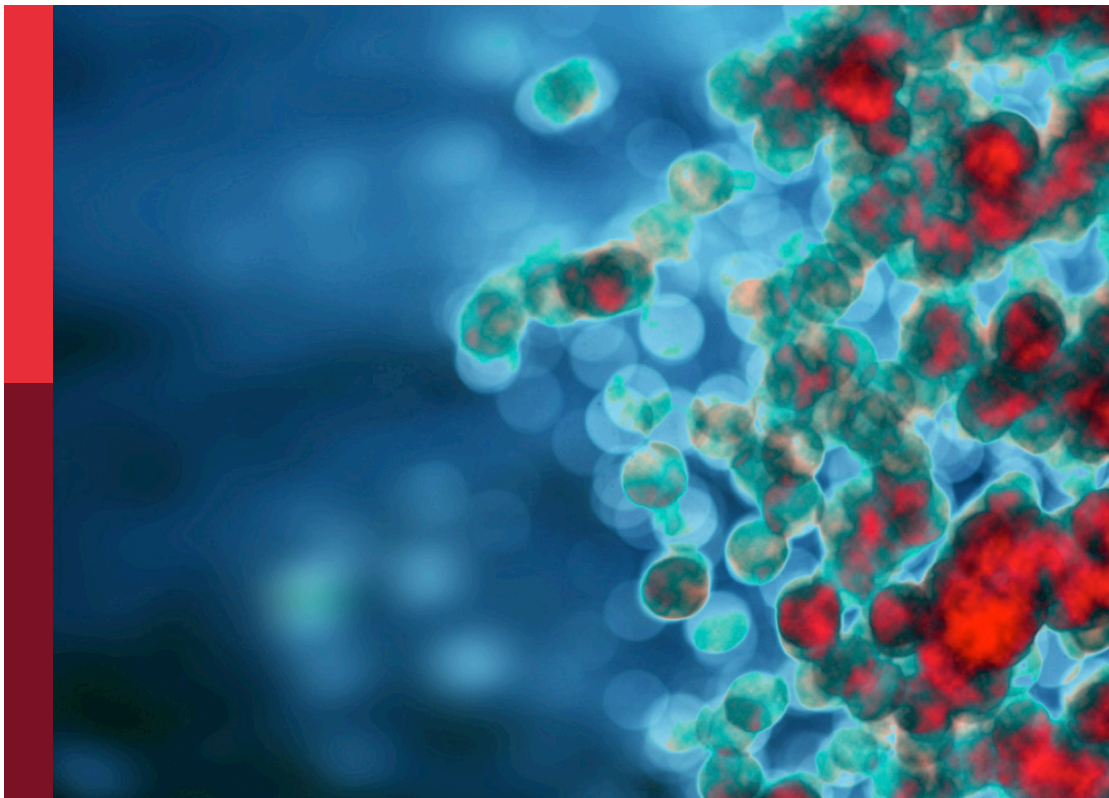
frontiers | Research Topics

Data-driven vaccine design for microbial-associated diseases

Edited by
Gurudeeban Selvaraj, Satyavani Kaliamurthi and Dongqing Wei

Published in
Frontiers in Immunology



ISSN 1664-8714
ISBN 978-2-8325-7440-9
DOI 10.3389/978-2-8325-7440-9

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Data-driven vaccine design for microbial-associated diseases

Topic editors

Gurudeeban Selvaraj — Department of Medical Biotechnology, Aarupadai Veedu Medical College and Hospital, India

Satyavani Kaliamurthi — Department of Biomaterials, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, India

Dongqing Wei — Shanghai Jiao Tong University, China

Citation

Selvaraj, G., Kaliamurthi, S., Wei, D., eds. (2026). *Data-driven vaccine design for microbial-associated diseases*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-7440-9

Table of contents

frontiers | Frontiers in Immunology

TYPE Editorial
PUBLISHED 15 January 2026
DOI 10.3389/fimmu.2025.1749184

OPEN ACCESS

EDITED AND REVIEWED BY
Srinivasa Reddy Bonam,
Indian Institute of Chemical Technology (CSIR), India

*CORRESPONDENCE
Gurudeeban Selvaraj
gurudeeban.selvaraj@avmc.edu.in

RECEIVED 18 November 2025
ACCEPTED 23 December 2025
PUBLISHED 15 January 2026

CITATION
Selvaraj G, Kaliamurthi S and Wei D (2026) Editorial: Data-driven vaccine design for microbial-associated diseases.
Front. Immunol. 16:1749184.
doi: 10.3389/fimmu.2025.1749184

Editorial: Data-driven vaccine design for microbial-associated diseases

Gurudeeban Selvaraj[1]*, Satyavani Kaliamurthi[2,3] and Dongqing Wei[4,5,6]

[1]Department of Medical Biotechnology, Faculty of Interdisciplinary Studies, Aarupadai Veedu Medical College and Hospital (AVMC&H), Vinayaka Mission Research Foundation (Deemed to be University), Puducherry, India, [2]Centre for Research in Molecular Modeling (CERMM), Department of Chemistry and Biochemistry, Concordia University, Montreal, QC, Canada, [3]Department of Biomaterials, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Chennai, India, [4]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [5]Qihe Laboratory, Qishui Guang East, Qibin District, Hebi, Henan, China, [6]Zhongjing Research and Industrialization Institute of Chinese Medicine, Nanyang, China

KEYWORDS

artificial intelligence, drug delivery system, gut microbiota, immunoinformatics, multi-epitope vaccine, structural modeling, tuberculosis, Zika virus

Editorial on the Research Topic
Data-driven vaccine design for microbial-associated diseases

Vaccinology is rapidly evolving, driven by the convergence of genomics, immunoinformatics, and artificial intelligence (AI). As infectious diseases continue to challenge global health whether through re-emerging bacterial pathogens, rapidly evolving viruses, or opportunistic microbiota computational tools are becoming central to vaccine discovery. This Research Topic brings together ten diverse yet thematically connected studies that collectively demonstrate how modern vaccinology is shifting toward precision, integration, and predictive modeling. Taken together, these studies illuminate the future direction of the field, data-driven, multi-targeted, and strategically optimized vaccine design.

To begin with, the long-standing shortcomings of the Bacillus Calmette-Guérin (BCG) vaccine in preventing adult pulmonary tuberculosis highlight the urgent need for improved alternatives. In this Research Topic, one study takes a rational genetic approach by creating progressively attenuated *M. tuberculosis* H37Rv strains. By examining macrophage transcriptomic responses to these engineered strains, the authors reveal strong activation of immune pathways including nuclear factor kappa B (NF-κB), tumor necrosis factor (TNF), chemokine signaling, and notably interleukin-17 (IL-17) signaling. Importantly, this upregulation across all vaccine strains suggests a capacity to elicit robust mucosal immunity, thus providing a promising foundation for next-generation TB vaccines (Veerapandian et al.). This work also exemplifies how integrating pathogen genomics with host response profiling can accelerate rational vaccine design.

Building on the theme of precision, the second study focuses on cervical cancer–associated high-risk *human papillomavirus* (HPV) subtypes (Cai et al.). Although current prophylactic vaccines provide broad protection, subtype-specific insights remain essential for refining immunogen design. Through detailed *in silico* profiling of HPV-31 and HPV-

52 E6/E7 proteins, the authors identify physicochemical properties, dominant B- and T-cell epitopes, and structural determinants of immunogenicity. These findings not only deepen our understanding of oncogenic HPV variants but also pave the way for subtype-tailored vaccine approaches.

Continuing with mycobacterial pathogens, another study addresses the rising burden of non-tuberculous mycobacteria. By analyzing complete genomes from *M. avium*, *M. intracellulare*, and *M. colombiense*, the authors design a multi-epitope vaccine based on conserved regions of the antigen 85 family. Furthermore, population coverage analysis ensures relevance across African populations, while immune simulations predict strong humoral and cellular responses (Kashiri et al.). Consequently, this cross-species construct represents a significant step toward broad-spectrum mycobacterial immunization strategies.

Similarly, the challenge of human immunodeficiency virus (HIV) vaccine development stems largely from viral variability. In this Research Topic, a study explores a dual strategy: mapping immunodominant epitopes and introducing targeted mutations to enhance recognition across HIV subtype C variants (Kumar Mishra et al.). Structural modeling, TLR3 docking, and long-timescale molecular dynamics simulations collectively demonstrate stable vaccine-receptor interactions. In addition, strong predicted immunoglobulin responses and favorable codon adaptation highlight its translational potential. Thus, this work underscores how rational mutation of epitopes may help overcome viral diversity.

Transitioning to viral encephalitides, one study presents a refined multi-epitope subunit vaccine targeting SLEV. Unlike earlier efforts that focused solely on the E protein, this work incorporates membrane protein M and anchored capsid protein anchC, thereby broadening antigenic coverage. The resulting constructs exhibits high structural stability and strong TLR-4 binding, and immune simulations further indicate robust immunogenicity (Ramalingam et al.). Hence, this expanded antigen strategy showcases how multi-protein approaches can enhance vaccine efficacy against complex RNA viruses.

Expanding beyond classical pathogens, another study examines *Ruminococcus gnavus*, a gut pathobiont implicated in inflammatory bowel disease. Through subtractive proteomics, the authors identify two key virulent proteins and construct a multi-epitope vaccine showing strong TLR4 interaction and structural stability (Dingding et al.). Although experimental validation remains necessary, this work importantly demonstrates how vaccinology can be extended to microbiota-associated diseases, potentially transforming future therapeutic approaches for chronic inflammatory conditions.

In addition, the Research Topic features a comprehensive computational pipeline for designing multi-epitope vaccines against human respiratory syncytial virus (hRSV). By mining conserved regions of F and G glycoproteins and evaluating antigenicity, allergenicity, and structural features, the authors identify promising candidates with strong docking affinity for TLR1 and TLR4 (Alnajran et al.). Coupled with immune simulations predicting high IgG, IgM, IL-2, and IFN-γ levels, the work offers a compelling alternative to the limited RSV vaccines currently available for older adults.

Meanwhile, the Zika virus continues to pose a threat in Asia, particularly India. This study identifies novel linear and conformational epitopes in both envelope and NS1 proteins of circulating Indian strains and evaluates their interactions with potent neutralizing antibodies (Roy et al.). The discovery of epitopes capable of strong engagement with monoclonal antibodies such as ZV-67 and Z3L1 provides critical information for developing next-generation, lineage-specific Zika vaccines.

