzed for flexibility. A population coverage analysis was performed to evaluate the broad-spectrum applicability of *B. pseudomallei*. The computed combined world population coverage was 99.74%. Molecular docking analysis was applied further to evaluate the binding efficacy of the designed vaccine construct with the human toll-like receptors-5 (TLR-5). Furthermore, the dynamic behavior and stability of the docked complexes were investigated using molecular dynamics simulation, and the binding free energy determined for Vaccine-TLR-5 was delta total −168.3588. The docking result revealed that the vaccine construct may elicit a suitable immunological response within the host body. Hence, we believe that the designed *in-silico* vaccine could be helpful for experimentalists in the formulation of a highly effective vaccine for *B. pseudomallei.*

KEYWORDS

***Burkholderia pseudomallei*, melioidosis, immunoinformatics, molecular dynamics simulation, vaccine**

Introduction

Burkholderia pseudomallei, a member of the family *Burkholderiaceae*, is a rod-shaped, gram-negative, motile, and multitrichous flagella bacterium typically of 1–5 μm long and diameter of 0.5–1.0 μm (1). Melioidosis, often known as Whitmore's infection, is caused by *B. pseudomallei* and is characterized by

high mortality and morbidity rates (2). The infection usually leads to abscess formation and sepsis. This pathogen infects both animals and humans when they come into contact with contaminated soil (3). The bacteria can be spread from an infected person to a non-infected person and can be acquired in the hospital environment as well. Despite the rarity of zoonotic transmission and animal-to-animal infection transfer, a wide variety of species have contracted the disease (4). The bacterium *B. pseudomallei* is thought to have a role in horizontal gene transfer (4). Melioidosis infection is prevalent in tropical regions while the main endemic regions include Southeast Asia and northern Australia (5, 6). Every year, around 165,000 cases of Melioidosis are reported worldwide, with 89,000 deaths (7). A high mortality rate of 42.6% was estimated in Thailand from 1997 to 2006 (6). By 2018, 30–35% of infections in Thailand's hospitals resulted in deaths (8). The bacterium is regarded as a biological weapon, and it has been linked to diabetes, chronic lung and renal diseases, malignancies, and heart diseases. Earlier diagnosis of the disease can decrease the mortality rate. Melioidosis chronic infection is frequently associated with the risk factors indicated above (9). The remarkable *B. pseudomallei*'s adaptative characteristics allow the bacteria to survive in the host body and neutralize host immune responses (2). As this infection is associated with many clinical manifestations, its diagnosis is considered difficult (10). The incubation period also varies but usually, it is 1–21 days (11). In case of infection, the infected person's blood test generally has elevated liver enzymes, higher levels of urea and creatinine, low blood glucose levels, and fewer white blood cell counts (2). Most of the infected individuals have pneumonia. The pathogen affects internal organs like the spleen, kidney, and liver (12). Toll-like receptors (TLRs) play an important role in inducing innate immunity. Infected patients showed increased expression of TLRs (13).

Early antimicrobial therapy against the infection proved effective in lowering the mortality rate in endemic regions (11). The pathogen shows resistance to several antibiotics, including cephalosporins, imipenem, meropenem, tetracyclines, sulphonamides, etc. The effectiveness of antimicrobial therapy depends upon the duration of the treatment, which in normal cases is up to 6 months (14). A licensed vaccine for Melioidosis has not been not reported to date, but recent studies confirmed the pre-clinical trials of *B. pseudomallei* vaccine in animal models (15). Several strategies for vaccine development against *Burkholderia* have been explored, but none of them induce sterilizing immunity for a long time. Some promising live attenuated and subunit vaccines against *Burkholderia* have been proposed. Many of the studies revealed short-term immune protection against the pathogen in animal models (15). The experimental vaccine development can be assisted with computational vaccine approaches that use genomic information of bacterial pathogens to predict novel antigenic peptides. The peptide vaccines are easy to produce and much cheaper compared to whole or subunit-based vaccines. This study highlights several vaccine targets against the pathogen using immunoinformatic approaches. A multi-epitope peptide vaccine is constructed along with MALP-2 (macrophage activating lipoprotein 2) adjuvant. Computational analysis of the vaccine highlighted its good ability to induce protective responses against the pathogen. This strategy is considered simple and cost effective and the findings might be useful for multi-epitope peptide vaccine development against *B. pseudomallei.*

The multi-epitopes vaccine (MEVC) construct has many advantages over the single peptide-based vaccine. MEVC elicits both humoral and cellular immune responses because it comprises both B and B-cell derived T-cell epitopes (16, 17). Other advantages of MEVC include probable antigenicity, less toxicity, non-allergenicity, and good water-soluble properties (18). The MEVC in the present study comprises multiple epitopes which were prioritized from vaccine targets. The vaccine proteins were filtered using pan-genome analysis and reverse vaccinology techniques. Extracellular, periplasmic, and outer membrane proteins were shortlisted for epitope mapping. B-cell derived T-cell epitopes were predicted and potential epitopes were used to design a vaccine construct which further underwent different computational analyses. Toll-like receptor 5 (TLR5), shows high expression in Melioidosis (15) and is a receptor for bacterial flagellin. This receptor was used for molecular docking advanced by molecular dynamics simulation and binding free energies.

Methodology

Core proteome retrieval

Completely sequenced 91 proteomes of *B. pseudomallei* were obtained from the National Center for Biotechnology Information (NCBI) (19) followed by the retrieval of the core proteomes of the species by applying the bacterial pan genome analysis tool (BPGA) (19). CD-HIT was then used to extract non-redundant sequences at a 90% threshold. CD-HIT is a cluster database used to analyze the proteomes by removing similar sequences. PSORTB v 3.0 (20) was used for subcellular localization. Subcellular localization refers to the protein localization prediction in *B. pseudomallei*. The outer membrane, extracellular, and periplasmic proteins were chosen for epitope mapping leading to multi-epitope vaccine designing. The protein sequences obtained includes 39 extracellular, 53 periplasmic, and 55 outer membranes which were then subjected to further analysis as presented in Figure 1.

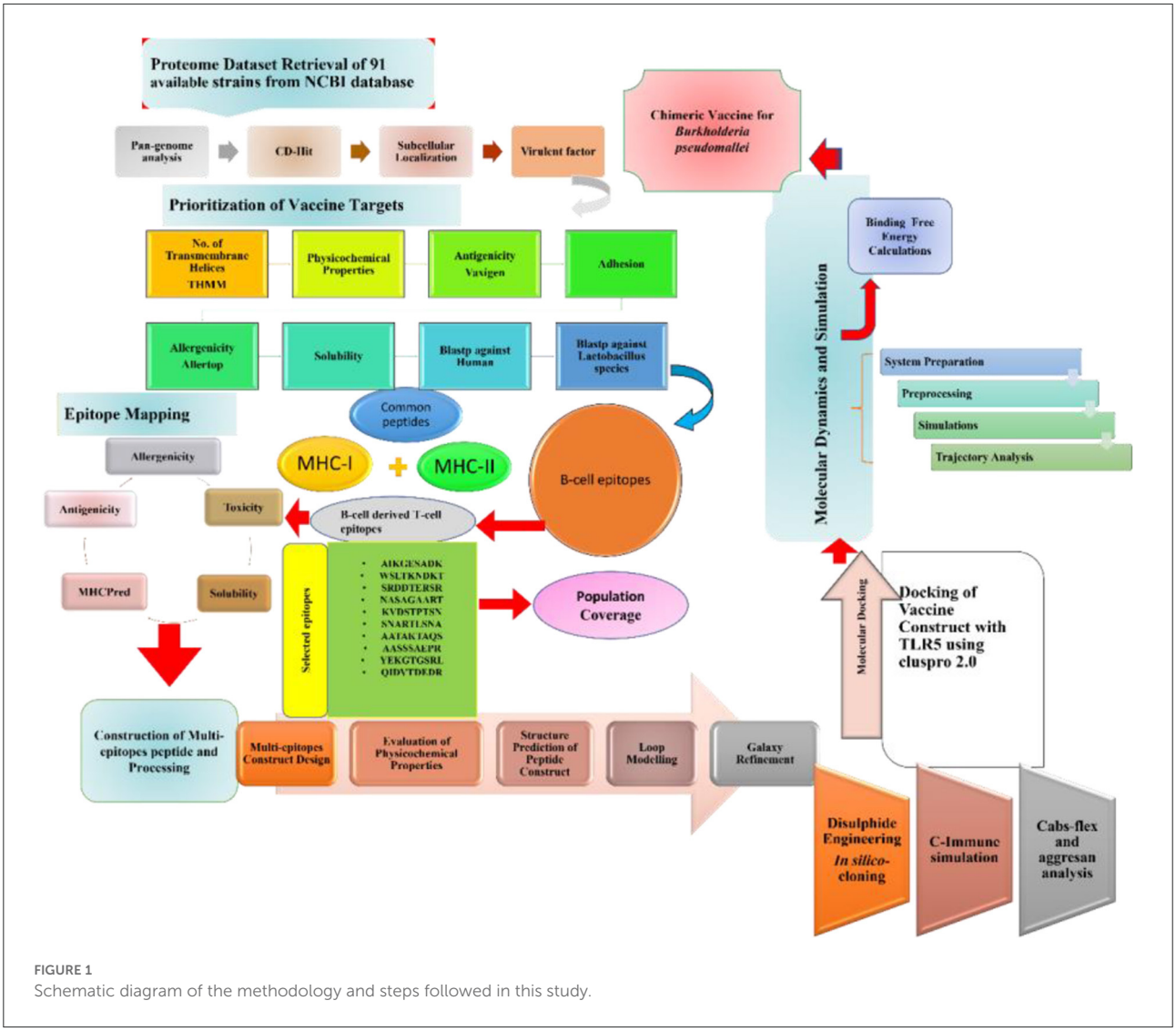


FIGURE 1
Schematic diagram of the methodology and steps followed in this study.

Identification of potential vaccine candidates

All the extracellular, periplasmic, and outer membrane proteins were checked for transmembrane helices using TMHMM 2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0). Physiochemical properties like molecular weight, theoretical index, instability index, and GRAVY (hydrophilicity check) for proteins with transmembrane helices < 2 were checked by using the Protparam tool (https://web.expasy.org/protparam/). The molecular weight of the proteins should be <100 kDa as they can be easily purified from the cell, and their instability index should be <45. Further, antigenicity and allergenicity of the proteins were checked *via* Vaxijen v2.0 (21) and AllerTOP v2.0, respectively (22). The Innovagen server was used to shortlist the soluble proteins. The adhesion probability of the proteins was evaluated using Vaxign webserver (23). Finally, BLASTp was performed on shortlisted proteins against human and lactobacillus species to avoid auto-immune responses and to restrict accidental inhibition of the probiotic bacteria, respectively. The sequence identity and bit score used in this process were $\leq$30 % and 100, respectively.

Epitope prediction and prioritization

All the protein sequences meeting the criteria of being a potential candidate and shortlisted from all the checks were then used for epitope mapping utilizing the Immune Epitope Database and Analysis Resource (IEDB) (24). The proteins subjected to epitope mapping were antigenic, non-allergen, soluble, and adhesion probability > 0.5. B-cell epitope mapping was done using the B-cell epitope prediction tool of IEDB. The predicted B-cell epitopes were then used as input sequences for

MHC-II (major histocompatibility complex II) binding epitopes prediction, and the anticipated MHC-II binding epitopes were used to predict MHC-I binding epitopes. A complete reference set of alleles were selected for T-cell epitope (MHC-I & MHC-II binding) mapping. The B-cell derived T-cell epitopes were further analyzed for their binding affinity by MHCpred v2.0 (http://www.ddg-pharmfac.net/mhcpred/MHCPred/). Epitopes with IC50 <100 nM were considered good binders and further scrutinized for antigenicity, allergenicity, toxicity, and solubility. Only epitopes that were antigens, non-allergen, non-toxic, and soluble were used to design a multi-epitope vaccine construct.

Population coverage analysis

Population coverage analysis of the designed vaccine was evaluated using the Population coverage tool of IEDB. The alleles which cover most of the world population were used for the analysis.

Multi-epitopes vaccine designing and 3D modeling

Shortlisted epitopes were linked together to design a vaccine construct. GPGPG linkers were used to link epitope to epitope while EAAAK linker was used to connect epitopes peptide with an adjuvant. The adjuvant molecule was used to boost vaccine immune responses. Here, we used four different adjuvants resulting in four different constructs which were further analyzed for physiochemical properties, antigenicity, allergenicity, adhesion probability, and secondary structures. Construct with Cholera toxin-B (CTB) adjuvant met the criteria to be considered best amongst the four. MALP-2 (GNNDESNISFKEK) was added to the construct, EAAAK linker was used to link the last epitope and MALP2 sequence. MALP-2 (macrophage activating lipoprotein) is a toll-like receptor (TLR) agonist mostly used to treat certain infections and also as an adjuvant for a vaccine. Physiochemical properties like molecular weight, hydrophilicity, theoretical, and instability index of the final vaccine construct were evaluated by the ProtParam tool (25). We predicted the 3D structure modeling of the vaccine using the Scratch predictor tool (25). The vaccine model was analyzed for the presence of the loops and subjected to loop modeling via Galaxy loop followed by refinement with Galaxy refine v 2.0 tool, respectively (26).

Cabs-flex analysis

The cabs-flex analysis is a computational approach to investigating the structural flexibility of the protein. The multi-epitopes peptide vaccine was subjected to coarse-grind simulations using Cabs-flex 2.0 (27). The analysis was run for 50 cycles and an RNG seed of 8335. The temperature range, weight of global side chain restraint, and the weight of global C-alpha restraints used in the analysis were 1.40, 1.0, and 1.0, respectively. This analysis proceeded by following the identification of aggregation-prone sites in protein sequence (vaccine) by AGGRESCAN 3D server v2.0 (http://bioinf.uab.es/aggrescan/).

Disulfide engineering and *in-silico cloning*

Disulfide by Design 2.0 was used to perform disulfide engineering of the vaccine model. It is usually done to identify unstable residues being mutated into cysteine residues narrating the induction of disulfide bonds making the structure stable (27).

Java Codon Adaptation Tool (JCat) (28) was used for the codon optimization of the vaccine construct. Through JCat, the vaccine protein sequence is translated into nucleotide sequence. Additionally, the CAI (codon adaptation index) value and GC content were determined, enabling *in-silico* cloning *via* Snapgene (29). The *in silico* cloning was done in vector pET-28a (+).

Computational immune simulation

A computational immune simulation is an approach to analyzing the vaccine construct for its immunogenicity and ability to induce immunity. This analysis was done through the C-ImmSim webserver (30). The webserver uses machine learning techniques while predicting the immune cells and antigen interaction.

Molecular docking

Toll-like receptor-5 (TLR-5) and vaccine model (ligand) were subjected to molecular docking using an online docking tool ClusPro 2.0 (31). Several docked solutions were generated and ranked based on binding energy. The solution with the lowest binding energy was chosen for further investigation. The interactions between the receptor and the vaccine were visualized through UCSF Chimera v 1.15, PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and Discovery Studio v2021.

Molecular dynamics simulation

Molecular dynamics simulation was performed for the docked complex of the vaccine model and TLR5 *via* (Assisted

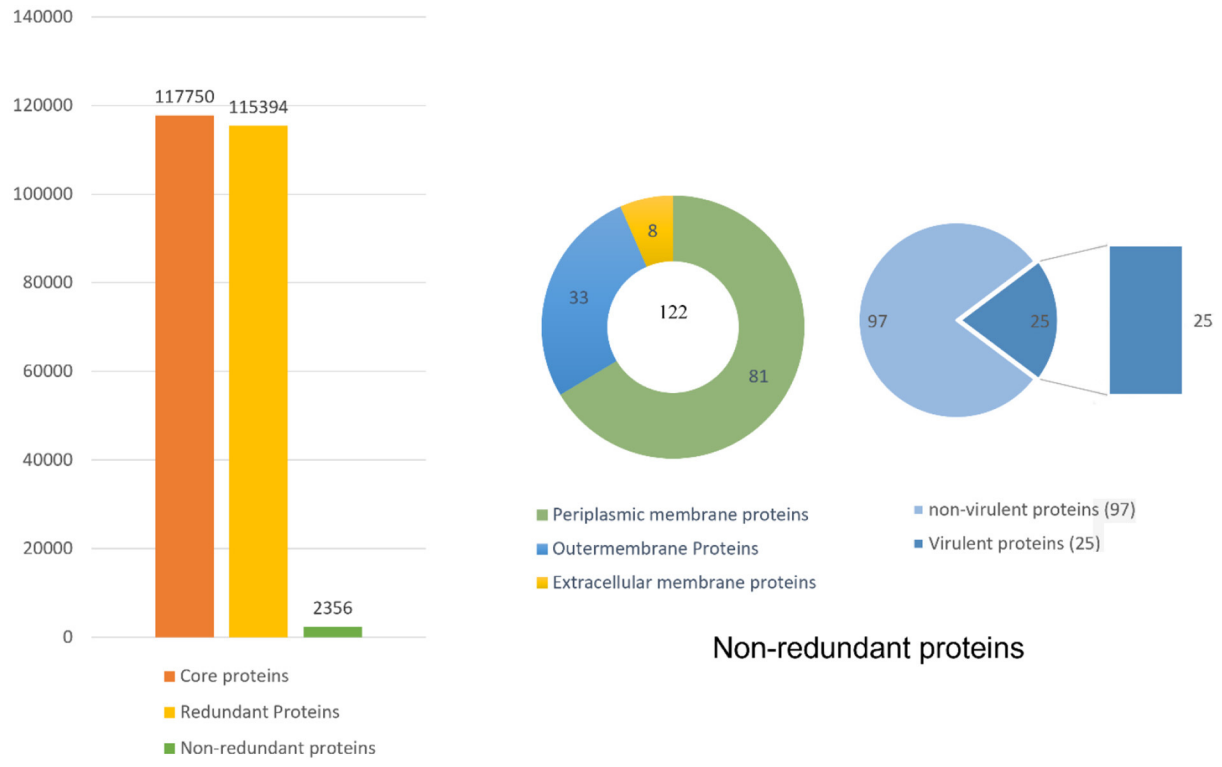


FIGURE 2
Complete detail of the total retrieved and shortlisted proteins.

Model Building with Energy Refinement tool) AMBER 20 software. The simulation was conducted for 50 ns to understand the dynamic behavior in an aqueous solution. The force field "ff14SB" was used to generate parameters for both the vaccine and the TLR5, followed by incorporation of complex into TIP3P water box by maintaining 12 Å padding distance. The Na^+ ions were added to neutralize the system. The carbon alpha atoms, non-heavy atoms, hydrogen atoms, and solvation box energies were minimized for 500, 1,000, 300, and 1,000 steps, respectively. Langevin dynamics were applied to maintain the system's temperature by the execution of system heating to 300 K for 20ps. The SHAKE algorithm was applied to constraint hydrogen bonds. The production run trajectories were analyzed by using the AMBER CPPTRAJ module. Binding free energies were predicted using the MMPBSA.py module (https://ambermd.org/tutorials/advanced/tutorial3/py_script/section4.htm).