Further reinforcing the theme of genomic integration, another study conducts a pangenome analysis to identify conserved virulence determinants in *Pseudomonas aeruginosa* (Elavarasu and K). Prioritizing the outer membrane protein *Lpt*F, the authors design a multi-epitope vaccine with stable TLR interactions and predicted high expression in *E.coli*. Immune simulations additionally indicate strong adaptive responses, including memory B-cell and T-cell activation. Therefore, this construct holds promise for addressing antibiotic-resistant *P. aeruginosa* infections.

Finally, moving from antigens to delivery systems, the Research Topic concludes with an innovative AI-driven framework for optimizing lipid nanoparticles formulations for mRNA vaccines (Di Salvatore et al.). By generating synthetic transcriptomic datasets to emulate tissue-specific responses, and integrating random forest modeling with a genetic algorithm, the authors identify nanoparticles designs with minimized off-target immune activation. As a result, this purely in silico pipeline offers a paradigm shift toward safer and more targeted mRNA vaccine delivery strategies.

These ten studies collectively showcase the transformative impact of computational biology in rational vaccine design. By integrating structural biology, immunoinformatics, molecular docking, AI, and immune simulations, each contribution extends the frontiers of vaccinology beyond traditional paradigms. Moreover, the spectrum of pathogens addressed from bacteria and viruses to gut microbiota demonstrates the versatility and applicability of these approaches across disease domains. As we face future outbreaks and emerging antimicrobial resistance, these studies lay the groundwork for agile, intelligent, and personalized vaccine development. I extend my gratitude to all contributing authors for their innovative efforts, and I am confident that this body of work will inspire further translational and experimental endeavors in infectious disease research.

Author contributions

GS: Conceptualization, Writing – original draft, Writing – review & editing. SK: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. DW: Conceptualization, Writing – original draft, Writing – review & editing.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 27 February 2025
DOI 10.3389/fimmu.2025.1534737

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Concordia University, Canada

REVIEWED BY
Prasanna Srinivasan Ramalingam,
Vellore Institute of Technology, India
Satyavani Kaliamurthi,
Saveetha Institute of Medical and Technical Sciences, India

*CORRESPONDENCE
Gajanan Sapkal
gajanansapkalniv@gmail.com
Pratip Shil
shilpratip@gmail.com

RECEIVED 26 November 2024
ACCEPTED 31 January 2025
PUBLISHED 27 February 2025

CITATION
Roy RR, Tadkalkar N, Deshpande GR, Atre NM, Shil P and Sapkal G (2025) Identification of B-cell epitopes of Indian Zika virus strains using immunoinformatics. *Front. Immunol.* 16:1534737. doi: 10.3389/fimmu.2025.1534737

Identification of B-cell epitopes of Indian Zika virus strains using immunoinformatics

Rohan Raj Roy[1], Nitali Tadkalkar[1], Gururaj Rao Deshpande[1], Nitin M. Atre[2], Pratip Shil[2*] and Gajanan Sapkal[1*]

[1]Diagnostic Virology Group, Indian Council of Medical Research (ICMR) - National Institute of Virology, Pune, India, [2]Bioinformatics and Data Management, Indian Council of Medical Research (ICMR) - National Institute of Virology, Pune, India

Introduction: The Zika virus is an emerging Flavivirus known to cause Zika infection in humans. It is associated with severe health problems such as microcephaly and Guillain-Barré syndrome post the Brazilian epidemic in 2015-16. The spread of the Zika virus to the Asian subcontinent, especially to India is a matter of great concern. Two recent co-circulating Indian Zika virus strains such as Rajasthan and Maharashtra detected in 2018 and 2021 were studied to identify B-cell epitopes in the envelope and non-structural 1 protein as these epitopes are major indicators of robust humoral immune response. The study aimed at identifying novel epitopes, followed by molecular docking with potent Zika virus-specific monoclonal antibodies. The novel epitopes identified in this study shall be essential in designing multi-epitope vaccines capable of inducing antibody response against Zika virus infection.

Methods: ABCpred, BepiPred 2.0 and Kolaskar-Tongaonkar methods were used for predicting the linear B-cell epitopes, and Discotope 2.0 and ElliPro were used for the prediction of conformational epitopes. Linear epitopes were further checked for protective antigenicity, allergenicity and toxicity. Based on the stringent study design criteria, only the novel epitopes were considered for molecular docking with complementary determining regions of potent Zika virus-specific monoclonal antibodies.

Results: Nineteen linear and five conformational epitopes were shortlisted based on protective potential, non-allergic and non-toxic properties for Zika virus E protein, from which nine linear and three conformational epitopes were identified as novel. Molecular docking studies revealed that the novel linear epitopes, one each from EDIII, EDII, EDI and EDI/DIII hinge were involved in epitope-CDR interactions with potent neutralizing Zika virus E-specific mouse monoclonal antibody ZV-67. Moreover, the novel EDII epitope was exclusively engaged in epitope-CDR interactions of potent neutralizing Zika virus E-specific human monoclonal antibody Z3L1. None of the linear epitopes of Zika virus NS1 were ascertained as novel based on our study criteria. Conformational epitopes were identified as novel for NS1 protein.

Conclusion: This study identified Zika virus-specific novel epitopes of envelope and non-structural -1 proteins in the currently co-circulating Indian strains. Furthermore, in-silico validation through molecular docking added insight into antigen-antibody interactions, paving way for future *in vitro* and *in vivo* studies.

KEYWORDS

Zika virus, immunoinformatics, B-cell epitopes, neutralization, monoclonal antibodies, molecular docking, complementary determining regions, 2-D interaction maps

1 Introduction

Zika virus (ZIKV) infection is an emerging Flavivirus transmitted by the *Aedes* mosquitoes, mainly *Aedes aegyptii* and *Aedes albopictus*. ZIKV originated in Africa and was first isolated from the serum of rhesus macaque at the Zika forest in Uganda in 1947 (1). Over the decades, ZIKV infection prevailed in the African continent and South Asia, happening sporadically with minimal symptoms and mild illness due to which it was not considered a serious health problem until the first major outbreak in 2007 in the Yap islands in Micronesia (2), followed by French Polynesia in 2013-14 (3). It was a matter of serious concern when the ZIKV accumulated various mutations and spread to South America through Brazil in late 2014 which caused a major outbreak associated with microcephaly and birth defects in newborns and infants (4). In 2016, there was a 20-fold increase of ZIKV infection cases in Brazil with around 304 cases of microcephaly (4) Moreover, ZIKV infection was also implicated in causing Guillain–Barré syndrome (GBS) (5). ZIKV also spread to Central America particularly, the United States. Therefore, the World Health Organization (WHO) declared ZIKV infection a Public Health Emergency of International Concern in 2016 (6, 7). The ZIKV surveillance was started in India in March 2016 through the network of Virus Research Diagnostic Laboratories (VRDLs) by the Indian Council of Medical Research (ICMR), following which sporadic cases of ZIKV were found in Gujarat (2016–17) and Tamil Nadu (2017). In late 2018, one hundred fifty-nine and one hundred twenty-seven ZIKV cases were reported in Rajasthan and Madhya Pradesh, respectively (8–10). The identification of the Rajasthan strain led to the first laboratory confirmation of the existence of the Zika Virus in India (8). Moreover, in the year 2021, Maharashtra reported their first cases of Zika virus infections which were confirmed in Belsar village in Pune district (11). Since then, Zika virus cases have expanded to Kerala, Uttar Pradesh and 16 states of India which is a matter of concern as currently there is no universal vaccine or antivirals. The symptoms of ZIKV are self-limited and resolve usually between 2-7 days. Small subgroups of patients may develop serious complications like GBS that require hospitalization and monitoring of mechanical ventilation, intravenous immunoglobulin and electrophoresis (12–14).

Humoral immune response to ZIKV infection is one of the major ways to accomplish protective immunity regulated by B-cells. ZIKV-specific B-cells are activated in response to infection which secrete IgG and IgA antibodies post-IgM antibodies produced during the acute phase of infection (15). Antigen-antibody interaction studies are important in understanding the immune response in viral infections (16, 17). B-cell epitopes are crucial for understanding protective immunity as well as immunopathogenesis (18). Neutralizing antibodies are mostly produced against ZIKV envelope (E) proteins which tend to identify and bind to specific B-cell epitopes to elicit robust protection (19). The most potent neutralizing antibodies bind to B-cell epitopes on the E domain III (EDIII) which is the least conserved region among different ZIKV strains. Hence, EDIII-specific epitopes are regarded as ZIKV type-specific (20). The other domains such as E domain I (EDI) and E domain II (EDII) are generally regarded as ZIKV/DENV and Flavivirus cross-reactive domains, respectively and epitopes of these domains also induce protective immunity to ZIKV (21). Among the non-structural proteins, Non-structural 1 (NS1) is the most enigmatic protein of the *Flaviviruses*. ZIKV NS1 has a multifunctional role in viral replication, pathogenesis and immune evasion (22). Recently, ZIKV NS1 has been seen as a potential vaccine candidate as it contains epitopes targeted by ZIKV-specific monoclonal antibodies (mAbs) (23, 24). The NS1 protein exists in two forms, a dimer or/and a hexamer (secretory NS1). Similar to the E protein, NS1 also possess three important domains such as β-roll, wing domain with three subdomains (α/β subdomain, long intertwined loops and discontinuous connector subdomain), and β-ladder which also contains the spaghetti loop (25, 26). Most ZIKV-NS1-specific mAbs, target epitopes on the wing-domain and β-ladder domains. To be precise, the epitopes which are a part of the exposed and outer surface of NS1, such as the spaghetti loop residues of the β-ladder and the first half of the intertwined loop of the wing domain are the most effective targets of the antibody response (26). However, the other NS1 domains and subdomains may also contain certain epitopes which may be of potential research interest regarding ZIKV infection and inhibition.

The emergence of immuno-informatics allows the use of various prediction tools and software to compare and analyze various aspects of virus-induced immune response in a less time-consuming, and cost-effective manner. As a result, various computational methods have been used to predict potential B-cell epitopes for arboviruses (16, 17, 27, 28).

In this study, we have undertaken a domain-specific approach to identify and compare the ZIKV E and NS1 B-cell epitopes in both the Indian ZIKV strains which are currently in co-circulation using the epitope prediction tools, present at the immune epitope database (IEDB; http://tools.iedb.org/main/bcell/). We have predicted both linear and conformational B-cell epitopes, the former composed of residues continuous in the sequence and the latter being distantly separated in the sequence but possessing spatial closeness. The epitopes predicted in this study have been compared with the known ZIKV-specific B-cell neutralizing epitopes enlisted in the IEDB. Epitopes with zero per cent identity with known ZIKV-specific neutralizing epitopes (IEDB) were considered novel. Subsequently, these epitopes were mapped to the three-dimensional (3-D) structures of the E and NS1 proteins of both the Indian ZIKV strains. Then these were compared to the prototype African ZIKV MR766 (African lineage) and the Brazilian ZIKV Natal RGN (Asian lineage) associated with microcephaly. Finally, these novel epitopes were used to study their binding interactions with complementary determining regions (CDRs) of ZIKV-specific highly neutralizing monoclonal antibodies (mAbs) by molecular docking analysis. The resulting 3-D and two-dimensional (2-D) interaction maps were used to select the novel epitopes possessing intermolecular bonding interactions such as conventional hydrogen (H) bonds with the CDRs of the neutralizing mAbs. This shall be essential in designing ZIKV-specific peptides for Indian ZIKV strains.

2 Materials and methods

2.1 ZIKV E and NS1 protein sequence retrieval and antigenicity prediction

The polyprotein sequences of both the Indian ZIKV strains i.e. ZIKV Rajasthan (ZIKV_RAJ; GenBank ID: AZS35409.1) and ZIKV Maharashtra (ZIKV_MAH; GenBank ID: UBI73854.1) were obtained from the NCBI protein database. The polyprotein sequences of the prototype African strain (ZIKV MR766; GenBank ID: YP 009227198.1) and the Brazilian ZIKV strain associated with microcephaly (ZIKV Natal RGN; GenBank ID: YP 009428568.1) were used as a reference for comparison. The amino acid sequences of the E and NS1 proteins of these ZIKV were extracted from the polyprotein sequences and subjected to antigenicity prediction in the VaxiJen version 2.0 (29). This server used an alignment-free approach for antigen prediction based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties.

2.2 Domain-specific antigenicity prediction

The retrieved ZIKV E and NS1 sequences were subjected to multiple sequence alignment using MEGA11 with ClustalW and MUSCLE alignment algorithms. The domain-wise antigenicity of the E and NS1 proteins of ZIKV_RAJ and ZIKV_MAH were calculated with the help of VaxiJen v2.0 having a threshold of 0.4. These antigenicity scores were also compared with the antigenicity of ZIKV MR766 and ZIKV Natal RGN.

2.3 Three-dimensional structure prediction

3-D structures of these proteins were predicted by homology modelling protocols as defined in MODELLER v10. 4. The templates for the E and NS1 proteins of ZIKV_RAJ and ZIKV_MAH were searched at NCBI using Protein Data Bank (PDB) via Position-Specific Iterated BLAST (PSI-BLAST). The PDB IDs which had the highest query cover (~100%), per cent identity (~100%) and lowest E values (=0) with the target sequences were selected as the templates. All 3-D models generated were validated via SAVES SERVER (https://saves.mbi.ucla.edu/) using PROCHECK. The best model was selected considering the amino acid occupancies in Ramachandran plots.

2.4 Domain-specific linear B-cell epitope prediction

Linear B-cell epitope prediction for both E and NS1 proteins was carried out by using ABCpred, BepiPred 2.0 and Kolaskar & Tongaonkar methods. ABCpred mediated epitope prediction used a machine-learning-based artificial neural network algorithm, whereas the IEDB B cell epitope prediction tools such as BepiPred 2.0 and Kolaskar Tongaonkar methods were based on random forest algorithm and semi-empirical antigenicity, respectively (30–32). The common epitope sequences from the three methods were considered for further analysis, with a minimal length ranging from 5-25 amino acids for each predicted epitope (33). To determine the exposed and buried residues, the surface accessibility of these epitopes was calculated using the Emini surface accessibility scale (34) and the hydrophilicity of the epitopes was determined by Parker hydrophilicity (32). Based on all the above parameters, the list of predicted B-cell epitopes specific to ZIKV E and NS1 was tabulated. Multiple sequence alignments of the proteins of ZIKV_RAJ, ZIKV_MAH, ZIKV MR766 and ZIKV NATAL RGN strains were used for comparative analysis (Supplementary Figures S1, S2). The ZIKV-specific B-cell epitopes of the E and NS1 proteins aligned to the respective domains were individually checked for being antigenic, non-allergic and non-toxic by Vaxijen v2.0 (29, 35–37), respectively.

2.5 Prediction of novel and overlapping domain-specific linear B-cell epitopes

The IEDB epitope database was searched for all the linear B-cell neutralizing epitopes of ZIKV. The search strategy for ZIKV-specific epitopes was linear epitopes with exact matches of the organism Zika virus (ID:64320); host as Homo sapiens (human) (ID:9606), Mus musculus (mouse) (ID:10090), and Mus musculus C57BL/6 (ID:10000067, c57) and the filter was set as "B-cell neutralization; biological activity (neutralization)" to obtain all the ZIKV-specific neutralizing epitopes submitted at IEDB till date. The predicted ZIKV

E-domain specific epitopes were compared individually with these IEDB-neutralizing epitopes using BLAST~70% to determine overlapping (BLAST~70% positive) and non-overlapping (BLAST~70% negative) epitopes. The non-overlapping epitopes were further checked for per cent identity based on amino acid composition with the known IEDB-neutralizing epitopes to determine the novel epitopes (Zero per cent identity).

To identify overlapping and non-overlapping ZIKV NS1-domain specific epitopes, the search strategy in IEDB was similar to that of the E epitopes, except for the filter, which was set as "B-cell antibody binding (any method)". Similarly, all the predicted ZIKV NS1-domain specific epitopes were compared individually with the ZIKV NS1 IEDB epitopes using Blast~70%.

2.6 Prediction of domain-specific conformational B-cell epitopes

The conformational B-cell epitopes were predicted by two IEDB-based methods i.e. Discotope 2.0 and ElliPro (38, 39). Three-dimensional structures were used as input for epitope prediction. Discotope 2.0 predicted the epitopes based on their solvent accessibility, contact numbers and propensity scores while ElliPro used surface protrusion, accessibility and flexibility.

2.7 Prediction of novel and overlapping domain-specific conformational B-cell epitopes

The IEDB epitope database was searched for all the conformational B-cell epitopes of ZIKV. The search strategy was linear epitope prediction as "discontinuous epitopes with exact matches", and the filter selected as "B-cell neutralization; biological activity (neutralization)" for both ZIKV E and NS1 protein. Epitopes which were not identical post-comparison with IEDB-neutralizing discontinuous epitopes were considered novel and the others were regarded as overlapping or identical.

2.8 Mapping and visualization of predicted epitopes to the ZIKV E and NS1 domains

BIOVIA Discovery Studio Visualizer is a free, feature-rich molecular modelling application for viewing, sharing and analyzing protein and small molecule data. All the relevant epitopes predicted were mapped onto their protein structure in their respective domains and visualized through the BIOVIA Discovery Studio Visualizer

2.9 Molecular docking of ZIKV E and NS1-specific novel epitopes with monoclonal antibodies

The ZIKV 3-D models of E and NS1 were subjected to molecular docking with the 3-D structures of highly neutralizing ZIKV mAbs by using HDock (http://hdock.phys.hust.edu.cn/) which incorporates a hybrid-docking algorithm (40). The best models were selected by analyzing their receptor-ligand interface residues required to identify epitope-CDR interactions. Among the top 10 models, the most suitable epitope-CDR interactions were selected by comparing all the relevant 3-D and 2-D receptor-ligand interaction maps in BIOVIA Discovery Studio Visualizer.

2.10 Study design

The overall study design is depicted in the following Figure 1.

3 Results

3.1 Protective antigenicity of both E and NS1 proteins of ZIKV

The predicted antigenicity scores of the E protein of both ZIKV_RAJ and ZIKV_MAH were found to be 0.6268 and 0.6417, respectively using VaxiJen version 2.0 (29). Similarly, the E proteins of ZIKV MR766 and ZIKV Natal RGN had antigenicity scores of 0.6276 and 0.6205, respectively. The same methodology was used for predicting the protectiveness of the ZIKV NS1 proteins calculated as ZIKV_RAJ (0.4487); ZIKV_MAH (0.4630); ZIKV MR766 (0.4455); and ZIKV Natal RGN (0.4607) Both the ZIKV E and NS1 proteins were considered as protective antigens with the E protein having higher antigenic scores than NS1.

3.2 Delineation and comparisons of ZIKV domain-specific antigenicity of E and NS1

The ZIKV E protein consisted of three domains namely domains 1,2 and 3 (EDI, EDII and EDIII), respectively. The positions of these three domains were mapped on the E protein of the ZIKV strains (41) (Supplementary Table S1). In the case of NS1, similarly, three different domains were identified as β-Roll, Wing-domain, and β-ladder, and their positions were mapped accordingly (25) (Supplementary Table S2). To dissect the domain-specific epitopes of the Indian ZIKV, it was necessary to evaluate the antigenicity of the respective domains (Supplementary Tables S1, S2). This gave a comparative analysis of the domains as well as differences in antigenicity between the ZIKV strains. These values were also compared with the antigenicity of ZIKV MR766 and ZIKV NATAL RGN strains.