Results

Retrieval of core proteome and determining the vaccine candidates

Pan-genome analysis was conducted for 91 completely sequenced proteomes of *B. pseudomallei* to get core proteome (19). The total number of the core proteins obtained were 310,128, which were analyzed through the CD-HIT analysis. The non-redundant sequences obtained were 3,647 while all the duplicate sequences were removed. As only extracellular, periplasmic, and outer membrane proteins were of interest so the subcellular localization of those 3,647 sequences was done by PSORTb. Sequences obtained were 39, 56, and 53 in number for extracellular, periplasmic, and outer membrane proteins, respectively (Figure 2).

The bacterial virulence factor was determined using the virulence factor database (VFDB). The protein sequences with a sequence identity of more than 30% and a bit score >100 were considered the best vaccine targets. Virulent proteins numbered 17 in the extracellular group, 18 in the periplasmic group, and 12 in the outer membrane group. The sequences were further analyzed for the presence of transmembrane helices that should be 0 or 1. The number of transmembrane helices was 0 for extracellular proteins, except for one. In the periplasmic group, only one had transmembrane helices out of 18 proteins while the remaining were lacking the transmembrane helices. Out of the 12 outer membrane proteins, only three (core/2096/1/Org1 Gene676, core/4074/1/Org1 Gene4523, and core/4761/73/Org73 Gene5861) had transmembrane helices, while the rest had 0. In all three groups, the proteins with no transmembrane

helices were subjected to physicochemical evaluation. The parameters evaluated were molecular weight, theoretical index, GRAVY (Grand average of hydropathy), and instability index. Molecular weight was <100 kDa, while the GRAVY calculated was negative and the proteins having an instability index >40 were considered unstable proteins. The instability index of 15 extracellular proteins, 13 periplasmic, and nine outer membrane proteins were stable and within the range of 40. All these stable proteins were subjected to antigenicity (threshold 0.5), allergenicity, solubility, and adhesion stability (threshold 0.5) checks. All the non-antigenic and allergen proteins were discarded including eight extracellular, four periplasmic, and two outer membrane proteins. Similarly, the poorly soluble and the proteins having adhesion probability <0.5 were discarded, and only antigenic, non-allergens, soluble, and having adhesion probability were shortlisted as potential vaccine candidates. Four proteins were filtered out of which three (core/7319/1/Org1_Gene2732, core/7521/1/Org1_Gene303, and core/8978/1/Org1_Gene1539) were from the extracellular group and only 1 (core/3042/1/Org1_Gene5247) was a periplasmic protein as listed in Table 1. The BLASTp against human and lactobacillus species was also run for the shortlisted proteins that reported no significant similarity between the proteins and human/probiotic proteomes. The core/7319/1/Org1_Gene2732 protein is a type VI secretion system involved in the transportation of effector molecules from a bacterial cell to the target cell, core/7521/1/Org1_Gene303 is an Hcp family type VI secretion system effector which is present in the outer membrane and play a role in pathogenicity. The core/8978/1/Org1_Gene1539 is a flagellar biosynthesis anti-sigma factor FlgM which acts as a negative regulator in synthesizing flagellin and core/3042/1/Org1_Gene5247 is D-alanyl-D-alanine endopeptidase (32–34).

TABLE 1 Properties for the potential vaccine candidates.

Gene Id	Name	Function	Length	VFDB	T.H	M.W	T. pI	I. I	GRAVY	Antigenicity	Allergenicity	Solubility	Adhesion
core/7319/1/Org1_Gene2732	Type VI secretion system	Transport the effector molecules from bacterial cell to target	167	46.71	0	18.33	6.43	37.04	−0.472	1.0872	Non-allergen	soluble	0.708
core/7521/1/Org1_Gene303	Hcp family type VI secretion system effector	Play role in pathogenicity	161	64.38	0	17.54	5.57	38.01	−0.501	1.146	Non-allergen	soluble	0.586
core/8978/1/Org1_Gene1539	Flagellar biosynthesis anti-sigma factor FlgM	Negative regulator in synthesizing flagellin	114	60.53	0	11.11	8.19	20.6	−0.239	0.728	Non-allergen	soluble	0.665
core/3042/1/Org1_Gene5247	D-alanyl-D-alanine endopeptidase, putative	Hydrolyzing the peptidoglycan	373	46.42	1	39.91	9.99	35.83	−0.231	0.534	Non-allergen	soluble	0.601

Epitope prediction and prioritization

In epitope mapping, T-cell epitopes were predicted from B-cell epitopes for eliciting both humoral and cellular immune responses. Using the shortlisted proteins as input sequences, the B-cell derived T-cell epitopes were predicted by IEDB. The B-cell epitope prediction tool of IEDB was used to predict potential peptides acting as B-cell epitopes. Peptides having a value ≥0.5 (threshold) were considered as B-cell epitopes. Thirteen peptides were predicted as B-cell epitopes. A total of 78 T-cell MHC class II binding epitopes were predicted from the B-cell epitopes. These epitopes were then used to predict the MHC class I binding epitopes on T-cells, and 120 of these epitopes with lengths of 9 or 10 residues were selected based on their percentile rank. The compete reference set of alleles was selected as a parameter for MHC class II and I binding epitopes prediction.

The predicted epitopes were prioritized using different parameters. First, they were subjected to MHCPred to select

epitopes with IC50 values <100 nM to confirm the epitopes as good binders. With the help of this analysis, 64 epitopes were chosen and further examined by VaxiJen for antigenicity; only 52 epitopes were antigenic, of which 24 were non-allergen. All the epitopes were non-toxins. Five epitopes were predicted by Innovagen to be poorly soluble and were deleted while overlapping ones were also removed. Finally, 10 epitopes listed in Table 2 were chosen for a downward analysis.

Population coverage

Population coverage analysis by IEDB was performed for the shortlisted epitopes to estimate their binding probability to MHC molecules covering the world population (35). World population coverage calculated for MHC class I was 98.55% (Figure 3A) and for MHC class II was 81.81% (Figure 3B). The combined population coverage for MHC molecules was 99.74% as in Figure 3C. MHC class combined world population coverage was also computed region-wise, as shown in Figure 4.

Designing the vaccine constructs and 3D modeling

The 10 shortlisted epitopes that were antigenic, non-allergen, non-toxic, and soluble were considered as potential epitopes to be a part of the vaccine construct. Epitopes were chosen using GPGPG linkers. We designed four constructs using four different adjuvants like TLR-4 agonist, B-defensin, Cholera B toxin (CTB), and 50 s ribosomal adjuvant. The rationale for using different adjuvant molecules was to check which adjuvants show the best compatibility with the designed vaccine molecule and generate strong and protective immune responses. The designed constructs were then evaluated for their physiochemical properties, antigenicity, allergenicity, adhesion probability, and prediction of secondary structure on four parameters (Figure 5). The construct with the adjuvant TLR4 agonist was marked as allergen while the construct with adjuvant 50S ribosomal had an adhesion probability of 0.3 which was less than threshold 0.5 so they both were eliminated. The remaining vaccine constructs one with the adjuvant B-defensin and the other with Cholera B toxin were compared. The vaccine construct with the adjuvant CTB was shortlisted based on its secondary structure being stable. We linked MALP-2 (macrophage activating lipoprotein 2) with construct *via* the EAAAK linker at the end presumed to be involved in enhancing the antigenicity of the construct. The CTB vaccine construct along with MALP 2 was again evaluated for physiochemical properties thus having a molecular weight of 29.01 kDa, theoretical index of 8.92, −0.65 GRAVY, and stable with a 23.28 instability index. It was non-allergen and antigen with antigenicity of 1.0059.

The finalized vaccine construct was modeled into a 3D structure (Figure 6A) using the Scratch predictor and visualized by UCSF chimera which highlights the loop regions in the structure. The structure was loop modeled by the Galaxy loop of the Galaxy web and was refined by the Galaxy refine of the Galaxy web. The top 10 galaxy refined models along with their structural information are shown in Table 3. 2D structure analysis and structural validation were done through PDBSum Generate tool (35). Figure 6B is highlighting the secondary structure elements of the vaccine model having 13 helices, 6 helix-helix, 46 beta turns, and 3 gamma turns. Ramachandran plot as shown in Figure 6C confirmed the presence of 199 residues in the Rama favored region, 14 were in additional allowed and 1 was in disallowed region, hence validating a good model for the vaccine.

Cabs-flex analysis

AGGRESCAN analysis was performed before a cabs-flex analysis (cabs-flex obtained model is shown in Figure 7A). The vaccine model was first subjected to AGGRESCAN for aggregation-prone regions. The residues having scores <0 are considered soluble while a positive value depicts the aggregations-prone residues as in Figure 7B. This step is followed by cabs-flex resulting in 10 models. The highest RSMF obtained was 7.117 angstrom for residue 282, and the lowest was 0.297 angstroms for residue 93 (Figure 7C).

Disulfide engineering and *in-silico* cloning

The presence of disulfide bonds in any model confirms its structural stability. Disulfide engineering refers to marking the residues in the model as being considered unstable and mutating them as cysteine pairs. There were 18 cysteine pairs identified (Figure 8A). Mutated cysteine pairs included 3LYS-36THR (energy value 5.71 kcal/mol, X3 angle +121.45), 6PHE-12VAL (energy value 3.60 kcal/mol, X3 angle +115.46), 16SER-27THR (energy value 1.33 kcal/mol, X3 angle +96.65), 105LYS-108VAL (energy value 3.11 kcal/mol, X3 angle +95.23), 109TRP-112LYS (energy value 2.7 kcal/mol, X3 angle +119.05), 130ALA-140PRO (energy value 4.85 kcal/mol, X3 angle +88.29), 160ASP-163GLU (energy value 2.57 kcal/mol, X3 angle +107.9), 168PRO-182PRO (energy value 4.5 kcal/mol, X3 angle +103.79), 173ALA-178ALA (energy value 1.41 kcal/mol, X3 angle +85.64), 186LYS-192THR (energy value 5.21 kcal/mol, X3 angle +89.68), 195GLY-216THR (energy value 1.92 kcal/mol, X3 angle −107.05), 206SER-256GLN (energy value 3.21 kcal/mol, X3 angle +98.68), 208ALA-215ALA (energy value

TABLE 2 Predicted B cell derived T cell epitopes with various checks.

Protein IDs	Protein names	B cell epitopes	MHC II binding epitopes	MHC I binding epitopes	MHC Pred	IC50 value	Antigenicity	Allergenicity	Toxicity	Solubility
core/7319/1/Org1_Gene2732	Type VI secretion system	PAIKGESADKDHE	AIKGESADKDH	AIKGESADK	AIKGESADK	75.16	1.7824	Non-allergen	Non-toxin	Soluble
		WKQTQQKIGGNQGGNTQGAWSLTKNDKTYA	WSLTKNDKTYA	WSLTKNDKTY	WSLTKNDKT	74.13	0.822	Non-allergen	Non-toxin	Soluble
core/7521/1/Org1_Gene303	Hcp family type VI secretion system effector	RPSGSRDDTERSRE	SGSRDDTERSR	GSRDDTERSR	SRDDTERSR	10.48	1.6539	Non-allergen	Non-toxin	Soluble
core/8978/1/Org1_Gene1539	Flagellar biosynthesis anti-sigma factor FlgM	MKVDSTPTSNARTLSNASAGAARTQAGQPAAAQTPAGAAGAPTGGDANV	LSNASAGAART	NASAGAART	NASAGAART	9.31	1.6869	Non-allergen	Non-toxin	Soluble
		MKVDSTPTSNARTLSNASAGAARTQAGQPAAAQTPAGAAGAPTGGDANV	MKVDSTPTSNA	KVDSTPTSNA	KVDSTPTSN	23.82	1.2218	Non-allergen	Non-toxin	Soluble
		MKVDSTPTSNARTLSNASAGAARTQAGQPAAAQTPAGAAGAPTGGDANV	TSNARTLSNAS	TSNARTLSNA	SNARTLSNA	9.46	0.5532	Non-allergen	Non-toxin	Soluble
core/3042/1/Org1_Gene5247	D-alanyl-D-alanine endopeptidase, putative	VAPADAFAATAKTAQSAKGKKSAAKK	AFAATAKTAQS	AATAKTAQS	AATAKTAQS	7.8	1.153	Non-allergen	Non-toxin	Soluble
		LRAASSSAEPRAKGAR	LRAASSSAEPR	AASSSAEPR	AASSSAEPR	37.85	1.52	Non-allergen	Non-toxin	Soluble
		KSPLTDQIDVTDEDRDYEKGTGSRL	RDYEKGTGSRL	YEKGTGSRL	YEKGTGSRL	4.7	1.9551	Non-allergen	Non-toxin	Soluble
		KSPLTDQIDVTDEDRDYEKGTGSRL	QIDVTDEDRDY	QIDVTDEDR	QIDVTDEDR	5.36	1.0121	Non-allergen	Non-toxin	Soluble

10.3389/fmed.2022.945938

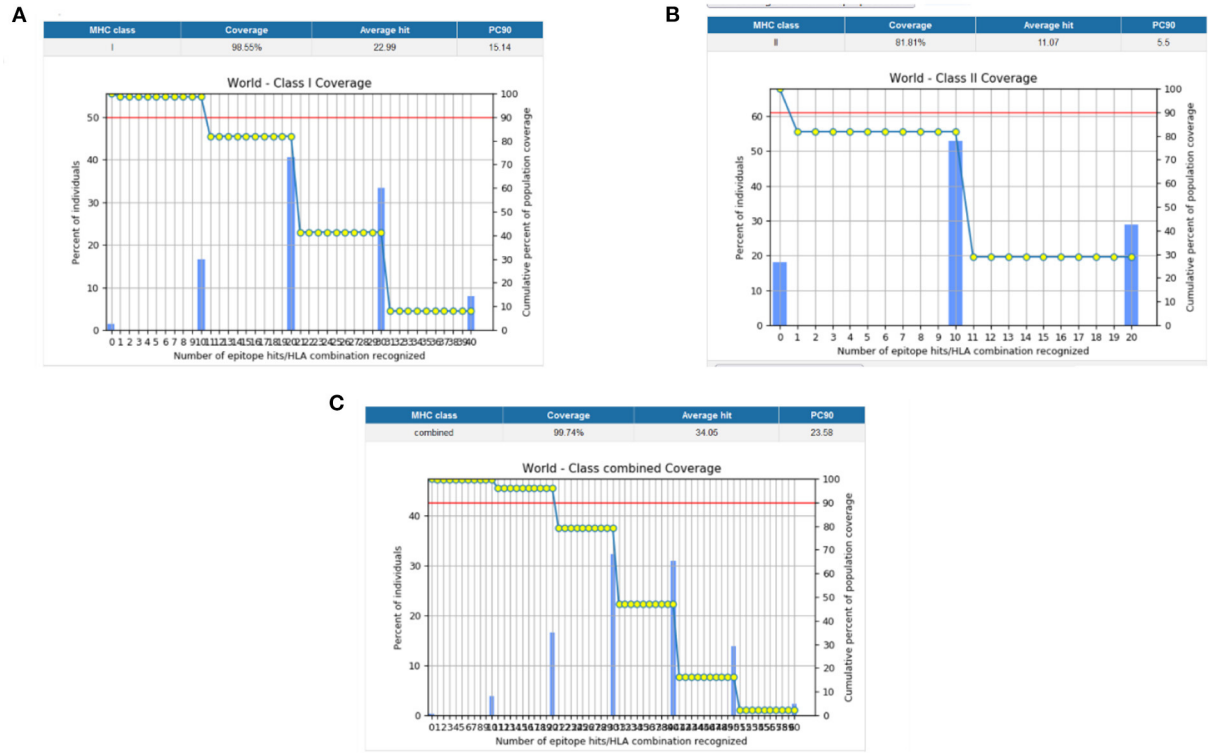


FIGURE 3
(A) MHC class I world population coverage **(B)** MHC class II world population coverage **(C)** MHC class I and II combined world population coverage.

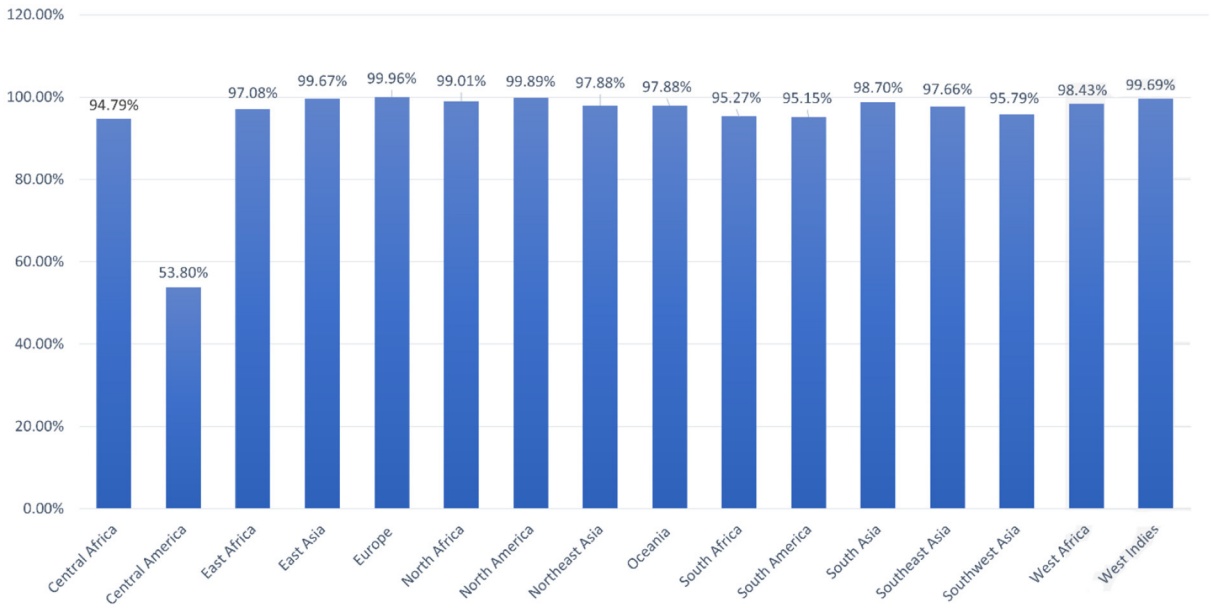


FIGURE 4
World population coverage of the designed vaccine construct.