It was evident from Supplementary Table S1 that all the domains of the E protein are antigenic with a minor non-antigenic region in EDII (52–131) in the case of both the Indian strains and Brazilian ZIKV NATAL RGN, unlike the African ZIKV MR766. Moreover, EDII was highly conserved across the ZIKV strains demonstrated by identical antigenicity scores., EDI had varying antigenicity scores, except positions (1–50, 39) which had identical antigenicity across all ZIKV strains. EDI (132–191) and EDI (280–295) had different antigenicity scores for both Indian

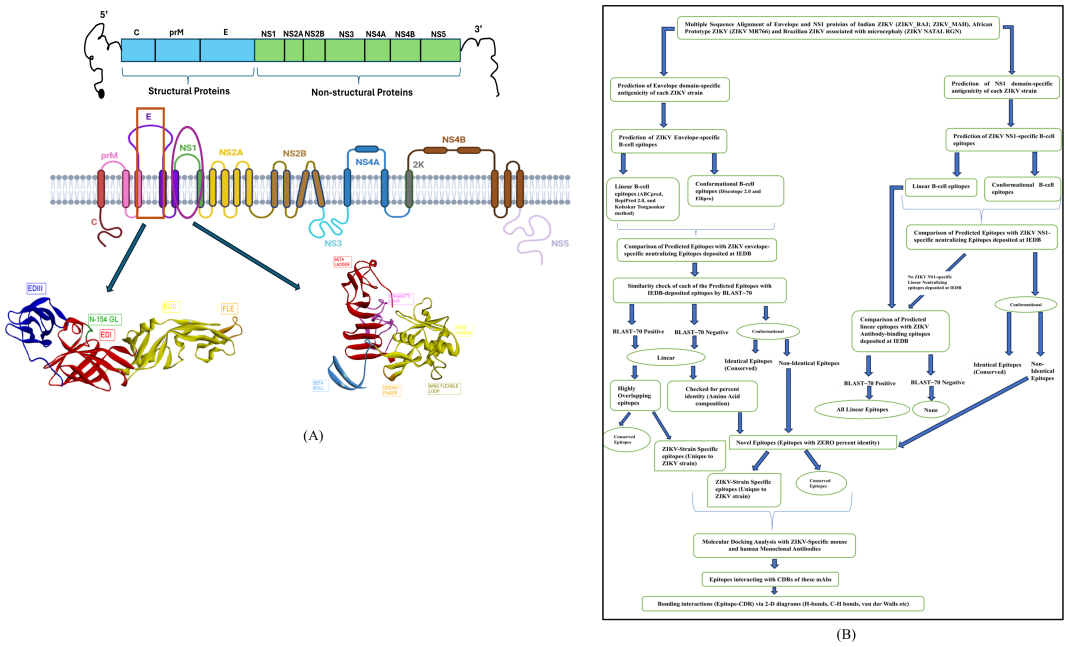


FIGURE 1
Schematic illustration and presentation of the study. **(A)** ZIKV genome highlighting the E and NS1 proteins, **(B)** Study design criteria.

ZIKV strains. The EDIII (296–403) had varying antigenicity scores for two ZIKV strains (ZIKV MR766 and ZIKV_MAH) and identical scores for the other two ZIKV strains (ZIKV_RAJ and ZIKV NATAL RGN), which corroborates this region as ZIKV type-specific (39).

Supplementary Table S2 summarizes the antigenicity analyses for NS1. The β-Roll domain was antigenic with identical scores for both the Indian ZIKV strains and the ZIKV NATAL RGN, unlike African ZIKV MR766. The wing-domain antigenicity scores varied among the Indian, Brazilian and African ZIKV strains and were found to be below the threshold of 0.4, identifying it as non-antigenic. However, the wing domain of ZIKV NS1 contributes to monoclonal antibody-mediated protection, hence epitopes of this region are of research interest (22). The β-ladder domain had varying antigenicity scores with ZIKV_MAH being the highest, followed by ZIKV NATAL RGN, ZIKV_RAJ and ZIKV MR766.

Overall, our analyses revealed that the domain-specific antigenicity of the E and NS1 proteins of Indian ZIKV was differential as well as identical with the African and Brazilian ZIKV strains. Moreover, similarities and differences in the antigenic domains were also observed between the Indian strains ZIKV_RAJ and ZIKV_MAH.

3.3 3D structure prediction of Indian ZIKV E and NS1 proteins

The 3-D structure ZIKV E protein ectodomain ZIKV_RAJ (1–403) and ZIKV_MAH (1–399) were predicted with the crystal structures of ZIKV E protein as templates: 7YW8.pdb (Supplementary Figure S3A) and 7YW7.pdb (Supplementary Figure S4B), respectively. Similarly, the templates for ZIKV MR766 and ZIKV NATAL RGN were identified as 7YW7.pdb and 7YW8.pdb, respectively. The best models were selected based on Ramachandran plot analysis: ZRE (ZIKV_RAJ; Figure 2A) and ZME (ZIKV_MAH; Figure 2B). The occupancy of amino acids in the most favorable and additionally allowed regions for ZRE (Figure 2C) was 99.7% (94.5% + 5.2%) and for ZME (Figure 2D) was 100% (93.1% and 6.1%). There was not a single amino acid occupying the disallowed regions in both the predicted structures. One striking difference was the presence of the N-154 (154-NDTG-157) glycan loop in ZRE (Figure 2A) and its deletion in ZME (Figure 2B). The N-linked glycosylation was also evident in ZIKV NATAL RGN and absent in ZIKV MR766. These results indicated differences in the E-protein structures of both the Indian ZIKV strains.

The 3D structures of ZIKV NS1 (Indian, African and Brazilian strains) were predicted using the known structure of ZIKV NS1 (5K6K.pdb; Supplementary Figure S4) (25). Figure 3 shows the predicted structures of NS1 for the Indian strains: ZRNS1 (ZIKV_RAJ; Figure 3A) and ZMNS1 (ZIKV_MAH; Figure 3B). ZRNS1 and ZMNS1 both had zero per cent residues in disallowed regions. The occupancy of amino acids in the most favored and allowed regions was found to be 100% (93.1% + 6.9%) and 99.7% (93.1% + 6.6%) for ZRNS1 (Figure 3C) and ZMNS1 (Figure 3D).

3.4 Prediction of ZIKV E-domain specific linear epitopes

The prediction of E-domain-specific epitopes was done with a combination of three immunoinformatic tools such as ABCpred,

10.3389/fimmu.2025.1534737

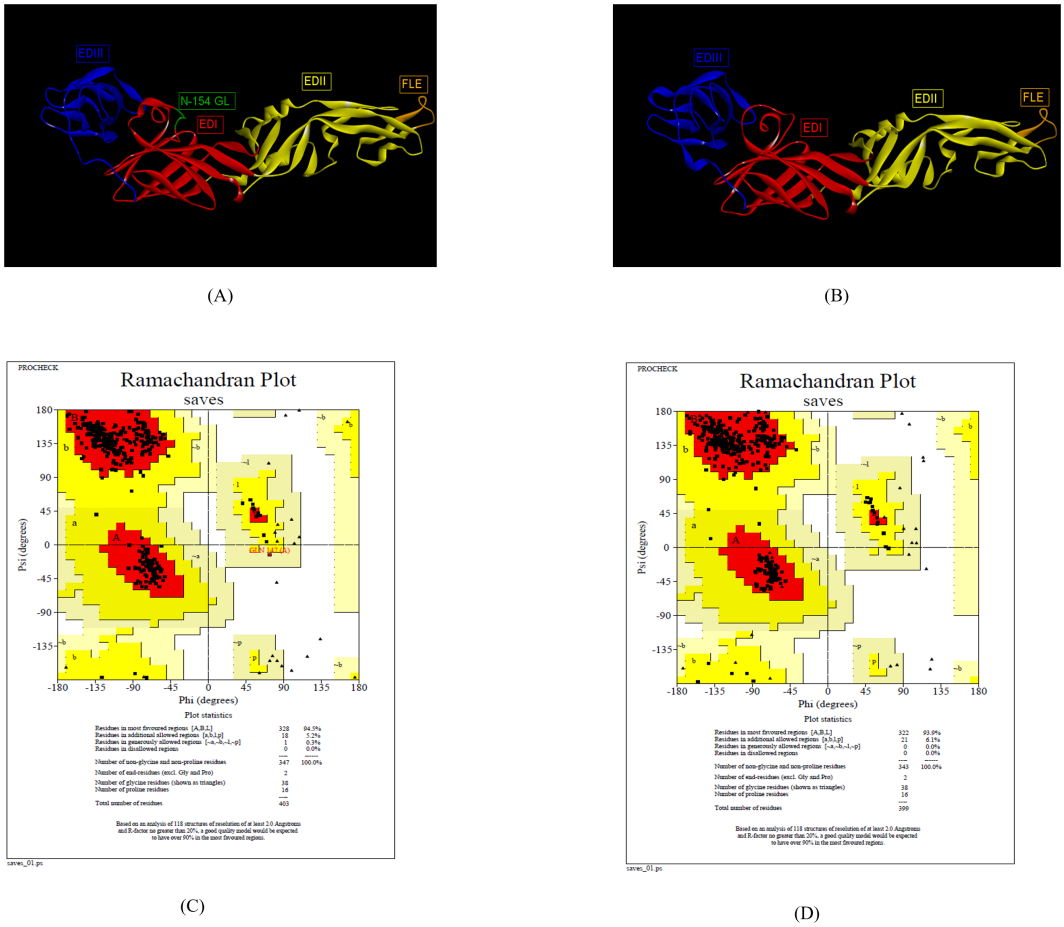


FIGURE 2
Reliable 3-D models for both the Indian ZIKV E and their validation. **(A)** ZRE (Predicted model for ZIKV_RAJ E protein) along with its Dope score **(B)** ZME (Predicted model for ZIKV_MAH E protein) along with its Dope score, **(C)** Ramachandran plot results of ZRE and **(D)** ZME.

BepiPred 2.0 and Kolaskar-Tongaonkar methods. Twenty-eight and twenty-five epitopes were predicted for ZIKV_RAJ and ZIKV_MAH, respectively via ABCpred with a threshold of 0.8 (Supplementary Table S3). In the case of BepiPred 2.0, 18 epitopes were predicted for ZIKV_RAJ and 22 for ZIKV_MAH at 0.5 as threshold (Supplementary Figure S4, Supplementary Table S4). Furthermore, epitope predictions by Kolaskar and Tongaonkar methods revealed 20 and 19 epitopes for ZIKV_RAJ (threshold= 1.026) and ZIKV_MAH (threshold=1.028), respectively (Supplementary Figure S5, Supplementary Table S5). Exposed and buried epitopes were predicted by Emini surface accessibility methods, where scores above 1.00 were considered positive (Supplementary Figure S6). The hydrophilic stretches were determined by Parker hydrophilicity predictions with thresholds of 1.701 and 1.726 for ZIKV_RAJ and ZIKV_MAH, respectively (Supplementary Figure S7). Overlapping shortlisted epitopes were further scanned for being potentially antigenic, non-toxic and non-allergic which identified 33 epitopes for ZIKV E protein (Table 1). To determine the non-overlapping and overlapping epitopes, the predicted epitopes were compared with the IEDB database for ZIKV-specific B-cell epitopes involved in neutralizing antibody response. It revealed two non-overlapping epitopes 5-GVSNRDFVEGMSGGTW-20 and 32-TVMAQDKPTVDIELVT-47 (Figures 4A, B) which had zero percentage identity with the IEDB neutralizing epitopes. Hence, these were designated as novel epitopes having antigenic scores 0.589 and 0.624, respectively. Both these epitopes were conserved across ZIKV_RAJ and ZIKV_MAH. Moreover, there was no amino acid mutation when these epitopes were compared with ZIKV MR766 and ZIKV Natal RGN. However, the other epitope 165-AKVEVTPNSPRAEATL-180 was present in ZIKV MR766 with an antigenic score of 0.6917 but a single point mutation from valine to isoleucine 165-AKVEITPNSPRAEATL-180 (V169I) in Indian ZIKV (ZIKV_RAJ and ZIKV_MAH) and ZIKV Natal RGN increased its antigenicity to 0.8044. These epitopes 165-AKVEVTPNSPRAEATL-180 and 165-AKVEITPNSPRAEATL-180 had a 12.5% per cent identity with the IEDB neutralizing epitopes and were considered overlapping epitopes.

There were four EDII-specific B-cell epitopes, out of which 61-YEASISDMASDSRCPT-76 and 98-DRGWGNGCGLFGK-110 were antigenic with scores as 0.435 and 0.486, respectively. 98-DRGWGNGCGLFGK-110 was considered a fusion loop epitope (FLE) as it spanned across the fusion loop domain. The other epitopes 197-DFSDLYYLTMNNKHWL-212 and 224-PWHAGADTGTPHWNNKE-240 had high antigenicity scores of 1.1716 and 1.1547, respectively, but the former was found to be an

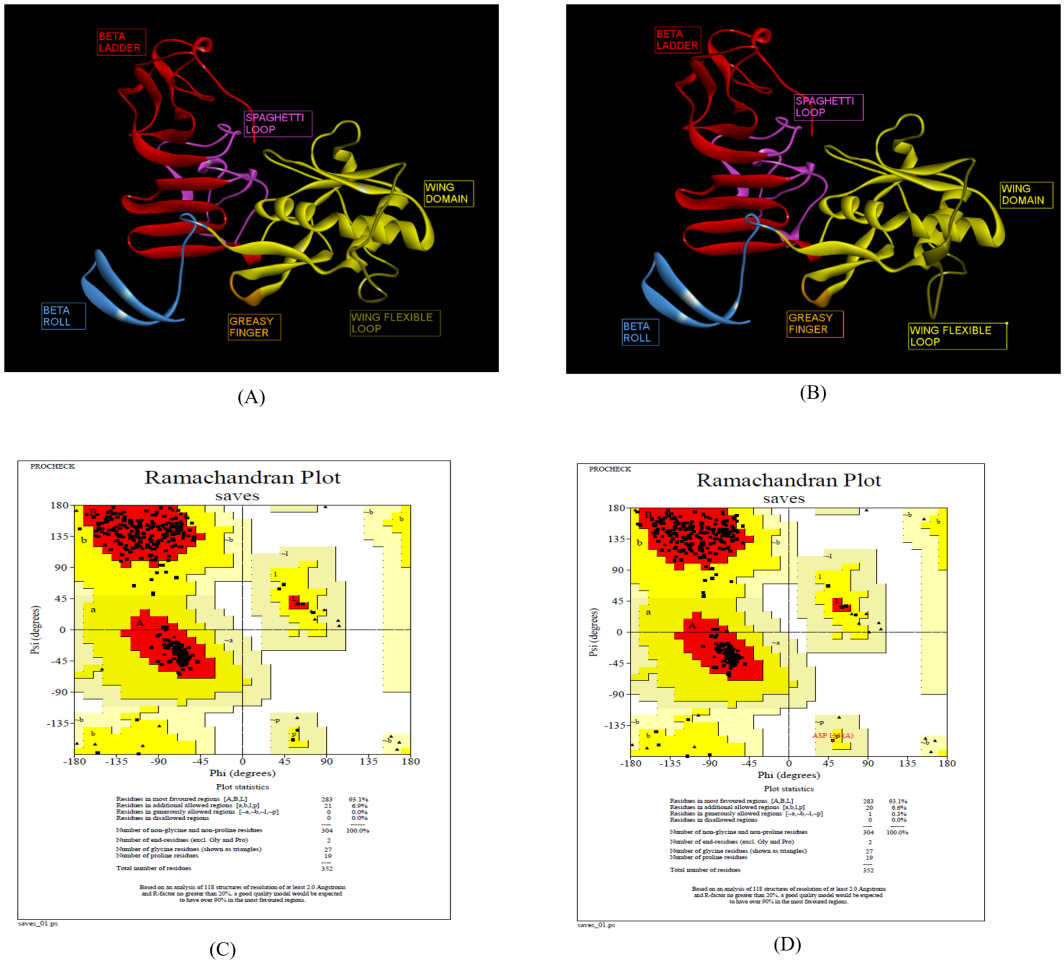


FIGURE 3
Reliable 3-D models for both the Indian ZIKV NS1 and their validation. **(A)** ZRNS1 (Predicted model for ZIKV_RAJ NS1 protein) along with its Dope score, **(B)** ZMNS1 (Predicted model for ZIKV_MAH NS1 protein) along with its Dope score, **(C)** Ramachandran plot results of ZRNS1 and **(D)** ZMNS1.

allergen and was excluded from the analysis. The epitope 224-PWHAGADTGTPHWNNKE-24 (Figures 4C, D) had zero per cent identity with IEDB neutralizing epitopes, hence considered as novel. The epitope 61-YEASISDMASDSRCPT-76 was non-overlapping (BLAST~70% negative), but had 25% per cent identity with the neutralizing epitopes at IEDB, hence was not regarded as novel. The FLE 98-DRGWGNGCGLFGK-110 was BLAST~70% positive with the IEDB neutralizing epitopes and considered as overlapping. Moreover, all the EDII-specific epitopes were conserved across ZIKV_RAJ, ZIKV_MAH, ZIKV MR766 and ZIKV NATAL RGN.

The highest number of ZIKV-specific linear B-cell epitopes were predicted for EDIII. eight epitopes were predicted, harboring point mutations across the ZIKV strains. Out of them, four epitopes, 323-HGTVTVEVQYA-333 (Figure 4E) and 323-HGTVTVEVQYS-333 (Figure 4F) as well as 338-PCKVPAQM-345 (Figure 4E) and 338-PCKIPVQM-345 (Figure 4F) were non-overlapping and had zero per cent identity with IEDB neutralizing epitopes, indicating these as novel. Epitope 323-HGTVTVEVQYA-333 was found in ZIKV_RAJ, ZIKV MR766 and ZIKV NATAL RGN, whereas 323-HGTVTVEVQYS-333 having A333S mutation was found exclusively in ZIKV_MAH. A333S mutation had increased the antigenicity from 1.2198 to 1.2569. Epitope 338-PCKIPVQM-345 was found only in ZIKV MR766 while 338-PCKVPAQM-345 was identified across both the Indian ZIKV (ZIKV_RAJ, ZIKV_MAH) and ZIKV Natal RGN with mutations I341V and V343A leading to increase in antigenicity from 0.450 to 0.478. The remaining epitopes, 363-PVITESTENSK-373 present in ZIKV_RAJ, ZIKV MR766 and ZIKV Natal RGN and 363-PVITESAENSK-373 (ZIKV_MAH), were BLAST~70%-positive and considered overlapping. In contrast, to the novel epitope mutations, the T369A mutation in epitope 363-PVITESAENSK-373 found only in ZIKV_MAH resulted in decreased antigenicity from 0.92 to 0.84. The epitope 384-DSYIVIGVGDKKITHHWHRS-403 in ZIKV_MAH, and 384-DSYIVIGVGEKKITHHWHRS-403 in ZIKV_RAJ were BLAST~70%-negative but had 10% identity with IEDB-neutralizing epitopes. Hence, these were not considered as novel. D393E mutation was also associated with decrease in antigenicity scores from 0.63 to 0.45. Moreover, 393D was found in ZIKV MR766 and 393E in ZIKV NATAL RGN. These EDIII epitopes with point mutations across the ZIKV strains are important for studying type-specific antibody responses to ZIKV infection.

10.3389/fimmu.2025.1534737

TABLE 1 ZIKV E domain-specific linear B-cell epitopes prediction.

Epitope sequence	ZIKV E-specific domain	Antigenicity	Allergenicity	Toxicity
5-GVSNRDFVEGMSGGTW-20	EDI	Antigenic 0.589	Non-allergen	Non-Toxin
32-TVMAQDKPTVDIELVT-47		Antigenic 0.624	Non-allergen	Non-Toxin
165-AKVEVTPNSPRAEATL-180		Antigenic 0.691	Non-allergen	Non-Toxin
165-AKVEITPNSPRAEATL-180		Antigenic 0.804	Non-allergen	Non-Toxin
61-YEASISDMASDSRCPT-76	EDII	Antigenic 0.435	Non-allergen	Non-Toxin
98-DRGWGNGCGLFGK-110		Antigenic 0.486	Non-allergen	Non-Toxin
197-DFSDLYYLTMNNKHWL-212		Antigenic 1.171	Allergen	Non-Toxin
224-PWHAGADTGTPHWNNKE-240		Antigenic 1.154	Non-allergen	Non-Toxin
323-HGTVTVEVQYA-333	EDIII	Antigenic 1.219	Non-allergen	Non-Toxin
323-HGTVTVEVQYS-333		Antigenic 1.256	Non-allergen	Non-Toxin
338-PCKVPAQM-345		Antigenic 0.450	Non-allergen	Non-Toxin
338-PCKIPVQM-345		Antigenic 0.478	Non-allergen	Non-Toxin
363-PVITESTENSK-373		Antigenic 0.634	Non-allergen	Non-Toxin
363-PVITESAENSK-373		Antigenic 0.452	Non-allergen	Non-Toxin
384-DSYIVIGVGDKKITHHWHRS-403		Antigenic 0.920	Non-allergen	Non-Toxin
384-DSYIVIGVGEKKITHHWHRS-403		Antigenic 0.844	Non-allergen	Non-Toxin
43-IELVTTTVSNMAEVRS-58	Hinge-Regions	Antigenic 0.813	Non-allergen	Non-Toxin
185-SLGLDCEPRTGLD-197		Antigenic 1.048	Allergen	Non-Toxin
118-KFTCSKKMTGKSIQPE-133		Antigenic 0.423	Non-allergen	Non-Toxin
118-KFACSKKMTGKSIQPE-133		Non-Antigenic 0.358	Non-allergen	Non-Toxin
126-TGKSIQPENLEYRIMLSV-143		Antigenic 1.088	Non-allergen	Non-Toxin
280-AKGRLSSGHLKCRLKMDK-297		Antigenic 0.488	Non-allergen	Non-Toxin
280-TKGRLSSGHLKCRLKMDK-297		Antigenic 0.566	Non-allergen	Non-Toxin
280-AKGRLFSGHLKCRLKMDK-297		Non-Antigenic 0.026	Non-allergen	Non-Toxin

Mutations of amino acids are colored with respect to the African prototype strain (ZIKV MR766; colored in green and corresponding mutations are colored in red; Novel Epitopes are underlined).