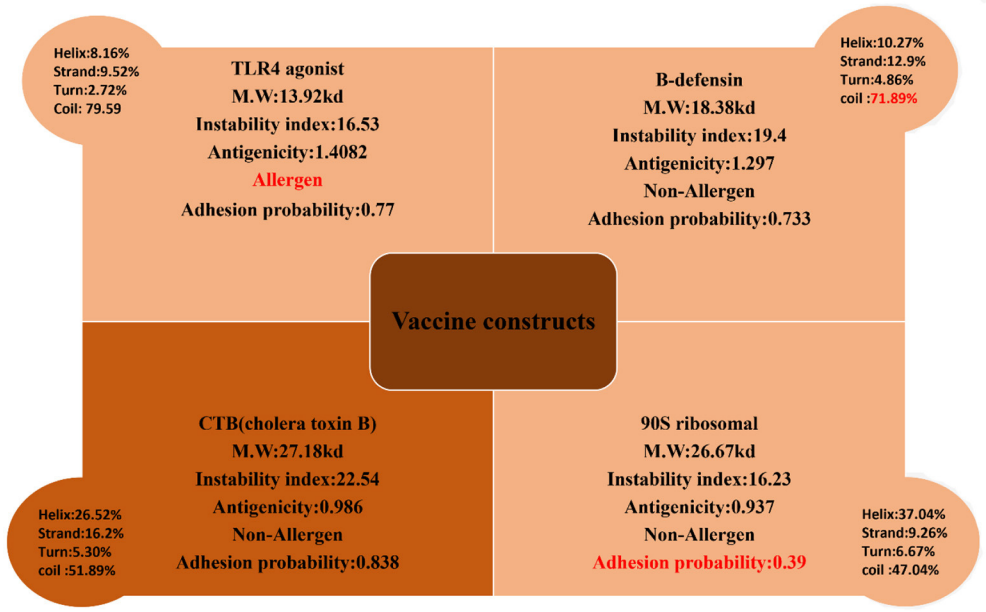


FIGURE 5
Four constructs with different adjuvant were analyzed for their physiochemical properties and secondary structure parameters. One construct with a TLR4 agonist was an allergen, the construct with B-defensin had a higher number of coils depicting the unstable structure, and the construct with 90S ribosomal had a very low adhesion probability; these were discarded. The construct with CTB met the criteria of a potential vaccine.

TABLE 3 Galaxy refined models with their properties.

Model	GDT-HA	RMSD	MolProb	Clash score	Poor rotamers	Rama favored
Initial	1	0	2.992	56.3	1.4	85.4
Model 1	0.9459	0.436	1.832	9.7	0.9	95.4
Model 2	0.9592	0.398	1.914	10.7	0.9	94.6
Model 3	0.9477	0.421	1.86	10.4	0.5	95.4
Model 4	0.9486	0.417	1.877	9.7	0.9	94.6
Model 5	0.9504	0.406	1.883	10.4	0.9	95

6.31 kcal/mol, X3 angle −64.05), 218LYS-227GLY (energy value 2.89 kcal/mol, X3 angle −70.65), 238PRO-243GLU (energy value 2.93 kcal/mol, X3 angle −103.67), 245GLY-255GLY (energy value 5.57 kcal/mol, X3 angle −79.03), 246THR-266PRO (energy value 5.16 kcal/mol, X3 angle +73.56), and 275SER-278SER (energy value 3.07 kcal/mol, X3 angle −91.98).

In-silico cloning for the vaccine construct was also performed following the codon adaptation in which the sequence for the vaccine construct was reverse-translated. The CAI value obtained for the sequence was 1.0 and the calculated GC content was 53.66%. The expression system was *E. coli* K12 and the vaccine sequence was expressed in expression vector pET-28a (+) colored red in Figure 8B. The reverse translated sequence being cloned in the vector was: ATGATCAAACTGAAATTCGGTGTTTTCTTCAC CGTTCTGCTGTCTTCTGCTTACGCTCACGGTACCCCGCA GAACATCACCGACCTGTGCGCTGAATACCACAACACCC AGATCTACACCCTGAACGACAAAATCTTCTCTTACACCG AATCTCTGGCTGGTAAACGTGAAATGGCTATCATCACCT TCAAAAACGGTGCTATCTTCCAGGTTGAAGTTCCGGGT TCTCAGCACATCGACTCTCAGAAAAAAGCTATCGAACG TATGAAAGACACCCTGCGTATCGCTTACCTGACCGAAG CTAAAGTTGAAAAACTGTGCGTTTGGAACAACAAAACC CCGCACGCTATCGCTGCTATCTCTATGGCTAACGAAGCT GCTGCTAAAGCTATCAAAGGTGAATCTGCTGACAAAGG TCCGGGTCCGGGTTGGTCTCTGACCAAAAACGACAAAA

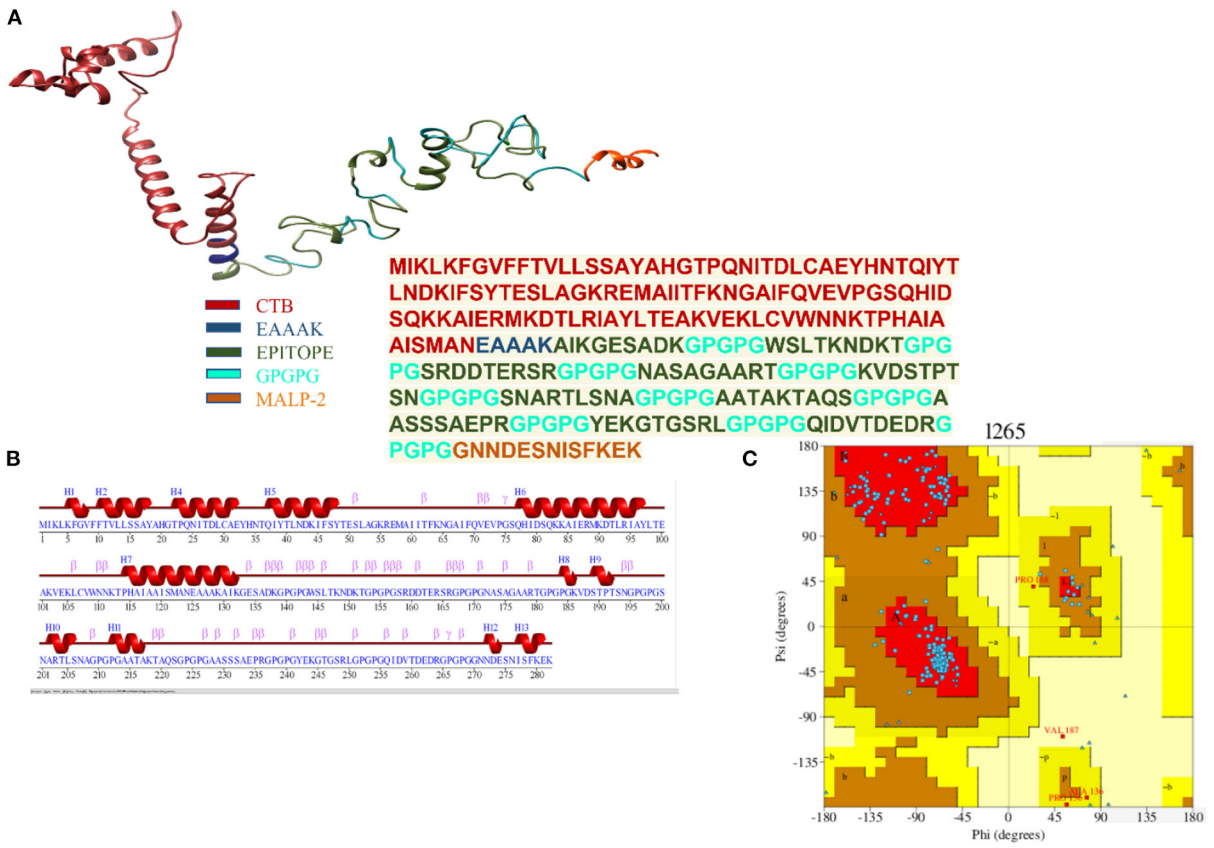


FIGURE 6
(A) Depiction of the details of 3D model of the multi-epitope vaccine with peptide sequence for the vaccine construct. **(B)** Secondary structure for the multi-epitope vaccine. **(C)** Ramachandran plot showing 92.6% of the residue in most favorable region.

CCGGTCCGGGTCCGGGTTCTCGTGACGACACCGAACGT
TCTCGTGGTCCGGGTCCGGGTAACGCTTCTGCTGGTGC
TGCTCGTACCGGTCCGGGTCCGGGTAAAGTTGACTCTA
CCCCGACCTCTAACGGTCCGGGTCCGGGTTCTAACGCT
CGTACCCTGTCTAACGCTGGTCCGGGTCCGGGTGCTGC
TACCGCTAAAACCGCTCAGTCTGGTCCGGGTCCGGGTG
CTGCTTCTTCTTCTGCTGAACCGCGTGGTCCGGGTCCG
GGTTACGAAAAAGGTACCGGTTCTCGTCTGGGTCCGGG
TCCGGGTCAGATCGACGTTACCGACGAAGACCGTGGTC
CGGGTCCGGGTGGTAACAACGACGAATCTAACATCTCT
TTCAAAGAAAAA.

Computational immune simulation

We used a C-ImmSim online webserver for the computational immune simulation demonstrating the immune responses expected against the designed multi-epitopes vaccine. Production of immunoglobulins and interleukins proved the effectiveness of the vaccine. Figure 9A shows that IgM production is 700,000 and IgM + IgG production is 10,000 in the first 15 days. Figure 9B is depicting the interleukins induction in response to the vaccine. The amount of IFN-g generated was >400,000 ng/ml. IFN-g levels are higher, indicating that the human immune system is activated (36).

Molecular docking of vaccine

The designed vaccine was docked with the receptor TLR 5. TLR5 recognizes the flagellin of the bacterium and is highly expressed in patients infected with *B. pseudomallei* (37). TLRs are involved in inducing immunity and cytokine induction. Docking refers to the prediction of interactions between ligand (vaccine) and the receptor (TLR 5) (38). Here, the docked models were obtained from the ClusPro server and the top 10 are shown in Table 4. These models are arranged based on their cluster size. The clustering step was done using the pairwise IRMSD (39). The best docked complex obtained is shown in Figure 10A. The interactions between the vaccine

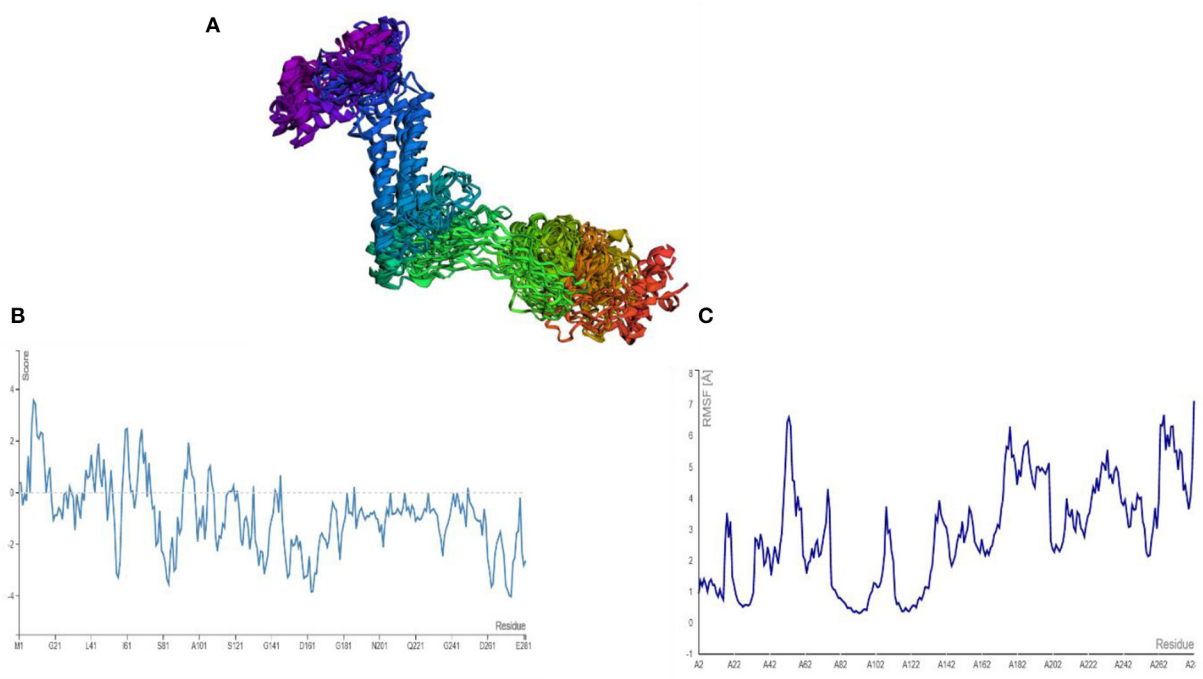


FIGURE 7
(A) Predicted models after a cabs-flex analysis. **(B)** AGGRESCAM plot which depicts that the residues with a score above zero are aggregation prone residues. **(C)** Cabs-flex fluctuations plot, fluctuations calculated as RSMF.

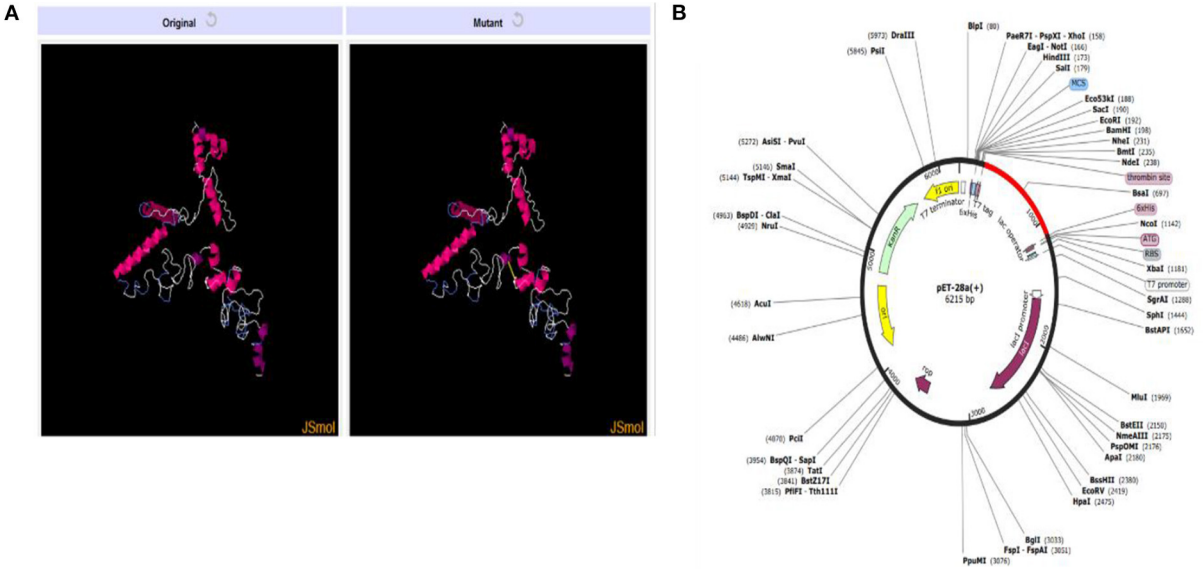


FIGURE 8
(A) Original and mutant vaccine model obtained by disulfide engineering. **(B)** Vaccine (in red) cloned in pET28a (+) vector.

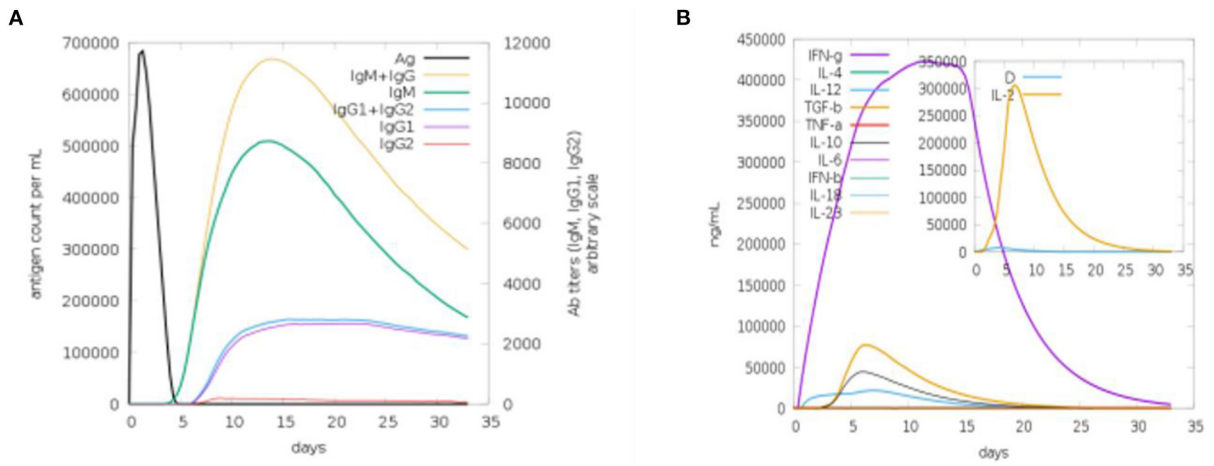


FIGURE 9
Computational immune simulation against the vaccine model. **(A)** Induction of immunoglobulins. **(B)** Induction of interleukins.

(chain A) and TLR-5 (chain B) in a docked complex are presented in Figure 10B. The interactions in red color are salt bridges, while those in yellow are disulfide, and hydrogen bonds are in blue color. The non-bonded interactions are in orange.

Molecular dynamics simulation analysis and binding free energies estimation

The stability and dynamic behavior of the docked Vaccine-TLR-5 complex was further assessed through molecular dynamics simulation analysis. The root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration (RoG) were computed for the system. RMSD analysis during simulation revealed that the complex reaches 7Å at 10.5 ns, however, this is due to the presence of loops. The graph became stable and showed maximum stability toward the end of simulation time as presented in Figure 11A. Next, in RMSF analysis the docked complex was analyzed to check the residue level fluctuations. In RMSF analysis, very low-level fluctuations were observed throughout the simulation time. The few fluctuations were due to the reason for vaccine adjustment at the proper docked site. However, these fluctuations did not affect the overall stability and binding mode of the vaccine to the receptor as shown in Figure 11B. Moreover, the intermolecular stability of the vaccine-TLR-5 can be also witnessed by the radius of gyration analysis, which predicted the maximum compactness of the complex as mentioned in Figure 11C. Additionally, the MM-GBSA analysis was utilized to estimate vaccine-TLR-5 complex free binding energies. The overall binding affinity of the vaccine-TLR-5 complex was a total of −168.35 kcal/mol. The net van der Waals and electrostatic energies are the most favorable in the docked complex formation. The gas phase energy of the complex dominates the overall energy of the system while polar

TABLE 4 Models for the docked complexes of TLR5 and designed vaccine.