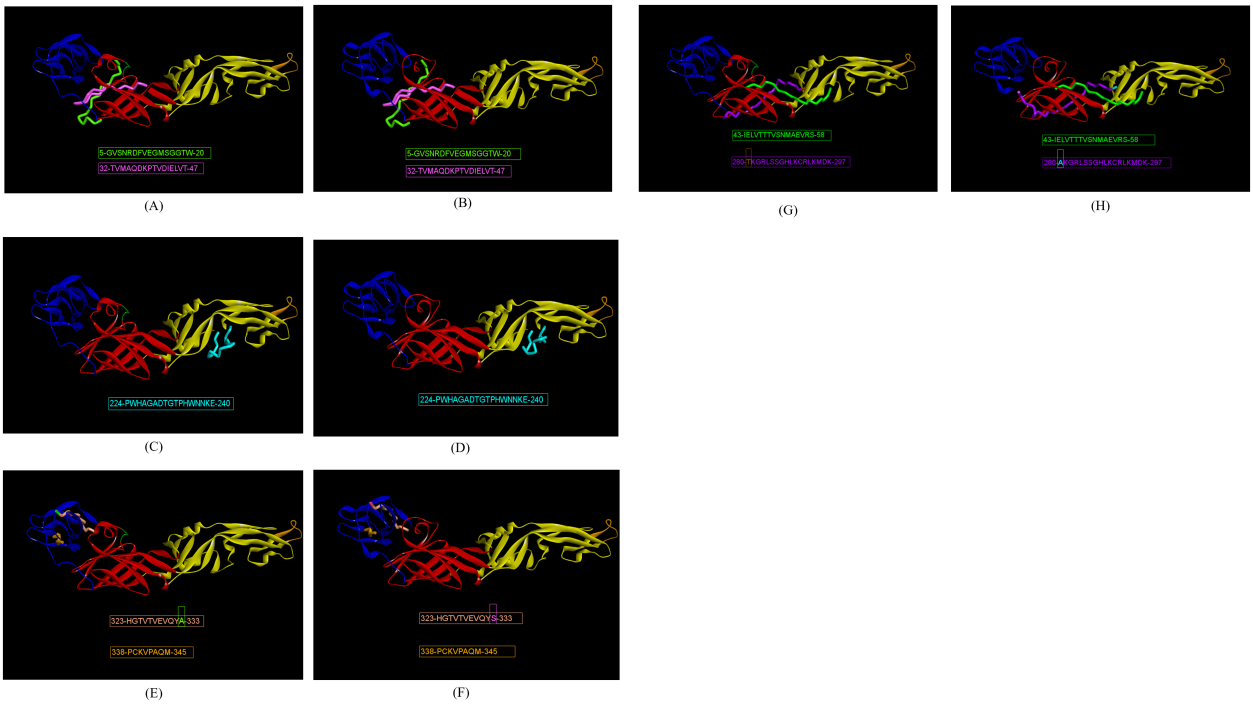


FIGURE 4
Indian ZIKV E domain-specific novel linear epitopes. **(A)** ZRE domain I-specific epitopes, **(B)** ZME domain I-specific epitopes, **(C)** ZRE domain II-specific epitopes, **(D)** ZME domain II-specific epitopes, **(E)** ZRE domain III-specific epitopes, **(F)** ZME domain III-specific epitopes, **(G)** ZRE hinge region-specific epitopes, and **(H)** ZME hinge region-specific specific epitopes.

The predicted B-cell linear epitopes also spanned across the hinge regions of the E protein as these regions are important in potently cross-neutralizing Flavivirus infections, especially ZIKV and DENV infections. A total of eight hinge region epitopes were predicted with epitope 126-TGKSIQPENLEYRIMLSV-143 (conserved in both Indian strains) having the highest antigenicity of 1.0884, followed by epitope 43-IELVTTTVSNMAEVRS-58 (conserved in both Indian strains) with antigenicity of 0.8137, both of which spanning EDI/II hinge and conserved in the Indian strains. Out of these two epitopes, 43-IELVTTTVSNMAEVRS-58 (Figures 4G, H) was BLAST~70% negative and had zero per cent identity with the IEDB neutralizing epitopes, hence regarded as novel. The other epitope 126-TGKSIQPENLEYRIMLSV-143 was BLAST~70% positive, overlapping with IEDB neutralizing epitopes. Epitope 118-KFTCSKKMTGKSIQPE-133 was found in ZIKV MR766 but the T120A mutation in both Indian ZIKV strains and ZIKV NATAL RGN led to a decrease in antigenicity from 0.42 to 0.35. These epitopes were BLAST~70% positive and considered overlapping epitopes. Moving ahead, the epitope 280-AKGRLFSGHLKCRLKMDK-297 in ZIKV MR766 was found to be highly non-antigenic with a score of 0.02, but the A280T and F285S mutations in 280-TKGRLSSGHLKCRLKMDK-297 (ZIKV_RAJ; Figure 4G) reversed its non-antigenicity, making it antigenic with a score of 0.56. Similarly, the F285S mutation in 280-AKGRLSSGHLKCRLKMDK-297 (ZIKV_MAH; Figure 4H) also made it antigenic with a score of 0.48. Both these epitopes, unique to Indian ZIKV strains were BLAST~70%-negative and had zero per cent identity with the IEDB neutralizing epitopes, indicating these as novel EDI/DIII epitopes. The epitope 185-SLGLDCEPRTGLD-197 of the EDI/II hinge region was antigenic but was excluded from the analysis due to its allergenicity. The change in antigenicity based on the mutations among the epitopes of the hinge regions would be essential to studying peptide-specific reactivity to ZIKV and related Flavivirus infections such as DENV.

3.4.1 ZIKV NS1 domain-specific linear B-cell epitopes

The prediction of linear B-cell epitopes for the NS1 protein of ZIKV was carried out in the same manner as that of the E protein. Following the E protein, the ZIKV antibody response is triggered by the NS1 protein. Hence, it is necessary to identify ZIKV NS1 domain-specific epitopes. The immunoinformatic tools for the prediction of linear B-cell epitopes of NS1 were the same as those used for the ZIKV E. ABCpred predicted a total of 24 and 26 epitopes with a threshold of 0.8 for ZIKV_RAJ and ZIKV_MAH, respectively (Supplementary Table S6). BepiPred 2.0 predicted 11 epitopes for each Indian ZIKV (ZIKV_RAJ and ZIKV_MAH) with a threshold of 0.5 (Supplementary Table S7, Supplementary Figure S7). Kolaskar and Tongaonkar method predicted 16 epitopes for each Indian ZIKV (ZIKV_RAJ and ZIKV_MAH) with a threshold of 1.018 and 1.023 for ZIKV_RAJ and ZIKV_MAH, respectively (Supplementary Table S8, Supplementary Figure S8). Similar to the E protein, these epitopes were checked for accessibility and hydrophilicity by Emini surface accessibility and Parker hydrophilicity, respectively (Supplementary Figures S9, S10). In total, sixteen epitopes were predicted for ZIKV occupying different NS1-specific domains (Table 2). Among them, nine epitopes were found in Indian ZIKV strains. Most of the epitopes predicted possessed point mutations among them and were ZIKV strain-specific which may be interesting to study via *in vitro* and *in vivo* peptide validation experiments.

The epitope 10-KKETRCGTGVFVYNDVE-26 in the beta roll domain of NS1 was conserved across both the Indian ZIKV strains (ZIKV_RAJ and ZIKV_MAH). This epitope was also conserved in the ZIKV Natal RGN. However, the ZIKV MR766 strain had I21 instead of V21. This I21V mutation in the rest of the strains resulted in increased antigenicity from 0.78 to 0.87.

The exposed surfaces of NS1 are composed of the wing domain, especially the residues except for the flexible loop, greasy finger. Considering this, five epitopes were predicted for the wing domain. All these linear epitopes were associated with point mutations. The

TABLE 2 ZIKV NS1 domain-specific linear B-cell epitopes prediction.

Epitope sequence	ZIKV NS1-specific domain	Antigenicity	Allergenicity	Toxicity
10-KKETRCGTGVFIYNDVE-26	β-Roll	Antigenic 0.780	Non-allergen	Non-Toxin
10-KKETRCGTGVFVYNDVE-26		Antigenic 0.875	Non-allergen	Non-Toxin
83-GVQLTVVVGSVKNP-96	Wing Domain	Antigenic 0.748	Allergen	Non-Toxin
83-GIQLTVVVGSVKNP-96		Antigenic 0.820	Non-allergen	Non-Toxin
141- ECPLEHRAWNSFLVED-157		Antigenic 0.587	Non-allergen	Toxin
141- ECPLKHRAWNSFLVED-157		Antigenic 0.505	Non-allergen	Non-Toxin
141- ECPLKHRAWNSFIVED-157		Antigenic 0.487	Non-allergen	Non-Toxin
191- REAAHSDLGYWIESEKND-208	β-Ladder	Antigenic 0.430	Non-allergen	Non-Toxin
191- KEAVHSDLGYWIESEKND-208		Antigenic 0.572	Allergen	Non-Toxin
248-AGPLSHHNTREGYRTQV-264		Antigenic 0.713	Non-allergen	Non-Toxin
248-AGPLSHHNTREGYRTQM-264		Antigenic 0.802	Non-allergen	Non-Toxin
331-YGMEIRPRKEPESNLVRSMV-350		Antigenic 0.623	Non-allergen	Non-Toxin
331-YGMEIRPRKEPESNLVRSVV-350		Antigenic 0.680	Non-allergen	Non-Toxin
176-SLECDPAVIGTAVKGREAA-194	Wing-β-Ladder	Antigenic 1.173	Non-allergen	Non-Toxin
176-SLECDPAVIGTAVKGKEAV-194		Antigenic 1.093	Non-allergen	Non-Toxin
176-SLECDPAVIGTAIKGKEAV-194		Antigenic 1.087	Non-allergen	Non-Toxin