Cluster	Members	Representative	Weighted score
0	50	Center	−1,137.2
		Lowest energy	−1,137.2
1	36	Center	−965.5
		Lowest energy	−1,168.3
2	28	Center	−997.5
		Lowest energy	−1,150.1
3	23	Center	−1,127
		Lowest energy	−1,339.4
4	23	Center	−1,032.9
		Lowest energy	−1,133
5	22	Center	−1,216.4
		Lowest energy	−1,234.3
6	19	Center	−1,109.5
		Lowest energy	−1,127.4
7	17	Center	−970.8
		Lowest energy	−1,204.7
8	17	Center	−1,173.4
		Lowest energy	−1,173.4
9	17	Center	−1,034.6
		Lowest energy	−1,139.9

Models are represented in order of their cluster size. Number of members depicting the number of neighboring structures. Weighted score is depicting the energy of the structure present at the center and the neighboring structure with lowest energy in a cluster.

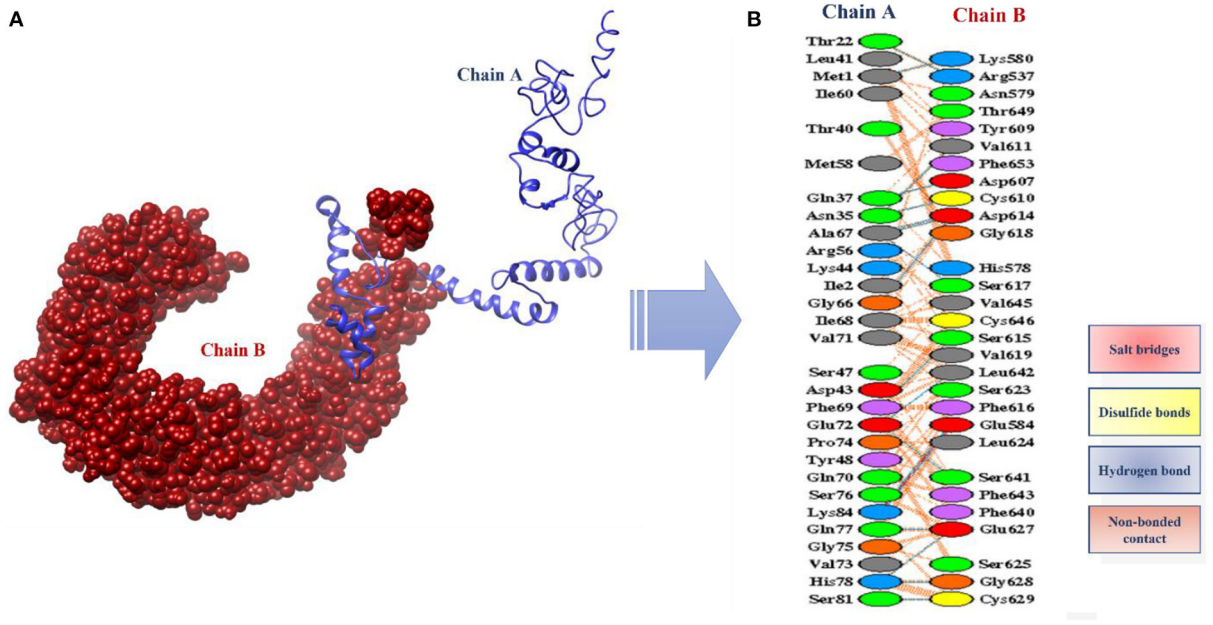


FIGURE 10
(A) Docked complex of vaccine model and TLR5 **(B)** Interactions between the chain A (vaccine) and the chain B (TLR5).

energy is non-favorable in complex formation. The terms for different binding free energies are tabulated in Table 5.

Discussion

B. pseudomallei is a gram-negative bacterium that causes melioidosis and can be lethal if not adequately treated. It can sometimes act as a facultative intracellular pathogen, emphasizing the pathogenesis of the infection (40). Long medications can be used to treat the infection; however, reports have shown that the bacterium has developed resistance to several antibiotics. Many vaccines against this infection were designed and some of them are in preclinical trials as well. Here, we applied cost-effective immunoinformatic and reverse vaccinology approaches to design a multi-epitope vaccine against *B. pseudomallei*. Immunoinformatic approaches for vaccine designing can develop effective vaccines in less time and can provoke both innate and humoral immune responses (40). The previous vaccine development techniques are considered less capable as compared to multi-epitope vaccines (41). Core proteome was retrieved from 91 *B. pseudomallei* strains and subjected to various filters resulting in potential vaccine candidates. This was done *via* a subtractive proteomics approach in which desired proteins were filtered (42, 43). The extracellular, periplasmic, and outer membrane protein sequences were used to design a multi-epitope vaccine. These are the surface proteins involved in bacterial pathogenesis and are very crucial in the development of vaccines (42, 43). These proteins also play a role in pathogen attachment, cell entrance, and disease prognosis (44). The virulence of a protein is very important for disease development (45). These proteins were checked for their capability to be a part of the vaccine and went through various filters like allergenicity, antigenicity, adhesion probability, and solubility. The shortlisted vaccine proteins were used in epitope mapping. B-cell derived T-cell epitopes were used in vaccine construct as B- and T-cells play vital roles in inducing cell-mediated immunity against the pathogen (46). Predicted epitopes were prioritized and shortlisted for vaccine designing. Similar studies were also conducted against *Acinetobacter baumannii* (18), *Enterobacter hormaechei* (16), and *Enterobacter xiangfangensis* (17) etc. These studies prioritized several vaccine targets against the pathogens and proposed chimeric epitope peptides that can better induce humoral and cellular immunity.

The vaccine construct was designed by linking the epitopes with GPGPG linkers and joining the epitope to the adjuvant *via* the EAAAK linker. The purpose of utilizing linkers is to facilitate the separation of epitopes and adjuvants while avoiding overlaps (46, 47). Four different adjuvants were utilized, and the final construct with adjuvant CTB was 3D modeled and subjected to further analysis. CTB is a non-toxic part of Cholera toxin which, when linked to antigen, results in boosting the immunity and immune responses (48). CTB is associated with the induction of CD4+ T-cell immune activation (49). The vaccine model was visualized through Chimera to identify

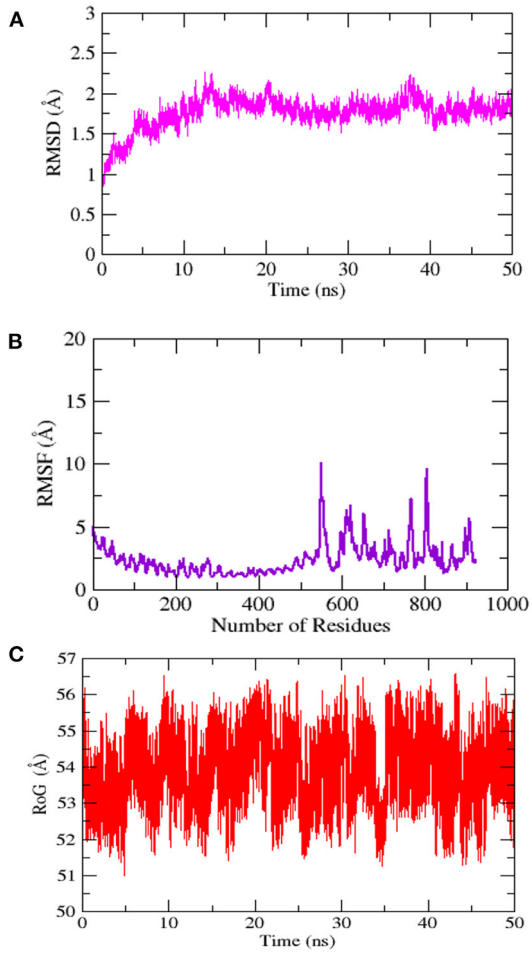


FIGURE 11
Simulation trajectories analysis **(A)** Root-mean-square deviation RMSD **(B)** Root-mean-square fluctuation (RMSF) and **(C)** Radius of gyration (RoG).

TABLE 5 Binding free energies calculated for vaccine-TLR-5 complex.

Energy Component	GB	PB
VDWAALS	−2,327.5146	−2,327.5146
EEL	−24,524.7331	−24,524.7331
EGB	−5,713.6747	−5,548.8917
ESURF	159.6436	110.8892
G gas	−26,852.2478	−26,852.2478
G solv	−5,554.0311	−5,438.0024
TOTAL	−32,406.2788	−32,290.2502
Receptor		
VDWAALS	−2,294.5732	−2,294.5732
EEL	−24,479.7925	−24,479.7925
EGB	−5,727.4367	−5,564.9699
ESURF	159.4314	110.6366
G gas	−26,774.3657	−26,774.3657
G solv	−5,568.0052	−5,454.3333
Vaccine construct		
VDWAALS	−6.8361	−6.8361
EEL	−29.9481	−29.9481
EGB	−11.9397	−12.0813
ESURF	3.1747	2.6863
Gas	−191.0979	−36.7841
Solv	22.7391	−9.3949
TOTAL	−45.5491	−46.1791
Differences (complex-Receptor-Vaccine construct)		
VDWAALS	−126.1053	−126.1053
EEL	−64.9926	−64.9926
EGB	25.7016	28.1595
ESURF	−2.9626	−2.4337
Delta G sol	−191.0979	−191.0979
Delta Solv	22.7391	25.7258
DELTA Total	−168.3588	−216.8237

the loop regions followed by loop modeling and refinement. Loop modeling is involved in predicting secondary structure elements for the loop. Loop regions play a crucial role in several biochemical functions (49, 50). The structure of the vaccine model was improved by refinement because some structures are not predicted accurately, hindering the protein from function properly, and causing difficulty in structure-based research (49, 50). Structural validation of the 3D model was done, and the Ramachandran plot (51) explained the stability of the structure having 92.6% of its residues in the most favored regions. Disulfide engineering and codon optimization were important steps performed to improve structural stability (52). The main aim of the codon optimization was to increase the vaccine's expression level resulting in its higher efficiency for experimental research (18). To generate proper immune responses, the designed vaccine construct should have the binding capability with immune cell receptors. Among immune cells receptors, the TLRs family plays a vital role in the generation of proper immune responses.

In this study, we performed a docking analysis to assess the vaccine's binding potency with TLR-5. The server generates 10 complexes with different binding energy scores. From the docking analysis, the designed vaccine can interact with TLR-5 molecules and can provoke proper immune reactions. For long-term immunity, the docking stability of the docked complexes must be maintained. To evaluate the dynamic movement of the docked complex, a molecular dynamics simulation approach was performed. The results of *in-silico* host immune simulation revealed that the chimeric vaccine can induce a proper immune response in the host body. The immune response was observed in the form of different antibodies and cytokines. According to our findings, the proposed vaccine is promising and capable of eliciting a proper immune response against *B. pseudomallei*.

Conclusions

Antibiotic abuse in animals, humans, and agriculture has resulted in the emergence of antibiotic-resistant bacterial infections, which has significantly raised hospitalization, community mobility, and mortality rates. Melioidosis, a condition that can be fatal, is caused by *B. pseudomallei*. There is no licensed vaccination to protect against *B. pseudomallei* infection. Immunoinformatics, bioinformatics, and reverse vaccinology were employed to speed up vaccine target discovery while also lowering costs and saving time. In the current research, a multi-epitopes-based vaccine was designed that may produce innate and adaptive immunity against *B. pseudomallei* using a variety of bioinformatics, immunoinformatic, and reverse vaccinology methodologies. For the above-said purpose, the first proteins were shortlisted from pathogen core proteins and utilized for probable antigenic, non-allergic, non-toxic, and good water-soluble epitopes prediction. Several analyses demonstrate that the developed vaccine model is capable of inducing a proper innate and adaptive immune response against the target pathogen. However, wet laboratory confirmation and validations are strongly advised to validate *in-silico* findings. Although the *in-silico* results are encouraging, additional efforts are required, such as the use of more accurate servers/tools to validate the results.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: NCBI GenBank.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The researcher would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Choh LC, Ong GH, Vellasamy KM, Kalaiselvam K, Kang WT, Al-Maleki AR, et al. Burkholderia vaccines: are we moving forward? *Front Cell Infect Microbiol.* (2013) 4:5. doi: 10.3389/fcimb.2013.00005

2. Wiersinga WJ, Virk HS, Torres AG, Currie BJ, Peacock SJ, Dance DAB, et al. Melioidosis. *Nat Rev Dis Primers.* (2018) 4:1–22. doi: 10.1038/nrdp.2017.107

3. Kaestli M, Schmid, M, Mayo M, Rothballer M, Harrington G, et al. Out of the Ground: Aerial and Exotic Habitats of the Melioidosis Bacterium Burkholderia Pseudomallei in Grasses in Australia. *Environ Microbiol.* (2012) 14:2058–70. doi: 10.1111/j.1462-2920.2011.02671.x

4. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, et al. Phylogeographic Reconstruction of a Bacterial Species with High Levels of Lateral Gene Transfer. *BMC Biol.* (2009) 7:1–14. doi: 10.1186/1741-7007-7-78

5. Limmathurotsakul D, Golding N, Dance DAB, Messina JP, Pigott DM, Moyes CL, et al. Predicted global distribution of burkholderia pseudomallei and burden of melioidosis. *Nat Microbiol.* (2016) 1:1–5. doi: 10.1038/nmicrobiol.2015.8

6. Limmathurotsakul D, Wongratanacheewin S, Teerawattanasook N, Wongsuvan G, Chaisuksant S, Chetchotisakd P, et al. Increasing incidence of human melioidosis in northeast Thailand. *Am J Trop Med Hyg.* (2010) 82:1113. doi: 10.4269/ajtmh.2010.10-0038

7. Currie BJ, Dance DAB, Cheng AC. The global distribution of burkholderia pseudomallei and melioidosis: an update. *Trans R Soc Trop Med Hyg.* (2008) 102:S1–4. doi: 10.1016/S0035-9203(08)70002-6

8. Currie BJ, Ward L, Cheng AC. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year darwin prospective study. *PLoS Negl Trop Dis.* (2010) 4:e900. doi: 10.1371/journal.pntd.0000900

9. Birnie E, Virk HS, Savelkoel J, Spijker R, Bertherat E, Dance DAB, et al. Global burden of melioidosis in 2015: a systematic review and data synthesis. *Lancet Infect Dis.* (2019) 19:892–902. doi: 10.1016/S1473-3099(19)30157-4

10. Limmathurotsakul D, Peacock SJ. Melioidosis: a clinical overview. *Br Med Bull.* (2011) 99:125–39. doi: 10.1093/bmb/ldr007

11. Cheng AC, Currie BJ. Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev.* (2005) 18:383. doi: 10.1128/CMR.18.2.383-416.2005

12. Wiersinga WJ, Wieland CW, Dessing MC, Chantratita N, Cheng AC, Limmathurotsakul D, et al. Toll-like receptor 2 impairs host defense in gram-negative sepsis caused by burkholderia (Melioidosis). *PLoS Med.* (2007) 4:1268–80. doi: 10.1371/journal.pmed.0040248

13. Hamad MA, Austin CR, Stewart AL, Higgins M, Vázquez-Torres A, Voskuil MI. Adaptation and antibiotic tolerance of anaerobic burkholderia pseudomallei. *Antimicrob Agents Chemother.* (2011) 55:3313–23. doi: 10.1128/AAC.00953-10

14. Morici L, Torres AG, Titball RW. Novel multi-component vaccine approaches for *Burkholderia Pseudomallei*. *Clin Exp Immunol.* (2019) 196:178–88. doi: 10.1111/cei.13286

15. Chantratita N, Tandhavanant S, Myers ND, Chierakul W, Robertson JD, Mahavanakul W, et al. Screen of whole blood responses to flagellin identifies TLR5 variation associated with outcome in melioidosis. *Genes Immunity.* (2013) 15:63–71. doi: 10.1038/gene.2013.60

16. Albekairi TH, Alshammari A, Alharbi M, Alshammary AF, Tahir ul Qamar M, Ullah A, et al. Designing of a novel multi-antigenic epitope-based vaccine against e hormaechei: an intergraded reverse vaccinology and immunoinformatics approach. *Vaccines (Basel).* (2022) 10:665. doi: 10.3390/vaccines10050665

17. Alshammari A, Alharbi M, Alghamdi A, Alharbi SA, Ashfaq UA, Tahir ul Qamar M, et al. Computer-aided multi-epitope vaccine design against enterobacter xiangfangensis. *Int J Environ Res Public Health.* (2022) 19:7723. doi: 10.3390/ijerph19137723

18. ud-din M, Albutti A, Ullah A, Ismail S, Ahmad S, Naz A, et al. Vaccinomics to design a multi-epitopes vaccine for acinetobacter baumannii. *Int J Environ Res Public Health.* (2022) 19:5568. doi: 10.3390/ijerph19095568

19. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep.* (2016) 6:1–10. doi: 10.1038/srep24373

20. Gardy JL, Spencer C, Wang K, Ester M, Tusnády GE, Simon I, et al. PSORT-B: Improving protein subcellular localization prediction for gram-negative bacteria. *Nucleic Acids Res.* (3613) 2003:31. doi: 10.1093/nar/gkg602

21. He Y, Xiang Z, Mobley HLT. Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol.* (2010) 2010:297505. doi: 10.1155/2010/297505

22. Dimitrov I, Flower DR, Doytchinova I. AllerTOP - a server for in silico prediction of allergens. *BMC Bioinform.* (2013) 14:1–9. doi: 10.1186/1471-2105-14-S6-S4

23. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol.* (2017) 8:278. doi: 10.3389/fimmu.2017.00278

24. *ProtParam - SIB Swiss Institute of Bioinformatics | Expasy*. Available online at: https://www.expasy.org/resources/protparam (accessed 24 February, 2022).

25. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH A protein structure and structural feature prediction server. *Nucleic Acids Res.* (2005) 33:W72. doi: 10.1093/nar/gki396

26. Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, et al. CABS-flex 20: a web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res.* (2018) 46:W338–43. doi: 10.1093/nar/gky356

27. Craig DB, Dombkowski AA. Disulfide by design 20: a web-based tool for disulfide engineering in proteins. *BMC Bioinform.* (2013) 14:1–7. doi: 10.1186/1471-2105-14-346

28. 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. doi: 10.1093/nar/gki376

29. *SnapGene | Software for Everyday Molecular Biology*. Available online at: https://www.snapgene.com/ (accessed 24 February 2022).

30. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE.* (2010) 5:e9862. doi: 10.1371/journal.pone.0009862

31. *ClusPro 2.0: Protein-Protein Docking*. Available online at: https://cluspro.bu.edu/login.php?redir=/queue.php (accessed 24 February 2022).

32. Zhou Y, Tao J, Yu H, Ni J, Zeng L, Teng Q, et al. Hcp family proteins secreted via the type VI secretion system coordinately regulate escherichia coli K1 interaction with human brain microvascular endothelial cells. *Infect Immun.* (2012) 80:1243. doi: 10.1128/IAI.05994-11

33. Wang J, Yang Y, Cao Z, Li Z, Zhao H, Zhou Y. The role of semidisorder in temperature adaptation of bacterial FlgM proteins. *Biophys J.* (2013) 105:2598–605. doi: 10.1016/j.bpj.2013.10.026

34. Hung WC, Jane WN, Wong HC. Association of a D-Alanyl-D-Alanine carboxypeptidase gene with the formation of aberrantly shaped cells during the induction of viable but nonculturable vibrio parahaemolyticus. *Appl Environ Microbiol.* (2013) 79:7305–12. doi: 10.1128/AEM.01723-13

35. Laskowski RA, Jabłońska J, Pravda L, Vareková RS, Thornton JM. PDBsum: structural summaries of PDB entries. *Protein Sci.* (2018) 27:129. doi: 10.1002/pro.3289

36. Birnie E, Weehuizen TAF, Lankelma JM, de Jong HK, Koh GCKW, van Lieshout MHP, et al. Role of Toll-like Receptor 5 (TLR5) in experimental melioidosis. *Infect Immun.* (2019) 87:e00409–18. doi: 10.1128/IAI.00409-18

37. Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. Performance and its limits in rigid body protein-protein docking. *Structure.* (2020) 28:1071. doi: 10.1016/j.str.2020.06.006

38. 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. doi: 10.1038/nprot.2016.169

39. Harley VS, Dance DAB, Drasar BS, Tovey G. Effects of burkholderia pseudomallei and other burkholderia species on eukaryotic cells in tissue culture. *Microbios.* (1998) 96:71–93.

40. Tahir Ul, Qamar M, Ismail S, Ahmad S, Mirza MU, Abbasi SW, et al. Development of a Novel multi-epitope vaccine against crimean-congo hemorrhagic fever virus: an integrated reverse vaccinology, vaccine informatics and biophysics approach. *Front Immunol.* (2021) 12:1. doi: 10.3389/fimmu.2021.669812

41. Ismail S, Shahid F, Khan A, Bhatti S, Ahmad S, Naz A, et al. Pan-vaccinomics approach towards a universal vaccine candidate against WHO priority pathogens to address growing global antibiotic resistance. *Comput Biol Med.* (2021) 136:104705. doi: 10.1016/j.compbiomed.2021.104705

42. Moxon R, Reche PA, Rappuoli R. Editorial: reverse vaccinology. *Front Immunol.* (2019) 10:2776. doi: 10.3389/fimmu.2019.02776

43. Grandi G. Bacterial surface proteins and vaccines. *F1000 Biol Rep.* (2010) 2:36. doi: 10.3410/B2-36

44. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of putative vaccine candidates against helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol.* (2015) 32:280–91. doi: 10.1016/j.meegid.2015.03.027

45. Bonilla FA, Oettgen HC. Adaptive immunity. *J Allergy Clin Immunol.* (2010) 125:S33–40. doi: 10.1016/j.jaci.2009.09.017

46. *Multiepitope Subunit Vaccine Design against COVID-19 Based on the Spike Protein of SARS-CoV-2: An in silico analysis* Available online at: https://www.hindawi.com/journals/jir/2020/8893483/#discussion (accessed 27 April 2022).

47. Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo CA. Novel design of a multi-antigenic, multistage and multi-epitope vaccine against helicobacter pylori: an in silico approach infection. *Genet Evol.* (2017) 49:309–17. doi: 10.1016/j.meegid.2017.02.007

48. Stratmann T. Cholera toxin subunit b as adjuvant—-an accelerator in protective immunity and a break in autoimmunity. *Vaccines.* (2015) 3:579–96. doi: 10.3390/vaccines3030579

49. Sun JB, Czerkinsky C, Holmgren J. Mucosally induced immunological tolerance, regulatory T cells and the adjuvant effect by cholera toxin B subunit. *Scand J Immunol.* (2010) 71:1–11. doi: 10.1111/j.1365-3083.2009.02321.x

50. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* (2013) 41:W384. doi: 10.1093/nar/gkt458

51. *The Ramachandran Plot and Protein Structure Validation | Biomolecular Forms and Functions*. Available online at: https://www.worldscientific.com/doi/abs/10.1142/9789814449144_0005 (accessed 27 April 2022).

52. Gupta S, Joshi L. *Codon Optimization. Computational Bioscience*. Arizona State University (2003).

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fmed.2022.1008527

OPEN ACCESS

EDITED BY
Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY
Sudha Chaturvedi,
Wadsworth Center, United States
Hamid Morovati,
Shiraz University of Medical Sciences, Iran

*CORRESPONDENCE
Mugdha Srivastava
mugdha.srivastava@uni-wuerzburg.de
Thomas Dandekar
dandekar@
biozentrum.uni-wuerzburg.de

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Infectious Diseases – Surveillance, Prevention, and Treatment, a section of the journal Frontiers in Medicine

RECEIVED 31 July 2022
ACCEPTED 17 October 2022
PUBLISHED 03 November 2022

CITATION
Gupta SK, Osmanoglu Ö, Minocha R, Bandi SR, Bencurova E, Srivastava M and Dandekar T (2022) Genome-wide scan for potential CD4+ T-cell vaccine candidates in *Candida auris* by exploiting reverse vaccinology and evolutionary information. *Front. Med.* 9:1008527. doi: 10.3389/fmed.2022.1008527

Genome-wide scan for potential CD4+ T-cell vaccine candidates in *Candida auris* by exploiting reverse vaccinology and evolutionary information

Shishir K. Gupta[1,2†], Özge Osmanoglu[1†], Rashmi Minocha[3†], Sourish Reddy Bandi[1,4], Elena Bencurova[1], Mugdha Srivastava[1,5*] and Thomas Dandekar[1,6*]

[1]Department of Bioinformatics, Biocenter, Functional Genomics and Systems Biology Group, University of Würzburg, Würzburg, Germany, [2]Evolutionary Genomics Group, Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany, [3]Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India, [4]Institute of Experimental Biomedicine, University Hospital Würzburg, Würzburg, Germany, [5]Core Unit Systems Medicine, University of Würzburg, Würzburg, Germany, [6]BioComputing Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Candida auris is a globally emerging fungal pathogen responsible for causing nosocomial outbreaks in healthcare associated settings. It is known to cause infection in all age groups and exhibits multi-drug resistance with high potential for horizontal transmission. Because of this reason combined with limited therapeutic choices available, *C. auris* infection has been acknowledged as a potential risk for causing a future pandemic, and thus seeking a promising strategy for its treatment is imperative. Here, we combined evolutionary information with reverse vaccinology approach to identify novel epitopes for vaccine design that could elicit CD4+ T-cell responses against *C. auris*. To this end, we extensively scanned the family of proteins encoded by *C. auris* genome. In addition, a pathogen may acquire substitutions in epitopes over a period of time which could cause its escape from the immune response thus rendering the vaccine ineffective. To lower this possibility in our design, we eliminated all rapidly evolving genes of *C. auris* with positive selection. We further employed highly conserved regions of multiple *C. auris* strains and identified two immunogenic and antigenic T-cell epitopes that could generate the most effective immune response against *C. auris*. The antigenicity scores of our predicted vaccine candidates were calculated as 0.85 and 1.88 where 0.5 is the threshold for prediction of fungal antigenic sequences. Based on our results, we conclude that our vaccine candidates have the potential to be successfully employed for the treatment of *C. auris* infection. However, *in vivo* experiments are imperative to further demonstrate the efficacy of our design.

KEYWORDS

***Candida auris*, T-cell epitope, epitope prediction, positive selection, evolution, immune-informatics**

Introduction

In recent years, life-threatening fungal diseases have increased, and new infections have emerged. Selection pressure of climate change has significantly contributed to the emergence of *Candida auris* as a pathogen (1). *C. auris,* a multidrug-resistant ascomycete, was first isolated in 2009 in Japan. Up to date, it has been detected in 32 countries over six continents (2). *C. auris* infects various tissues and organs, including central nervous system (3), cardiovascular system (4), respiratory tract (5), bones and joints (6), and possesses the potential to cause nosocomial infections (7). The diagnosis of *C. auris* is difficult due to non-availability of specific laboratory techniques which could rapidly and accurately detect it in clinical samples. Furthermore, treatment of infections caused by *C. auris* poses a real challenge due to its high multidrug-resistivity pattern. The reduced susceptibility to conventional antifungal drugs such as azoles and amphotericin B leads to high mortality rate, rising up to 60% (8, 9). Furthermore, *C. auris* can persist on abiotic surfaces such as healthcare instruments for several weeks, which can facilitate prolonged pathogen persistence and high transmissibility (10). This also explains why transmission of *C. auris* to hosts frequently occurs in hospitals. Its persistence in the environment is a trait that differentiates it from other *Candida* species.

Development of anti-fungal vaccines which can trigger the host immune response to generate immunological memory against fungi and their spores remains challenging. Nevertheless, three approaches of immunization are commonly used in development of anti-fungal vaccines; (i) vaccination with live-attenuated strains, which can be hazardous due to possible disease development in immunocompromised patients (11), (ii) immunization with recombinant proteins containing the immunogenic sequences (epitopes), or (iii) immunization with polysaccharides, which are present only in fungal cell wall (12).

To date, there is no human vaccine approved against *C. auris* infections, however, several research groups have made significant progress in its development. Currently, most studies are focused on the Als3, a member of agglutinin-like sequence (Als) family of proteins with adhesive and invasive properties (13). Researchers have proposed Als3 as a promising therapeutic target for *Candida* albicans vaccine development (14–17). In a recent study, Singh et al. (18) identified three adhesin/invasin proteins in *C. auris*, that shared sequence and structural homology to Als3 protein of *C. albicans*. The NDV-3A vaccine proposed by Singh et al. (18) which was based on the N-terminus of Asl3-protein sequence significantly blocked biofilm production ability of *C. auris in vitro*. These authors also found that when combined with the antifungal drug micafungin, the NDV-3A vaccine augmented the protective efficacy of this drug against *C. auris* infection in neutropenia mice. It further induced regulatory CD4+ T helper (TH) cells in infected mice, which comprised of Th1, Th2, and Th17 subcellular population (18). In another study, immunizing mice with rAls3p-N vaccine induced T-cell mediated protection in *C. albicans,* which further signifies the crucial role of CD4+ T lymphocytes as well as associated cytokines such as IFN-γ in acquired immune response against the pathogen (19, 20).

The objective of the current study was to identify novel putative CD4+ T-cell epitopes with vaccine potential against *C. auris* infection by harnessing the evolutionary information combined with reverse vaccinology approach. To this end, we extensively scanned the family of proteins coded by *C. auris* genome and identified the potential vaccine candidates. In search of epitope-based vaccine candidates, one could expect that the pathogen may acquire substitutions in epitopes targeted by immune memory over the years. For instance, hemagglutinin protein of influenza A evolves under strong selection from antibodies (21). To avoid such possibilities, we computed non-synonymous and synonymous distances and then tested for sites with statistical evidence where the accumulation of non-synonymous substitutions exceeds that of the synonymous substitutions. We thus discarded all the rapidly evolved genes of *C. auris* with positively selected sites. We further used the highly conserved regions of protein sequence alignment of multiple *C. auris* strains, so as to identify major candidates for designing the new potential vaccine. This subsequently decreased the possibility of epitopes to rapidly evolve and escape the immune recognition. Moreover, using the stringent *in silico* analysis, we finally identified two conserved immunogenic and antigenic CD4+ T-cell epitopes that could be used for efficient immune memory generation against *C. auris*.

Materials and methods

Genome sequences and quality control

The protein coding genes and translated transcriptomes of five *C. auris* strains (*C. auris* 6,684 from India, *C. auris* B8441 from Pakistan, *C. auris* B11220 from Japan, *C. auris* B11221 from South Africa, and *C. auris* B11243 from Venezuela) were retrieved from NCBI GenBank release 231.0 (22). The annotation completeness of the genome assemblies was accessed with BUSCO v3 (23) using Fungi odb9 database that contains single-copy orthologs (SCOs) selected from OrthoDB v9 (24, 25). **Supplementary Table 1** lists the version and details of the genomic assembly.

Identification of orthologs

Clustering of orthologous genes was performed using Orthofinder v2.3.3 (26) with Diamond v0.9.24.125 (27) (under default settings) based on translated proteomes of selected

five strains. The orthologous clusters that did not contain orthologous genes from all the selected five strains were discarded to only keep genes conserved in all strains. Among the genes in each conserved orthologous clusters, genes that belong to *C. auris* B11220 (4,860 genes in total) were used as reference for further analysis due to higher quality and annotation completeness of its assembly.

Filtering of candidate genes

Conserved *C. auris* B11220 genes were filtered to eliminate the unlikely vaccine candidates. **Table 1** list the tools used for filtering. The first round of elimination was performed based on their secretion or cellular localization. Secretion of the proteins was predicted by checking whether they have a signal peptide or a glycosylphosphatidylinositol (GPI)-anchor or a transmembrane (TM) domain. Signal peptides were identified by using TargetP (28), SignalP (29), Phobius (30), and FunsecKB2 (31). GPI-anchor were identified using PredGPI (32). To identify TM domain, TMHMM (33) and Phobius (30) were used.

Next, proteins with any of the features predicting their sorting *via* classical secretion pathway were further filtered based on the presence of endoplasmic reticulum (ER) retention signal that may restrict their secretion to extracellular space or cell membrane. This step of filtering through the detection of PROSITE pattern PS00014 was performed by ScanProsite (34). Last step of the first round of elimination was done by sorting the proteins by the presence of a signal peptide or a TM domain or a GPI anchor. Then the cellular localization of the proteins was assessed by Deeploc (35). Accordingly, the proteins with a signal peptide were filtered further by their cellular localization being cell membrane, while the proteins with TM domain or GPI anchor were reduced to the ones with extracellular localization.

Second round of filtering was based on fungal effectors and adhesins which possess a higher probability of being a vaccine target (36). The filtering was performed using prediction tools EffectorP (37), FungalRV (38), and FaaPred (39) on both sets of proteins that were subject to classical secretion pathway either to the cell membrane or extracellular space.

Positive selection analysis

Proteins identified from the previous step were aligned with their orthologs from other *C. auris* strains under study. T-Coffee (40) was used for the multiple sequence alignment (MSA) which combines the output of different aligners to enhance the MSA accuracy. The aligned amino acid sequences together with the corresponding nucleotide sequences of each ortholog group were converted into nucleotide alignments at the codon level using the program PAL2NAL (41). Since removal of unreliable regions increases the power to detect positive selection, we used stringent Gblocks filtering (type = codons; minimum length of a block = 4; no gaps allowed) to remove gap-rich regions from the alignments (42). We used site model implemented in codeml program from the PAML 4.2b package (43) to detect the sites in the alignment under positive selection. The Bayes empirical Bayes approach was employed to estimate the probabilities of positive selection for specific codons under the likelihood framework (44). FEL (45) and MEME (46) were used to evaluate for positive selection among filtered genes. The cut off parameter was set to $p < 0.01$. If the positive selection was detected by any of mentioned three programs, the genes were removed for the subsequent analysis.

Highly conserved regions of ortholog clusters

Conservation of amino acid residues in the aligned pathogen sequences was estimated by Shannon entropy function (47) using the Protein Variability Server.[1] We selected the Shannon entropy variability threshold of 1.0 to extract the highly conserved consensus subsequences of length >9 mers. The Shannon entropy (S) for every position in the sequence alignment was calculated as

$$S = -\sum_{i=0}^{N} Fi \log_2 Fi$$

where *Fi* is the fraction of residues of amino acid type *i*, and *N* is the number of amino acid types (48).

CD4+ T-cell binding peptide prediction

The peptides that trigger major histocompatibility complex (MHC) class-II responses are often longer than class-I peptides. These MHC binders are <13–15 amino acids longer with a core sequence of about 9 mers, usually three anchor residues, and their ends extend beyond the peptide-binding groove (49, 50). The MHC class-II binding 13–15 mer peptides for alleles were predicted using Immune Epitope Database (IEDB) (51) recommended prediction method and on full HLA reference set with percentile rank.

Host similarities and antigenicity of MHC class-II binders

To assess the similarity of the predicted MHC class-II binders with human sequences, Blast similarity search

1 http://imed.med.ucm.es/PVS

TABLE 1 Web-servers and databases used for sequence filtering.

Name	Description	Method	References
DeepLoc 2.0	Eukaryotic protein subcellular localization	Machine learning-deep neural networks	97
EffectorP 2.0	Fungal effector proteins	Machine learning	37
PredGPI	GPI-anchor prediction	Support vector machine (SVM) and hidden markov model (HMM)	32
TargetP	Eukaryotic protein subcellular localization	Neural network	98
FungalRV	Fungal adhesin prediction	SVM	38
Phobius	Transmembrane topology and signal peptide prediction	Hidden markov model (HMM	99
SignalP 6.0	Signal peptide prediction	Deep neural network	100
TMHMM 2.0	Transmembrane helices prediction	Hidden markov model	33
FAAPred	Fungal adhesins and adhesin-like protein prediction	Support vector machine (SVM)	39
FunSecKB2	A fungal protein subcellular location knowledgebase		101
ScanProsite	PROSITE signature detection		34

(52) was performed. The Blast parameters were tailored appropriately for this analysis, in accordance with the short size of the peptides (such as matrix PAM30; word-size 2). We first screen the binders against human assembly from ENSEMBL (assembly version GRCh38.p13) and then against NCBI non-redundant human sequence database. Highly similar sequences were discarded from further evaluation. We further performed Blast analysis against a customized sequence database of experimentally verified autoimmune class-II epitopes downloaded from IEDB database (51) to avoid the induction of potential autoimmune reactions. Only the sequence with less than 35% sequence identity over 80% query coverage were evaluated for the antigenicity using antigenicity prediction server[2] and Vaxijen server (53). VaxiJen is based on auto cross covariance (ACC) transformation (53) of protein sequences into uniform vectors of principal amino acid properties, with a threshold value 0.5 for prediction of fungal antigenic sequences. Only consensus predictions from the antigenicity prediction server were considered.

Population coverage

Before population coverage analysis the filtered epitopes binding affinity for HLA class II alleles was accessed using MHC-II prediction tool at IEDB using the neural align method 2.3 (NetMHCII 2.3) and epitope length 11–15 amino acids. Resulting alleles were sorted according to their IC50 (inhibition concentration 50) values and only alleles with IC50 equal to or less than 500 nM were selected for further analysis. For the calculation of the population coverage, selected alleles were examined by IEDB Population coverage tool. This analysis was performed for 16 different geographical regions and global population.