Mutations of amino acids are colored with respect to the African prototype strain (ZIKV MR766; colored in green and corresponding mutations are colored in red).

epitope 83-GVQLTVVVGSVKNP-96 was specific to both African ZIKV (ZIKV MR766) and Brazilian ZIKV (ZIKV Natal RGN) but V84I mutation in the same epitope 83-GIQLTVVVGSVKNP-96 was specific to both the Indian ZIKV strains (ZIKV_RAJ and ZIKV_MAH). This epitope in African and Brazilian strains was predicted as an allergen, but the V84I mutation in the Indian strains made the epitope non-allergen (by AllerTop). This mutation also increased the antigenicity from 0.74 to 0.82. Similarly, another set of predicted epitopes as 141-ECPLEHRAWNSFLVED-157 was specific to ZIKV MR766 and ZIKV Natal RGN. Point mutations were seen in both the Indian ZIKV strains such as E145K and L153I (ZIKV_RAJ) and only L153I (ZIKV_MAH). The E145K mutation in ZIKV_MAH was associated with decreased antigenicity from 0.58 to 0.50 while both E145K and L153I mutations in ZIKV_RAJ led to a further decrease in antigenicity to 0.48. Interestingly, the epitope 141-ECPLEHRAWNSFLVED-157 specific to ZIKV MR766 and ZIKV Natal RGN was predicted to be toxic (ToxinPred), whereas in the Indian strains, the epitopes 141-ECPLKHRAWNSFIVED-157 (ZIKV_RAJ) and 141-ECPLKHRAWNSFIVED-157 were found to be non-toxic.

The beta ladder is another domain of NS1 which is a target for ZIKV NS1-specific neutralizing mAbs as it is mostly exposed, especially the spaghetti loop residues. The highest number of predicted epitopes belonged to this domain. The epitope 191-REAAHSDLGYWIESEKND-208 in ZIKV MR766 possessed mutations: R191K and A194V in both Indian strains. This resulted in an increase in antigenicity from 0.43 to 0.57. Though there was an increase in antigenicity, the epitope 191-KEAVHSDLGYWIESEKND-208 was found to be an allergen (by AllerTop) and was not considered for further analysis. Another epitope 248-AGPLSHHNTREGYRTQV-264 was conserved in ZIKV_MAH and ZIKV_MR766 while the epitope 248-AGPLSHHNTREGYRTQM-264 was conserved in ZIKV_RAJ and ZIKV NATAL RGN. This V264M mutation in ZIKV_RAJ was associated with an increase in antigenicity from 0.71 to 0.80. The epitope 331-YGMEIRPRKEPESNLVRSMV-350 was conserved in ZIKV MR766 and ZIKV RAJ whereas the M349V mutation observed in ZIKV_MAH and ZIKV Natal RGN was associated with an increase in antigenicity from 0.62 to 0.68.

Moving forward, the epitope 176-SLECDPAVIGTAVKGVEAA-194 present in ZIKV MR766 was highly antigenic (antigenicity=1.17; VaxiJen) This was located as part of the connector residues linking the wing domain to the beta ladder domain. Mutations V188I, R191K, and A194V were observed in the ZIKV_MAH strain while mutations R191K and A194V were observed in ZIKV_RAJ with respect to ZIKV MR766. Epitope 176-SLECDPAVIGTAVKGKEAV-194 in ZIKV_RAJ and ZIKV Natal RGN had an antigenicity of 1.09 and 176- SLECDPAVIGTAIKGKEAV-194 in ZIKV_MAH had an antigenicity of 1.08.

According to our study design, the IEDB ZIKV-neutralizing epitopes database did not contain any linear B-cell ZIKV NS1-specific epitope. Comparing the predicted ZIKV NS1-specific B-cell linear epitopes with the IEDB ZIKV-antibody binding epitopes revealed that all these B-cell epitopes were BLAST~70% positive. This suggests that all the predicted linear epitopes of the ZIKV NS1 overlap with linear ZIKV NS1-specific antibody-binding epitopes. However, these epitopes may be potentially validated by *in vitro* neutralization assays to determine the antibody response against ZIKV NS1.

3.5 ZIKV E domain-specific conformational B-cell epitopes

The humoral immune response to ZIKV infection mostly depends on discontinuous or conformational B-cell epitopes. Alongside linear B-cell epitopes of the ZIKV E, identification of conformational B-cell epitopes is also necessary to evaluate neutralizing antibody responses. The predictions were made by two immunoinformatic tools available at IEDB; Discotope 2.0 and ElliPro. A total of 19 and 14 epitopes were identified by Discotope 2.0 for ZIKV_RAJ and ZIKV_MAH respectively. Ellipro predictions identified 28 and 24 epitopes for ZIKV_RAJ and ZIKV_MAH, respectively. The epitopes common in both prediction tools for ZIKV_RAJ were identified as G383, D384, H401 and S403 (Supplementary Table S9). Out of these, D384 was also found in the IEDB list of ZIKV E-specific neutralizing conformational epitopes, whereas the other three epitopes, G383, H401 and S403 were found to be unique, hence considered as novel epitopes for ZIKV_RAJ (Figure 5). All these three epitopes were located in EDIII. Similarly, in the case of ZIKV_MAH, epitope W101 of EDI (Supplementary Table S9) was found to be the common epitope from both the prediction tools and was also found in the IEDB list of ZIKV E-specific neutralizing conformational epitopes, hence was not considered as novel (Figure 5). The presence of different conformational epitopes across the ZIKV E domains and the identification of the above-mentioned novel epitopes is essential in understanding the neutralizing antibody response associated with ZIKV-specific mAbs identifying these epitopes.

3.6 ZIKV NS1 domain-specific conformational B-cell epitopes

Immunoinformatics tools Discotope 2.0 and ElliPro were used for the predictions. A total of 37 and 19 conformational epitopes

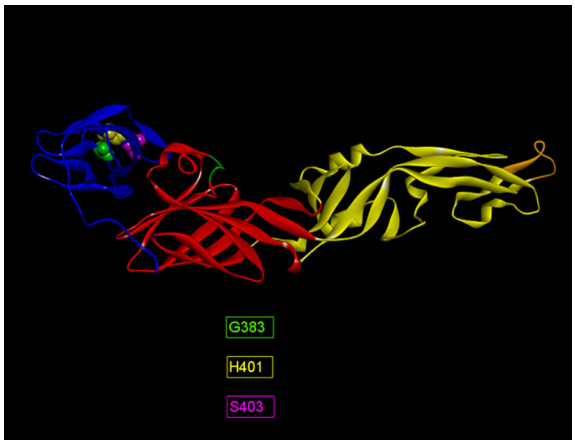


FIGURE 5
Novel ZIKV E-Specific B-cell conformational epitopes (ZRE).

were predicted for ZIKV_RAJ and ZIKV_MAH, respectively with the Discotope 2.0 whereas 75 and 68 epitopes were predicted for ZIKV_RAJ and ZIKV_MAH, respectively by ElliPro. Fourteen epitopes for ZIKV_RAJ and 6 for ZIKV_MAH, common in both prediction tools, were considered for further analysis. (Supplementary Table S10). Further comparison of these epitopes with the existing IEDB database of conformational B-cell neutralizing epitopes of ZIKV NS1, revealed the identification of novel epitopes as F8, S9, K10, K11, K116, A117, W118, G119, K120, P341 and S343 specific to ZIKV_RAJ (Figure 6A). Moreover, F8, S9, K10, K11, and S343 were considered as novel epitopes specific to ZIKV_MAH (Figure 6B).

3.7 Molecular docking analysis of ZIKV E-specific novel linear epitopes

In the case of ZIKV E protein, there were 7 novel linear B-cell epitopes and 3 novel conformational B-cell epitopes specific to ZIKV_RAJ. Similarly, there were 7 novel linear B-cell epitopes specific to ZIKV_MAH, however, there were no novel conformational B-cell epitopes specific to ZIKV_MAH. The molecular docking of the ZIKV E protein was carried out with two highly potent neutralizing ZIKV-specific mAbs, ZV-67 and Z3L1. ZV-67 is a mouse mAb that has potent neutralization against both African and Asian ZIKV strains, whereas Z3L1 is a human mAb that has effective neutralization against both African and Asian strains. Moreover, both these mAbs neutralized ZIKV *in vitro* and *in vivo* (21, 41). Molecular docking of the E protein using HDock for Indian ZIKV strains with these mAbs revealed that out of the 7 novel linear B-cell epitopes, 4 epitopes specific to both ZIKV_RAJ and ZIKV_MAH showed intermolecular interactions with the complement determining regions (CDRs) of the mAb ZV-67 (Table 3). Of these 4 epitopes, one was located in EDIII, one each belonged to EDI and EDII, and one belonged to the hinge region (DI/DIII:280-295). The epitopes of EDIII and hinge regions had point mutations between both the Indian ZIKV strains whereas the other two epitopes belonging to EDI and EDII were conserved between both ZIKV_RAJ and ZIKV_MAH. However, there was only one novel EDII epitope which interacted with the CDRs of the mAb Z3L1. The inter-molecular bonding interactions between the epitope and CDR region residues were analyzed by 2-D interaction maps.

In the case of EDIII, epitope 323-HGTVTVEVQYA-333, unique to ZIKV_RAJ (Figure 7A) and epitope 319-HGTVTVEVQYS-329 unique to ZIKV_MAH (Readers should note that due to a deletion of 4-amino acids in ZIKV_MAH strain at N154, the nomenclature is: amino acid 319 for ZIKV_MAH is equivalent to 323 for ZIKV_RAJ which applies to all the amino acids post 154th position in ZIKV_MAH) (Figure 7B), the epitope residues 323-HG-324 (ZIKV_RAJ; Figure 7C) and 319-HG-323 (ZIKV_MAH; Figure 7D) interacted with CDRs of VH and VL regions of mAb ZV-67. The 2-D interaction maps for these epitopes revealed intermolecular hydrogen bonding interactions with CDRs (Table 3) of ZV-67, wherein the epitope-CDR3(VH) interactions were found to be the most suitable. In the case of the epitope unique to ZIKV_RAJ (323-HGTVTVEVQYA-333), the 2-D interaction maps showed that 323H formed one carbon-hydrogen (C-H) bond and one pi-donor hydrogen bond and 324G formed van der Walls interaction with CDR3-VH residues, respectively (Figure 7E). However, in the case of the epitope unique to ZIKV_MAH (319-HGTVTVEVQYS-329), the results were slightly different with 319H forming one conventional hydrogen (H) bond and 320G forming one carbon-hydrogen (C-H) bond with CDR3-VH of ZV-67, respectively (Figure 7F). The mAb ZV-67 is known to bind to the ZIKV-E protein in the EDIII domain (21).

In addition, the footprint of mAb ZV-67 also covered parts of EDI and EDII domains and interactions with novel epitopes were observed. Out of the two novel EDI epitopes, epitope 5-GVSNRDFVEGMSGGTW-20 interacted with CDR regions of the mAb ZV-67. This epitope was conserved in both the Indian ZIKV strains. The most suitable docking interactions shown by this EDI

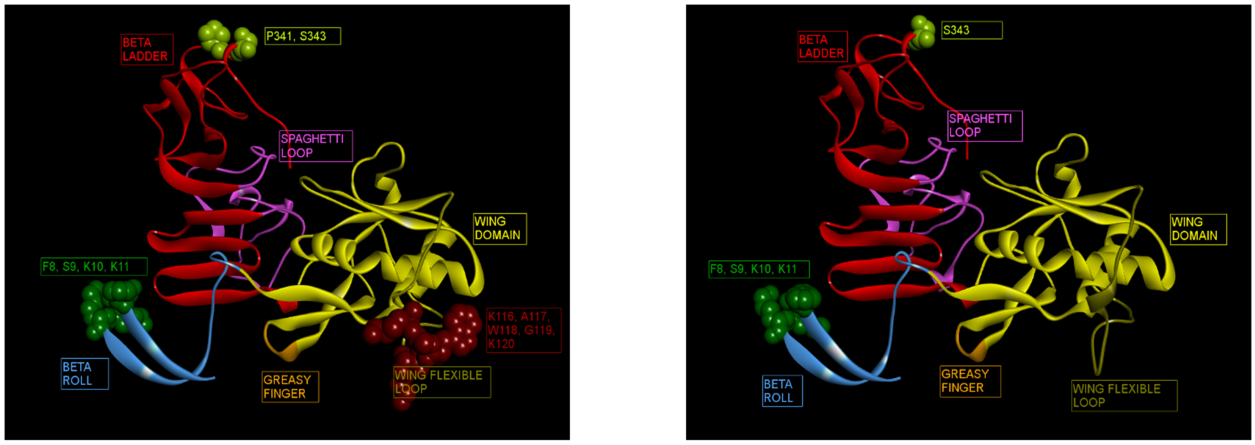


FIGURE 6
Novel ZIKV NS1 domain-specific B-cell conformational epitopes. **(A)** ZIKV_RAJ NS1 domain-specific B-cell conformational epitopes, **(B)** ZIKV_MAH NS1 domain-specific B cell conformational epitopes.

TABLE 3 H-Dock based molecular docking analysis of ZIKV-specific novel linear epitopes of the E-protein.

Novel Linear Epitopes	CDR Interactions	Indian ZIKV-specificity	Epitope/ CDR Interactions	2-D Interaction maps	Docking/ Binding scores
323-HGTVTVEVQYA-333A (EDIII)	Yes	ZIKV_RAJ (Unique)	323-HG-324/ CDR3-VH	323H: 1 C-H bond and 1 pi-donor H-bond 324G: van der Walls	-306.07
319-HGTVTVEVQYS-329 (EDIII)	Yes	ZIKV_MAH (Unique)	319-HG-320 CDR3-VH	319G: 1 H-bond and 320G: 1 C-H bond	-291.28
338-PCKVPAQM-345 (EDIII)	No	ZIKV_RAJ and ZIKV_MAH	N.A.	N.A.	
5-GVSNRDFVEGMSGGTW-20 (EDI)	Yes	ZIKV_RAJ and	10-DFVE-13/ CDR1-VL (ZIKV_RAJ)	13E: 1 H-bond 12V: 1 C-H bond 10D & 11F: van der Walls 10D: 1 H-bond	-285.93
		ZIKV_MAH	7S; 9-RD-10/CDR3-VL (ZIKV_MAH)	7S: 1 C-H bond and 1 pi-donor H-bond, 9R: van der Walls	-282.31
32-TVMAQDKPTVDIELVT-47 (EDI)	No	ZIKV_RAJ and	N.A	N.A.	
		ZIKV_MAH	N.A.	N.A.	
224-PWHAGADTGTPHWNNKE-240; (EDII)	Yes	ZIKV_RAJ and	235H; 237-NN-238; 239-KE-240/CDR3-VL (ZIKV_RAJ)	235H: 1 H-bond 237N: 1 H-bond 238N: 1 H-bond 239K: van der Walls	-276.14
220-PWHAGADTGTPHWNNKE-236 (EDII)		ZIKV_MAH	233-NN-234; 231H, 235K/CDR3-VL (ZIKV_MAH)	233N: 1 H-bond 234N: 1 H-bond 231H: van der Walls 235K: van der Walls	-277.38
43-IELVTTTVSNMAEVRS-58 (Hinge Region)	No	ZIKV_RAJ and ZIKV_MAH	N.A	N.A	
280-TKGRLSSGHLKCRLKMDK-297 (Hinge Region)	Yes	ZIKV_RAJ	288H;290K/CDR1-VL	288H: van der Walls 290K: van der Walls	-286.01
276-AKGRLSSGHLKCRLKMDK-293 (Hinge Region)	Yes	ZIKV_MAH	284H/CDR3-VH	284H: van der Walls	-283.06

Molecular Docking interactions of the novel epitopes with potent mouse monoclonal antibody ZV67; N.A. (Not Available).
Amino acid mutations are coloured in Red with respect to ZIKV Prototype strain MR766 (Green).

epitope were with CDR1-VL in the case of ZIKV_RAJ and CDR3-VL in the case of ZIKV_MAH, respectively (Figures 8A–D). The 2-D interaction maps showed that for the ZIKV_RAJ EDI, the amino acids 13E and 12V showed one conventional hydrogen (H)-bonding interaction and one carbon-hydrogen (C-H) bonding interaction, respectively with CDR1-VL residues (Figure 8E) Amino acids 10D and 11F displayed van der Walls interactions (Figure 8E). However, for ZIKV_MAH EDI, amino acids 10D showed one conventional hydrogen (H)-bond, and 7S displayed one carbon-hydrogen (C-H) bond and one pi-donor hydrogen bond with CDR3-VL residues (Figure 8F). In this case, amino acid 9R showed van der Walls interactions (Figure 8F). Further, the novel EDII epitope 224-PWHAGADTGTPHWNNKE-240 which was conserved in both ZIKV_RAJ and ZIKV_MAH, interacted with CDR3-VL of the mAb (Figures 9A–D). For ZIKV_RAJ EDII, the amino acids 235H, 237N, and 238N had 1 conventional H-bond each and amino acids 239K and 240E had van der Walls interaction. (Figure 9E). However, in the case of ZIKV_MAH EDII (220-PWHAGADTGTPHWNNKE-236) the amino acids 233N and 234N had 1 conventional H-bond each, with 234N also displaying an unfavorable interaction. Further, 231H and 235K possessed van der Walls interaction (Figure 9F). There were two novel epitopes identified in the hinge regions of Indian ZIKV strains, among which the epitope 280-TKGRLSSGHLKCRLKMDK-297 (ZIKV_RAJ) and

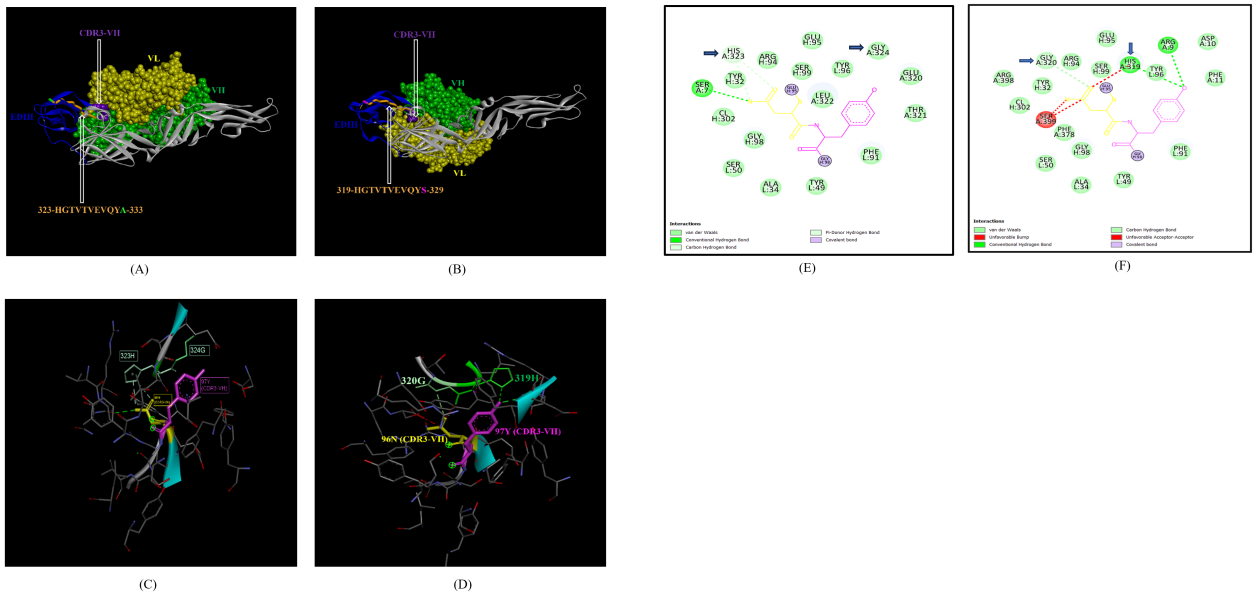


FIGURE 7
Molecular docking of novel linear EDIII epitopes with CDRs of ZV-67 mAb. **(A)** 3-D model of docked ZRE and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VH (coloured in purple) interacting with EDIII (coloured in blue) epitope 323-HGTVTVEVQYA-333 (coloured in light brown) except alanine (coloured in light green). **(B)** 3-D model of docked ZME and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VH (coloured in purple) interacting with EDIII (coloured in blue) epitope 319-HGTVTVEVQYS (coloured in teal orange) except serine (coloured in blue). **(C)** Epitope-CDR interaction in 3-D for ZRE with CDR3-VH of ZV67; 323-HG-324 (coloured in shades of light green) are the amino acids of EDIII epitope interacting