2 http://imed.med.ucm.es/Tools/antigenic.pl

Results

With the aim to identify novel putative CD4+ T-cell epitopes against *C. auris* infection, a pipeline was developed including several steps based on orthology, cellular localization, positive selection, MSAs, T-cell binding potential, antigenicity, and population coverage.

Candida auris B11220 as representative of five strains

Genome assemblies of *C. auris* strains originating from different geographical locations were used for the first step of assessing possible common genes between different strains to be adopted as potential vaccine targets. All the used strains had almost similar sizes in the range of 12.1–12.7 Mb with GC-content of ∼45%. These genomes were originally assembled at scaffold level with scaffold N50 ranging between 60 and 2400 $\times$ 10^3 bp (**Supplementary Table 1**). Next, the Benchmarking Universal Single-Copy Orthologs (BUSCO) (23) software was used for assessment of the quality and completeness of assembly and annotation of the used *C. auris* strains. Employing hidden Markov model (HMM) (54) profiles for a match between given proteome and the OrthoDBv9 (24, 25) SCO dataset for fungi, BUSCO determines the fraction of complete, fragmented, and missing genes for each genome. In other words, it assesses the quality of the genomic data based on expected gene content, which is determined/guided by the fungal SCO dataset. As mentioned above, low quality caused by technical issues in genome assemblies can be detected by metrics like N50 for contigs or scaffolds. However, quality of the genome resource that can be affected by, for instance, contamination, is also influential for the comparative downstream analyses. Therefore, to detect any quality issues with the genomic resource, BUSCO uses an expected gene set of SCO as marker genes for genome/assembly completeness (55).

Genes that were marked as complete were further classified in single-copy and duplicated and quantification of all fractions was thereby assessed by BUSCO. The *C. auris* 6,684 strain assembly showed the lowest completeness level with 33 missing genes and five fragmented genes, whereas the highest quality of assembly and annotation was observed for *C. auris* strains B11220 and B11243 with 285 out of 290 complete genes (**Figure 1**). We identified a total number of 5,331 orthologous groups (OG) among five strains of *C. auris*. In our analysis, 4,752 out of 5,331 OGs were found to encompass all five strains and 4,616 OG were found to be single copy orthologs clusters. B12221 strain of *C. auris* was determined to have the highest number of duplication events with 49 duplications and the highest number of genes with 118 genes to have more than one copy. *C. auris* 6,684 strain was found to be missing in 405 OG, missing in highest number of OG among five strains.

Based on our orthology analysis, assembly quality, and genome completeness, *C. auris* strain B11220 was selected for further analysis and steps of filtering of orthologous gene clusters were followed as shown in **Figure 2**.

Filtering pipeline identifies candidate genes with possible extracellular or membrane localization

First, filtering was performed for identification of genes with possible signaling routes by controlling following features: the existence of a signal peptide with TargetP (28), SignalP (29), Phobius (30), and FunsecKB2 (31), the existence of a GPI anchor with PredGPI (32), and the existence of a TM domain with TMHMM (33) and Phobius (30). Elimination of genes with no probability of being signaled into or through membranes led to a reduction to 2,381 genes that were further eliminated in case of the presence of ER retention signal. Further classification of these signaling proteins was performed through assessment of location by testing the presence of a TM domain or a GPI-anchor (30, 32, 33). Almost half (1,180) of 2,366 possibly signaled proteins were found to possess TM domains or a GPI-anchor. These 1,180 proteins detected here were then further filtered by their location determined by DeepLoc (35). A total of 280 proteins were detected with no cell membrane, and possibly secreted *via* the classical secretion pathway. On the other hand, the other half without TM domains or a GPI-anchor (1,186 proteins) were filtered down to 91 proteins by their location being extracellular. Next step identified 23 proteins for each class as effector or adhesion proteins that may have a higher probability to be a good vaccine target. Out of 23 membrane effector or adhesion proteins, 4 were filtered due to high similarity to human proteins while for extracellular effector or adhesion proteins only 3 were eliminated in this step. This last step ended in 19 membrane and 20 extracellular proteins with low or no similarity to human proteins that reduces the risk of cross-reactivity with host's self-proteins.

Next, the 39 proteins clustered in 38 orthogroups were checked for signs of positive selection. Out of 38 gene alignments, 5 were identified by our pipeline with sites under positive selection (**Supplementary Dataset 1**). PAML (43) identified sites under selection in orthogroups OG0003318 and OG0003714. Selection in OG0000005 was identified by both FEL (45) and MEME (46). In addition, MEME also identified selection in OG0000033 and OG0004411. These genes were removed for the subsequent analysis. PAML identifies the sites that could be under continuous changes while MEME identifies the episodic positive selection which implies that site changes are kept in clade to provide the advantage in the new environment. As a consequence of selection pressure from the host environment, *C. auris* like other pathogens can undergo evolutionary changes for enhanced survival.

Conserved regions reveal immunogenic and antigenic peptides

For the proteins coded from the genes with no sign of positive selection, we then focused on highly conserved regions of their MSA. This is particularly important to increase the coverage of vaccine candidates as the regions of low Shannon entropy would theoretically remain conserved even upon adding the new *C. auris* strains in the MSA. Therefore, we selected the candidate proteins from conserved regions using Shannon entropy as previously used (56–58). Shannon entropy reflects the degree of variability of protein sequence fragments and supports their evolutionary stability inferences. Stable peptides are characterized by low entropy and an entropy value of 2.0 indicates conserved fragments. With increasing variability of a site, entropy increases and is influenced by both the number of variants at that site and their respective frequency. The estimated average conservancy less than 1.0 postulated that the proposed epitopes would be highly conserved among used *C. auris* strains. We used these conserved fragments originated from 38 alignments and in total identified 7,149 CD4+ T-cell binders.

If the predicted peptides share a higher identity with host genomic regions, they can be imitated as self-molecules. The ability to prevent immunological responses against self-antigens is advantageous which would liberate the vaccine from the risk of inducing autoimmunity. Therefore, to avoid the possibility of causing autoimmunity because of the homology of predicted MHC class-II binders with humans, we performed the blast similarity search analysis (see Materials and methods). For searching against IEDB auto IEDB we used stringent cutoff and only selected the MHC-binders that have less than 35% identity over 80% sequence coverage. This removed >99% of predicted MHC-binders and only 6 MHC-binders could cross

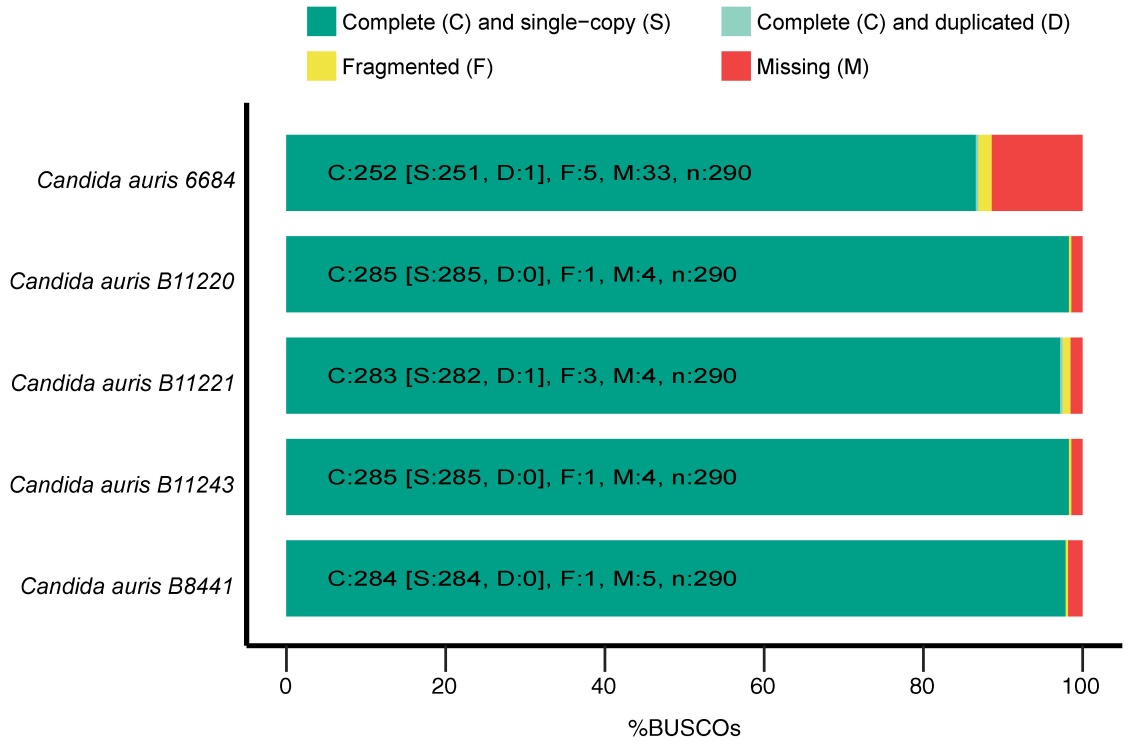


FIGURE 1
Genome statistics. The graph depicts the genome completeness of the investigated *Candida auris* strains as determined by benchmarking universal single-copy orthologs (BUSCO), with BUSCOs for Fungi (odb9) serving as the background.

this threshold (**Supplementary Table 2**). Such reduction was not surprising because of the close evolutionary distance between the eukaryotic pathogen *C. auris* and humans (1,105 Mya) compared to bacterial (such as *Mycobacterium tuberculosis*; 4,290 Mya) pathogens as estimated with TimeTree (59).

Furthermore, only the ability to bind to the MHC receptor does not guarantee that the binder peptide is antigenic. Out of six, only two predicted MHC class-II binders were found to be antigenic. VaxiJen antigenicity score of QTTCFQTEYYDPYIS and FVDPKKCCCDPKMIK was 0.85 and 1.88, respectively with probable fungal antigen annotation at 0.5 threshold (**Figure 3A**). The 3D structure of these epitopes is shown in **Figure 3B**.

High population coverage is required for the recognition of epitopes

The MHC class II molecules are known to be highly polymorphic and are critically involved in recognition and defense against pathogens. The distribution of MHC alleles differs among various ethnic groups worldwide. Therefore, a peptide that acts as a T-cell epitope in majority of a population with a particular MHC alleles distribution may not work in a population with a different MHC alleles distribution. Thus, allelic distribution is crucial also for the adaptation of the population to environmental changes and it various among different geographical regions (60). Both identified epitopes were therefore submitted to IEDB server as described in methodology section. Only MHC class II binders with strong to intermediate binding affinity (IC50 less or equal to 500 nM) were selected to calculate population coverage (**Figure 3C** and **Table 2**). The analysis of both epitopes revealed high coverage for the global population (99.09%). Therefore, we assume that the identified epitopes can be recognized by sufficient HLA alleles and cover most of the world's population. However, we noted significantly lower binding abilities in Central American (57.2%) and South African (45.98%) regions (**Figure 4**), thus only approximately half of the population can recognize selected epitopes. The low binding ability is due to the small number of the epitopes (only 2 epitopes were examined) and low IC50 value. The IC50 value $\leq$500 was selected to obtain more accurate results when even lower load of fungi can be recognized by immune cells, which is crucial mainly for the immunocompromised patients. However, selecting more benevolent values, the number of the binding alleles would increase and thus also the probability for the better coverage in Central American and South African population, where candidemia caused by *C. auris* accounts for 10% of all cases (61).

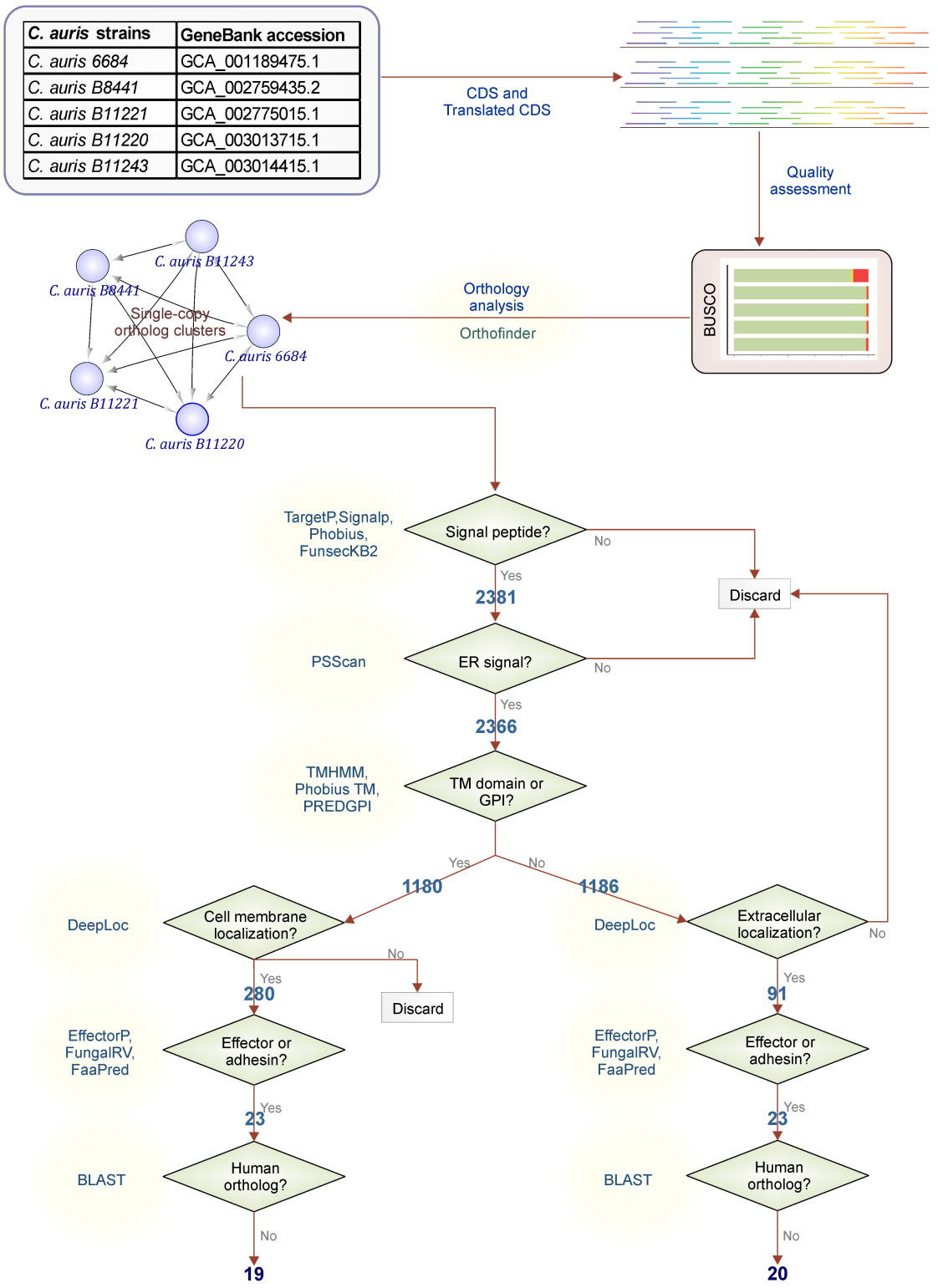


FIGURE 2
Filtering workflow. Steps used for *C. auris* protein filtering adapted from Vivek-Ananth et al. (79) The *C. auris* proteins with orthologs in all strains were filtered to obtain membrane proteins or extracellular secreted proteins. Secretion to the extracellular space or to the membrane was determined by existence of a signal peptide or glycosylphosphatidylinositol (GPI) anchor or a transmembrane (TM) domain. The proteins with a endoplasmic reticulum (ER) signal peptides were filtered out to ensure the localization of the protein to the membrane or extracellular space. Further steps were taken to predict the effector or adhesins and to filter out proteins that may invoke cross-reactivity.

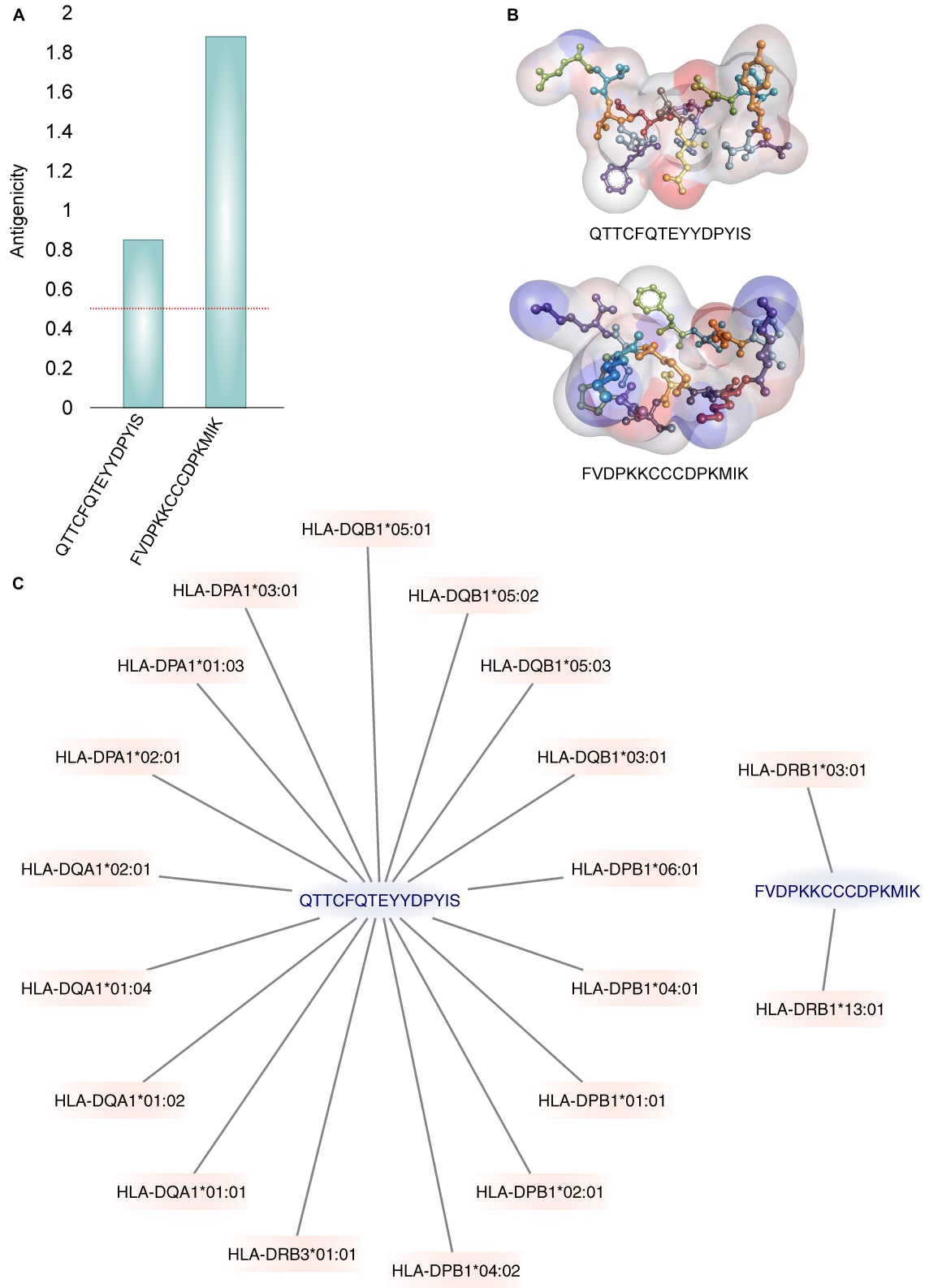


FIGURE 3
The best two identified epitopes. **(A)** The Vaxijen server's calculated antigenic potential (96) is displayed. A horizontal red line indicates the antigenicity threshold. **(B)** Illustrations of the 3D structures of epitopes. The residues are shown under a soft surface and are colored by atom charges. **(C)** Graphs are used to show how binding with MHC-II alleles is displayed. Only the alleles with binding IC50 values ≤500 nM are displayed.

TABLE 2 List of selected alleles used for the population coverage analysis.

Epitope	Allele	IC50	Percentile rank	Allele locus	Prediction method
QTTCFQTEYYDPYIS	HLA-DRB3*01:01	205.6	6.2	HLA-DR	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:01/DQB1*05:01	17.2	0.02	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:02/DQB1*05:02	278.5	1.8	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:04/DQB1*05:03	290.9	0.31	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*02:01/DQB1*03:01	359.7	34	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*06:01	17.40	2.2	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*04:01	35.10	0.51	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*02:01/DPB1*01:01	198.60	2	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*02:01	256.00	6.9	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*03:01/DPB1*04:02	370.4	6.5	HLA-DP	NN-align 2.3
FVDPKKCCCDPKMIK	HLA-DRB1*13:01	368.1	50	HLA-DR	NN-align 2.3
FVDPKKCCCDPKMIK	HLA-DRB1*03:01	430.5	6.3	HLA-DR	NN-align 2.3

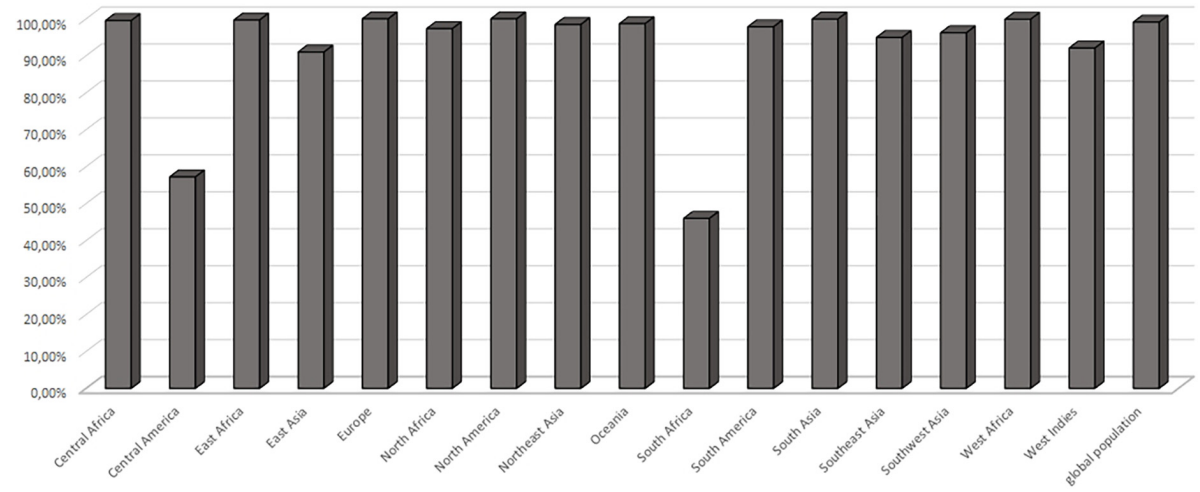


FIGURE 4
Population coverage. Cumulative population coverage for both predicted epitopes QTTCFQTEYYDPYIS and FVDPKKCCCDPKMIK.

Discussion

The best strategy to control an infectious disease in a given population and one of the most efficient, quickest, and affordable ways to promote public health is through vaccination. Given the fact that *C. auris* can be resistant to nearly all available antifungal medications, the most effective way to fight this evolving and fatal pathogen is potentially by using a vaccine targeted against it. Notably, epitope-based vaccines are known to generate a stronger immune response against a pathogen as compared to the whole protein vaccines (62). Previously, vaccine development was solely dependent on experimental techniques which required extensive resources and time. However, with recent advancements in the fields of bioinformatics and reverse vaccinology, as pioneered by the work of Rino Rappouli, (63), we have efficient *in silico* approaches to screen the genome for best epitopes which can thus be employed to design novel vaccines in a cost and time effective manner. These approaches have been already successfully employed by many researchers for identifying potent vaccine candidates (56, 64–67), which further endorses the potential of *in silico* approaches in vaccinology. Reverse vaccinology is now especially poised to produce more vaccines after the promising discovery of the first accepted multicomponent meningococcal serogroup B (MenB) vaccine (4CMenB; Bexsero®) which was proposed from computational approaches and later received huge success in laboratory experiments (63, 68).

Th1 cells (a subtype of CD4+ TH cells) are known to play a central role in providing immunity against fungi and effective fungal vaccines. They also induce generation of pro-inflammatory signature cytokines IFN-γ and TNF-α and play a role in production of opsonizing antibodies which promotes enhanced phagocytosis at the sites of infection (69). In addition,

emerging evidence states that Th17 cells (another subtype of CD4+ TH cells) have also been implicated in the generation of specialized immune response against fungal infection and usually a balance between the Th1 and Th17 associated responses is desired (69, 70). In memory pool of humans, anti-fungal Th17 memory cells have been found and they have also been known to play a role in induced vaccine protection in mice (71). Furthermore, Th17 cells act majorly at mucosal surfaces and thus, Th17 cells inducing vaccines are adequate to protect against deadly lung infection caused by *Blastomyces dermatitidis* in mice and also against three major systemic mycoses in North America (72). Altogether, it is highly suggested to target Th17 induction in the potential vaccine designs against systemic fungal infections (73).

We have previously developed and used different strategies to increase the ability of computer-aided vaccine designing approaches to identify vaccine candidates such as (1) epitope selection from conserved regions (74), (2) profile similarity method to analyze biased-ness of predicted epitopes toward profiles of experimentally validated epitopes (74), (3) introducing cleavage sites for targeted cleavage of multi-epitopic vaccine (56, 75), (4) B-cell epitope prediction by docking (74), (5) structure-based epitope identification (57), (6) CpG optimization (76), (7) sequence homology search against host (77), (8) adding adjuvant such as IL-12 (78), and (9) population coverage analysis (58, 77). In this study, we integrated the evolutionary analysis with our reverse vaccinology, i.e., bioinformatical genome and epitope search pipeline (74, 77) to identify vaccine candidates. The approach we used can be particularly helpful for designing vaccine against any emerging pathogens. Here we used this for epitope identification for potential vaccine against *C. auris*.

Currently, only the NDV-3A vaccine has shown potential so far in immunizing mice against *C. auris* infection (18). We extensively scanned the whole *C. auris* genome coded candidate proteins for their potential as vaccine candidates. We focused on secreted proteins and cell membrane proteins as they are particularly important for fungal host-pathogen interactions. During pathogenesis, these proteins play a role in the interaction with the host immune cells. Moreover, they are good candidates for targeting because of easier access and low expected-resistance mechanisms (79). Therefore, we followed the workflow in **Figure 2** [adapted from Vivek-Ananth et al. (79)] to filter *C. auris* proteins based on existence of a signal peptide, a GPI anchor or a TM domain. The filtered proteins were then classified in two separate groups of secreted and membrane proteins.

Five major clades of *C. auris* have been identified up to date with tens to hundreds of thousands of single nucleotide polymorphisms, which can be potentially linked to various infection strategies and outbreaks (80, 81). It had been shown that highly related isolates play important role in local and ongoing transmission, however, undergone clonal expansion was detected in each clade (82). Moreover, as fungus has evolved as a pathogen under strong selection pressure imposed by climate change, it is important to recognize the rapidly evolved genes of *C. auris* (83). This knowledge is critical not only to understand the transmission dynamics, but also to design the suitable vaccine. In comparison of other genes, rapidly evolved genes have higher possibility to mutate to facilitate the pathogen survival in critical environment. We identified the positive selection signals and removed the proteins coded by such genes. Studies have shown that mutations in viral epitopes can promote hindrance in viral recognition by CD4+ and CD8+ T-cell receptors leading to escape of viruses from immune surveillance, thus escalating pathologic conditions (84). For development of a successful *C. auris* vaccine, it is crucial to have a memory pool of T-cells which can recognize *C. auris* and provide immunity despite its continuous evolution. To maintain high affinity, T-cell memory pool epitopes which are highly conserved and resistant to mutations should be considered. Thus, we focused on the conserved regions of MSAs after elimination of ortholog clusters with signs of positive selection.

Finally, we determined candidates from the conserved regions that are highly immunogenic, antigenic, and that show no significant similarity with host proteins. Furthermore, we also analyzed if the identified MHC class-II epitopes showed any similarity to other pathogenic organisms including fungi (taxid4751), bacteria (taxid2) and parasites from clade of *Protostomia* (taxid33317) that include parasitic worms *Nematoda* and *Platyhelminthes* and other parasites like *Trypanosomatida* (txid2704949) and *Plasmodium* (txid5820). We performed our Blast search with the parameters adjusted for short sequence length (see Host similarities and antigenicity of MHC class-II binders). Our search against bacteria and chosen parasite sequences showed no similar hits to both epitopes (*e*-value < 0.01). On the other hand, we have found seven fungal hits for epitope 1 with the *e*-value lower than 0.01. All hits showed 85.71% identity and 93.3% query coverage (**Supplementary Figure 1**). Six of seven hits are from *Hanseniaspora* species, a yeast genus while 1 is from *Brettanomyces*, a non-spore forming yeast genus. Thus, we also performed more specific Blast search for more possible hits in yeast first against the *Saccharomycetes* (txid4891) class and then directly with the *Candida* genus (txid5475, excluding *C. auris* from the search) of the same class. We found no hits with *e*-value < 0.01 in other *Candida* species for any of the epitopes. However, we found that 42 yeast proteins show similarity to epitope 1 (*e*-values < 0.01) and further investigated these hits. Top seven hits showed query coverage higher than 90% and identity score higher than 85%. When we compared these hits to the previous fungal proteins, we found a 100% overlap. The 17 of the remaining hits also had identity score over 78% and query coverage over 85% for epitope 1. Finally, the 18 remaining hits only showed identity scores higher than 70% and query

coverage higher than 60% and thus were not determined to be similar to epitope 1.

In our work, we used 5 *C. auris* strains spanning 4 clades namely *C. auris* 6,684 Clade I from India, *C. auris* B8441 Clade I from Pakistan, *C. auris* B11220 Clade II from Japan, *C. auris* B11221 Clade III from South Africa, and *C. auris* B11243 Clade IV from Venezuela. For further analysis to determine the clade coverage of our epitopes we also performed a protein sequence comparison against *Candida auris* (taxid:498019). This is especially important, since our analysis did not include the Iran strain IFRC2087, belonging to the recently identified clade V. We found 5 hits that showed 100% identity and coverage to epitope 1 from strains 6,684 (India, Clade I), and JCM 15448 (Japan, Clade II), B11221 (South Africa, Clade III), B11243 (Venezuela, Clade IV), and IFRC2087 (Iran, Clade V). For epitope 2, we also found 5 hits from all clades, respectively from strains 6,684 (India, Clade I), B11220 (Japan, Clade II), B11221 (South Africa, Clade III), B11243 (Venezuela, Clade IV), and IFRC2087 (Iran, Clade V) (**Supplementary Table 3**). Furthermore, since many isolates from different strains only have their genomes sequenced and no protein annotation yet, we have also performed a blast search in the nucleotide database with tblastn algorithm. Our analysis revealed 43 hits for each epitope (with 100% coverage and identity). Moreover, we found these hits to be from the same strains for both epitopes. Among these strains, 32 belong to clade I (28 from Lebanon: Beirut, 1 from China: Beijing, 1 from India, 1 from Italy: Genoa, and 1 from United Arab Emirates), 5 belong to clade II (1 from each Canada: Alberta, China: Shenyang, Japan, South Korea, and USA: New York), 4 belong to clade III (1 from each Canada: Quebec, South Africa, USA: California, USA: Indiana), 2 belong to clade IV (Colombia: Cartagena, and Venezuela) and 1 belongs to clade V from Iran (**Supplementary Figure 2**).

Taking all these results and comparison into account, we further decided to combine these two epitopes using the linker KFERQ (56, 85) so as to generate a more potent immune response from our vaccine construct. We thus performed *in silico* simulations for possible immune responses using the C-IMMSIM server (86, 87). Two injections were given with 4 weeks interval and immune response over a 6 months-long period was predicted. We have used the alleles that we determined to cover larger portion of the global population (namely HLA-DRB1*13:01, HLA-DRB1*03:01, and HLA-DRB3*01:01 since only these three were included in the database). Predicted response for all three combinations of the HLA alleles showed high innate response after the first injection and even higher IgM counts after the second injection. This demonstrates long-term immune response which is still retained upto 6 months (**Supplementary Figure 3**). We also observed an increase in the memory CD4+ T-cells' number to as high as 1,600 cells per mm^3 after the second injection and still a population of 100 cells per mm^3 even after 6 months (**Supplementary Figure 4**).

Since 2019, SARS-CoV-2 belongs to one of the most prevalent viral pathogens worldwide. It is well documented that COVID-19 patients often suffer from various infectious diseases, which also include fungal infections (88–91). This is due to the weakened immune system of the patients, as well as the treatment used to cure COVID-19 infection. A study from 2020 showed that 91.8% of COVID-19 patients were reported to suffer from secondary infections caused by bacterial agents, while 23.3% suffered from fungal co-infections (92). Coronavirus-associated *Candida* infection (CACa) is the third most prevalent fungal disease associated with SARS-CoV-2, with a high mortality rate of 67.849% (91). However, it is necessary to state that the majority of these patients also had other complications or diseases, such as diabetes mellitus, hypertension, or obesity (93). Therefore, due to lack of documentation, it is unclear whether *C. auris* infection alone was responsible for the higher mortality rate or other risk factors significantly contributed toward it. Moreover, a recent meta-analysis showed that men are more frequently affected by CACa as compared to women, with a 3.7 times greater risk of co-infection. Another important fact is that COVID-19 infections may significantly alter host immune response, as a result of which a fungal pathogen can resist its disposal with usual antifungal drugs (94). However, there may also be the possibility that it is a consequence of drug-drug interactions, which are administrated against the COVID-19 and co-infections.

Overall, this study presents a refined strategy to improve the current computer-aided vaccine design. The major limitation of our study is the current lack of the experimental validation, including functional analysis to reveal if our epitopes can really evoke the immune responses in the cell lines and in appropriate animal models. This process is though very challenging, however, we hope that with the right research collaboration, we can switch our work from theoretical study to the experimental work. Our results are very promising despite getting only two epitopes from our conservative approach. Furthermore, we achieved good immunogenicity, antigenicity, and high populations coverage in our predicted vaccine candidates with least possibility of epitopes to get evolved or mutated.

Conclusion

Vaccine design against invasive fungal pathogens has been challenging (95). In recent years, epitope-based vaccines have proven to provoke a potent and efficient defense response of the immune system in a targeted manner. They are synthesized at a lower cost, have good stability and relative safety, and have no limitations in target diseases. Thus, robust computational methods that have considerable predictive and analytical information combined with reverse vaccinology can encourage the prediction of novel epitopes that can serve as powerful vaccine candidates. Here, combining evolutionary

information and employing an immune-informatics approach, we identified two potent vaccine candidates for designing a *C. auris* vaccine. This strategy provides a new approach to identify highly conserved putative T-cell epitopes for other emerging pathogens. We screened the whole *C. auris* genome and rejected the genes showing adaptive evolution. Using the highly conserved regions from candidate proteins, we further enabled the identification of two MHC class-II epitopes QTTCFQTEYYDPYIS and FVDPKKCCCDPKMIK, which besides showing high strain level conservancy, showed good immunogenicity, antigenicity, and no significant similarity with humans. In the follow up study, the results need to be experimentally confirmed by the peptide vaccine formulation in the laboratory followed by clinical trials.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Author contributions

SG and MS designed the project and had the main conceptual ideas and made the first manuscript draft. ÖO, RM, SB, and EB performed the data analysis. SG, ÖO, and MS wrote the manuscript with feedback and the manuscript edits from all authors. TD provided the critical feedback, supervision, and helped to shape the research. All authors contributed to the article and approved the submitted version.

Funding

We gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) project number 210879364 and the CRC/Transregio124 FungiNet (Project B1 – SG and TD). MS would like to thank for financial support by the Gleichstellungsbüro, University of Würzburg, Germany.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.1008527/full#supplementary-material

References

1. Casadevall A, Kontoyiannis DP, Robert V. On the emergence of *Candida auris*: climate change, azoles, swamps, and birds. *mBio.* (2019) 10:e1397–1319. doi: 10.1128/mBio.01397-19

2. Lone SA, Ahmad A. *Candida auris*-the growing menace to global health. *Mycoses.* (2019) 62:620–37. doi: 10.1111/myc.12904

3. Singhal T, Kumar A, Borade P, Shah S, Soman R. Successful treatment of C. auris shunt infection with intraventricular caspofungin. *Med Mycol Case Rep.* (2018) 22:35–7. doi: 10.1016/j.mmcr.2018.08.005

4. Schelenz S, Hagen F, Rhodes JL, Abdolrasouli A, Chowdhary A, Hall A, et al. First hospital outbreak of the globally emerging *Candida auris* in a European hospital. *Antimicrob Resist Infect Control.* (2016) 5:35. doi: 10.1186/s13756-016-0132-5

5. Azar MM, Turbett SE, Fishman JA, Pierce VM. Donor-derived transmission of *Candida auris* during lung transplantation. *Clin Infect Dis.* (2017) 65:1040–2. doi: 10.1093/cid/cix460

6. Roberts SC, Zembower TR, Bolon MK, Kadakia AR, Gilley JH, Ko JH, et al. Successful treatment of a *Candida auris* intra-articular infection. *Emerg Microbes Infect.* (2019) 8:866–8. doi: 10.1080/22221751.2019.1625287

7. Sherry L, Ramage G, Kean R, Borman A, Johnson EM, Richardson MD, et al. Biofilm-forming capability of highly virulent, multidrug-resistant *Candida auris*. *Emerg Infect Dis.* (2017) 23:328–31. doi: 10.3201/eid2302.161320

8. Chakrabarti A, Sood P, Rudramurthy SM, Chen S, Kaur H, Capoor M, et al. Incidence, characteristics and outcome of ICU-acquired candidemia in India. *Intensive Care Med.* (2015) 41:285–95. doi: 10.1007/s00134-014-3603-2

9. Calvo B, Melo AS, Perozo-Mena A, Hernandez M, Francisco EC, Hagen F, et al. First report of *Candida auris* in America: clinical and microbiological aspects of 18 episodes of candidemia. *J Infect.* (2016) 73:369–74. doi: 10.1016/j.jinf.2016.07.008

10. Welsh RM, Bentz ML, Shams A, Houston H, Lyons A, Rose LJ, et al. Survival, persistence, and isolation of the emerging multidrug-resistant pathogenic yeast *Candida auris* on a plastic healthcare surface. *J Clin Microbiol.* (2017) 55:2996–3005. doi: 10.1128/JCM.00921-17

11. Cassone A. Fungal vaccines: real progress from real challenges. *Lancet Infect Dis.* (2008) 8:114–24. doi: 10.1016/S1473-3099(08)70016-1

12. Tso GHW, Reales-Calderon JA, Pavelka N. The elusive anti-*Candida* vaccine: lessons from the past and opportunities for the future. *Front Immunol.* (2018) 9:897. doi: 10.3389/fimmu.2018.00897

13. Liu Y, Filler SG. *Candida albicans* Als3, a multifunctional adhesin and invasin. *Eukaryot Cell.* (2011) 10:168–73. doi: 10.1128/EC.00279-10

14. Spellberg BJ, Ibrahim AS, Avanesian V, Fu Y, Myers C, Phan QT, et al. Efficacy of the anti-*Candida* rAls3p-N or rAls1p-N vaccines against disseminated and mucosal candidiasis. *J Infect Dis.* (2006) 194:256–60. doi: 10.1086/504691

15. Brena S, Omaetxebarria MJ, Elguezabal N, Cabezas J, Moragues MD, Ponton J. Fungicidal monoclonal antibody C7 binds to *Candida albicans* Als3. *Infect Immun.* (2007) 75:3680–2. doi: 10.1128/IAI.01840-06

16. Schmidt CS, White CJ, Ibrahim AS, Filler SG, Fu Y, Yeaman MR, et al. NDV-3, a recombinant alum-adjuvanted vaccine for *Candida* and *Staphylococcus aureus*, is safe and immunogenic in healthy adults. *Vaccine.* (2012) 30:7594–600. doi: 10.1016/j.vaccine.2012.10.038

17. Edwards JE Jr., Schwartz MM, Schmidt CS, Sobel JD, Nyirjesy P, Schodel F, et al. A fungal immunotherapeutic vaccine (NDV-3A) for treatment of recurrent vulvovaginal candidiasis-A phase 2 randomized, double-blind, placebo-controlled trial. *Clin Infect Dis.* (2018) 66:1928–36. doi: 10.1093/cid/ciy185

18. Singh S, Uppuluri P, Mamouei Z, Alqarihi A, Elhassan H, French S, et al. The NDV-3A vaccine protects mice from multidrug resistant *Candida auris* infection. *PLoS Pathog.* (2019) 15:e1007460. doi: 10.1371/journal.ppat.1007460

19. Spellberg B, Ibrahim AS, Lin L, Avanesian V, Fu Y, Lipke P, et al. Antibody titer threshold predicts anti-candidal vaccine efficacy even though the mechanism of protection is induction of cell-mediated immunity. *J Infect Dis.* (2008) 197:967–71. doi: 10.1086/529204

20. Spellberg B, Ibrahim AS, Yeaman MR, Lin L, Fu Y, Avanesian V, et al. The antifungal vaccine derived from the recombinant N terminus of Als3p protects mice against the bacterium *Staphylococcus aureus*. *Infect Immun.* (2008) 76:4574–80. doi: 10.1128/IAI.00700-08

21. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* (2004) 305:371–6. doi: 10.1126/science.1097211

22. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* (2007) 35:D21–5. doi: 10.1093/nar/gkl986

23. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* (2015) 31:3210–2. doi: 10.1093/bioinformatics/btv351

24. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, et al. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* (2015) 43:D250–6. doi: 10.1093/nar/gku1220

25. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* (2017) 45:D744–9. doi: 10.1093/nar/gkw1119

26. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* (2015) 16:157. doi: 10.1186/s13059-015-0721-2

27. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* (2015) 12:59–60. doi: 10.1038/nmeth.3176

28. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* (2000) 300:1005–16. doi: 10.1006/jmbi.2000.3903

29. Armenteros JJA, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* (2019) 37:420–3. doi: 10.1038/s41587-019-0036-z

30. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* (2004) 338:1027–36. doi: 10.1016/j.jmb.2004.03.016

31. Meinken J, Asch DK, Neizer-Ashun KA, Chang G-H, Cooper CR JR., Min XJ. FunSecKB2: a fungal protein subcellular location knowledgebase. *Comput Mol Biol.* (2014) 4:1–17.

32. Pierleoni A, Martelli PL, Casadio R. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics.* (2008) 9:392. doi: 10.1186/1471-2105-9-392

33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* (2001) 305:567–80. doi: 10.1006/jmbi.2000.4315

34. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* (2006) 34:W362–5. doi: 10.1093/nar/gkl124

35. Almagro Armenteros JJ, Sonderby CK, Sonderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics.* (2017) 33:3387–95. doi: 10.1093/bioinformatics/btx431

36. Nami S, Mohammadi R, Vakili M, Khezripour K, Mirzaei H, Morovati H. Fungal vaccines, mechanism of actions and immunology: a comprehensive review. *Biomed Pharmacother.* (2019) 109:333–44. doi: 10.1016/j.biopha.2018.10.075

37. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol.* (2018) 19:2094–110. doi: 10.1111/mpp.12682

38. Chaudhuri R, Ansari FA, Raghunandanan MV, Ramachandran S. FungalRV: adhesin prediction and immunoinformatics portal for human fungal pathogens. *BMC Genomics.* (2011) 12:192. doi: 10.1186/1471-2164-12-192

39. Ramana J, Gupta D. FaaPred: a SVM-based prediction method for fungal adhesins and adhesin-like proteins. *PLoS One.* (2010) 5:e9695. doi: 10.1371/journal.pone.0009695

40. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* (2000) 302:205–17. doi: 10.1006/jmbi.2000.4042

41. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* (2006) 34:W609–12. doi: 10.1093/nar/gkl315

42. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* (2007) 56:564–77. doi: 10.1080/10635150701472164

43. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* (2007) 24:1586–91. doi: 10.1093/molbev/msm088

44. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* (2005) 22:1107–18. doi: 10.1093/molbev/msi097

45. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* (2005) 22:1208–22. doi: 10.1093/molbev/msi105

46. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* (2012) 8:e1002764. doi: 10.1371/journal.pgen.1002764

47. Oliveira L, Paiva PB, Paiva AC, Vriend G. Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins.* (2003) 52:544–52. doi: 10.1002/prot.10490

48. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) 27:379–423.

49. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol.* (1995) 13:587–622. doi: 10.1146/annurev.iy.13.040195.003103

50. Sinigaglia F, Hammer J. Motifs and supermotifs for MHC class II binding peptides. *J Exp Med.* (1995) 181:449–51. doi: 10.1084/jem.181.2.449

51. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* (2019) 47:D339–43. doi: 10.1093/nar/gky1006

52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* (2009) 10:421. doi: 10.1186/1471-2105-10-421

53. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* (2007) 8:4. doi: 10.1186/1471-2105-8-4

54. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* (2011) 7:e1002195. doi: 10.1371/journal.pcbi.1002195

55. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc.* (2021) 1:e323. doi: 10.1002/cpz1.323

56. Gupta SK, Singh A, Srivastava M, Gupta SK, Akhoon BA. In silico DNA vaccine designing against human papillomavirus (HPV) causing cervical cancer. *Vaccine.* (2009) 28:120–31. doi: 10.1016/j.vaccine.2009.09.095

57. Gupta SK, Srivastava M, Akhoon BA, Gupta SK, Grabe N. In silico accelerated identification of structurally conserved CD8+ and CD4+ T-cell epitopes in high-risk HPV types. *Infect Genet Evol.* (2012) 12:1513–8. doi: 10.1016/j.meegid.2012.02.022

58. Singh KP, Verma N, Akhoon BA, Bhatt V, Gupta SK, Gupta SK, et al. Sequence-based approach for rapid identification of cross-clade CD8+ T-cell vaccine candidates from all high-risk HPV strains. *3 Biotech.* (2016) 6:39. doi: 10.1007/s13205-015-0352-z

59. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* (2017) 34:1812–9. doi: 10.1093/molbev/msx116

60. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool.* (2005) 2:16. doi: 10.1186/1742-9994-2-16

61. Govender NP, Magobo RE, Mpembe R, Mhlanga M, Matlapeng P, Corcoran C, et al. *Candida auris* in South Africa, 2012–2016. *Emerg Infect Dis.* (2018) 24:2036.

62. De Groot AS, McMurry J, Marcon L, Franco J, Rivera D, Kutzler M, et al. Developing an epitope-driven tuberculosis (TB) vaccine. *Vaccine.* (2005) 23:2121–31. doi: 10.1016/j.vaccine.2005.01.059

63. Gorringe AR, Pajon R. Bexsero: a multicomponent vaccine for prevention of meningococcal disease. *Hum Vaccin Immunother.* (2012) 8:174–83. doi: 10.4161/hv.18500

64. Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R. Defining putative T-cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. *Vaccine.* (2005) 23:1265–72. doi: 10.1016/j.vaccine.2004.08.046

65. Bencurova E, Gupta SK, Oskoueian E, Bhide M, Dandekar T. Omics and bioinformatics applied to vaccine development against *Borrelia*. *Mol Omics.* (2018) 14:330–40. doi: 10.1039/c8mo00130h

66. Jain R, Jain A, Verma SK. Prediction of epitope based peptides for vaccine development from complete proteome of novel corona virus (SARS-COV-2) using immunoinformatics. *Int J Pept Res Ther.* (2021) 27:1729–40. doi: 10.1007/s10989-021-10205-z

67. Sethi G, Sethi S, Krishna R. Multi-epitope based vaccine design against *Staphylococcus epidermidis*: a subtractive proteomics and immunoinformatics approach. *Microb Pathog.* (2022) 165:105484. doi: 10.1016/j.micpath.2022.105484

68. Martin NG, Snape MD. A multicomponent serogroup B meningococcal vaccine is licensed for use in Europe: what do we know, and what are we yet to learn? *Expert Rev Vaccin.* (2013) 12:837–58. doi: 10.1586/14760584.2013.814862

69. Lin L, Ibrahim AS, Xu X, Farber JM, Avanesian V, Baquir B, et al. Th1-Th17 cells mediate protective adaptive immunity against *Staphylococcus aureus* and *Candida albicans* infection in mice. *PLoS Pathog.* (2009) 5:e1000703. doi: 10.1371/journal.ppat.1000703

70. Bartemes KR, Kita H. Innate and adaptive immune responses to fungi in the airway. *J Allergy Clin Immunol.* (2018) 142:353–63. doi: 10.1016/j.jaci.2018.06.015

71. Romani L. Immunity to fungal infections. *Nat Rev Immunol.* (2011) 11:275–88. doi: 10.1038/nri2939

72. Wuthrich M, Gern B, Hung CY, Ersland K, Rocco N, Pick-Jacobs J, et al. Vaccine-induced protection against 3 systemic mycoses endemic to North America requires Th17 cells in mice. *J Clin Invest.* (2016) 126:795. doi: 10.1172/JCI85788

73. Wuthrich M, Gern B, Hung CY, Ersland K, Rocco N, Pick-Jacobs J, et al. Vaccine-induced protection against 3 systemic mycoses endemic to North America requires Th17 cells in mice. *J Clin Invest.* (2011) 121:554–68. doi: 10.1172/JCI43984

74. Gupta SK, Srivastava M, Akhoon BA, Smita S, Schmitz U, Wolkenhauer O, et al. Identification of immunogenic consensus T-cell epitopes in globally distributed influenza-A H1N1 neuraminidase. *Infect Genet Evol.* (2011) 11:308–19. doi: 10.1016/j.meegid.2010.10.013

75. Ranjbar MM, Gupta SK, Ghorban K, Nabian S, Sazmand A, Taheri M, et al. Designing and modeling of complex DNA vaccine based on tropomyosin protein of *Boophilus* genus tick. *Appl Biochem Biotechnol.* (2015) 175:323–39. doi: 10.1007/s12010-014-1245-z

76. Baloria U, Akhoon BA, Gupta SK, Sharma S, Verma V. In silico proteomic characterization of human epidermal growth factor receptor 2 (HER-2) for the mapping of high affinity antigenic determinants against breast cancer. *Amino Acids.* (2012) 42:1349–60. doi: 10.1007/s00726-010-0830-x

77. Gupta SK, Smita S, Sarangi AN, Srivastava M, Akhoon BA, Rahman Q, et al. In silico CD4+ T-cell epitope prediction and HLA distribution analysis for the potential proteins of *Neisseria meningitidis* Serogroup B–a clue for vaccine development. *Vaccine.* (2010) 28:7092–7. doi: 10.1016/j.vaccine.2010.08.005

78. Akhoon BA, Slathia PS, Sharma P, Gupta SK, Verma V. In silico identification of novel protective VSG antigens expressed by *Trypanosoma brucei* and an effort for designing a highly immunogenic DNA vaccine using IL-12 as adjuvant. *Microb Pathog.* (2011) 51:77–87. doi: 10.1016/j.micpath.2011.01.011

79. Vivek-Ananth RP, Mohanraj K, Vandanashree M, Jhingran A, Craig JP, Samal A. Comparative systems analysis of the secretome of the opportunistic pathogen *Aspergillus fumigatus* and other *Aspergillus* species. *Sci Rep.* (2018) 8:6617. doi: 10.1038/s41598-018-25016-4

80. Muñoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, et al. Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. *Nat Commun.* (2018) 9:5346. doi: 10.1038/s41467-018-07779-6

81. Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. Potential fifth clade of *Candida auris*, Iran, 2018. *Emerg Infect Dis.* (2019) 25:1780–1. doi: 10.3201/eid2509.190686

82. Chow NA, Muñoz JF, Gade L, Berkow EL, Li X, Welsh RM, et al. Tracing the evolutionary history and global expansion of *Candida auris* using population genomic analyses. *mBio.* (2020) 11:e3364–3319. doi: 10.1128/mBio.03364-19

83. Jackson BR, Chow N, Forsberg K, Litvintseva AP, Lockhart SR, Welsh R, et al. On the origins of a species: what might explain the rise of *Candida auris*? *J Fungi.* (2019) 5:58. doi: 10.3390/jof5030058

84. Ciurea A, Hunziker L, Martinic MMA, Oxenius A, Hengartner H, Zinkernagel RM. CD4(+) T-cell-epitope escape mutant virus selected *in vivo*. *Nat Med.* (2001) 7:795–800.

85. Chiang HL, Terlecky SR, Plant CP, Dice JF. A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins. *Science.* (1989) 246:382–5. doi: 10.1126/science.2799391

86. Castiglione F, Deb D, Srivastava AP, Liò P, Liso A. From infection to immunity: understanding the response to SARS-CoV2 through in-silico modeling. *Front Immunol.* (2021) 12:646972. doi: 10.3389/fimmu.2021.646972

87. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One.* (2010) 5:e9862. doi: 10.1371/journal.pone.0009862

88. Ezeokoli OT, Gcilitshana O, Pohl CH. Risk factors for fungal co-infections in critically ill COVID-19 patients, with a focus on immunosuppressants. *J Fungi.* (2021) 7:545. doi: 10.3390/jof7070545

89. Hoenigl M, Seidel D, Sprute R, Cunha C, Oliverio M, Goldman GH, et al. COVID-19-associated fungal infections. *Nat Microbiol.* (2022) 7:1127–40. doi: 10.1038/s41564-022-01172-2

90. Song G, Liang G, Liu W. Fungal co-infections associated with global COVID-19 pandemic: a clinical and diagnostic perspective from China. *Mycopathologia.* (2020) 185:599–606. doi: 10.1007/s11046-020-00462-9

91. Vaseghi N, Sharifisooraki J, Khodadadi H, Nami S, Safari F, Ahangarkani F, et al. Global prevalence and subgroup analyses of Coronavirus disease (COVID-19) associated *Candida auris* infections (CACa): a systematic review and meta-analysis. *Mycoses.* (2022) 65:683–703. doi: 10.1111/myc.13471

92. Zhu X, Ge Y, Wu T, Zhao K, Chen Y, Wu B, et al. Co-infection with respiratory pathogens among COVID-2019 cases. *Virus Res.* (2020) 285:198005. doi: 10.1016/j.virusres.2020.198005

93. Vinayagamoorthy K, Pentapati KC, Prakash H. Prevalence, risk factors, treatment and outcome of multidrug resistance *Candida auris* infections in Coronavirus disease (COVID-19) patients: a systematic review. *Mycoses.* (2022) 65:613–24. doi: 10.1111/myc.13447

94. Arastehfar A, Carvalho A, van de Veerdonk FL, Jenks JD, Koehler P, Krause R, et al. COVID-19 associated pulmonary aspergillosis (CAPA)—from immunology to treatment. *J Fungi.* (2020) 6:91. doi: 10.3390/jof6020091

95. Medici NP, Del Poeta M. New insights on the development of fungal vaccines: from immunity to recent challenges. *Memor Inst Oswaldo Cruz.* (2015) 110:966–73.

96. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* (2007) 8:4. doi: 10.1186/1471-2105-8-4

97. Thumuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* (2022) 50:W228–34. doi: 10.1093/nar/gkac278

98. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance.* (2019) 2:e201900429. doi: 10.26508/lsa.201900429

99. Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.* (2007) 35:W429–32. doi: 10.1093/nar/gkm256

100. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason 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. doi: 10.1038/s41587-021-01156-3

101. Lum G, Min XJ. FunSecKB: the fungal secretome knowledgebase. *Database (Oxford).* (2011) 2011:bar001. doi: 10.1093/database/bar001

Frontiers in Medicine

Translating medical research and innovation into improved patient careA multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact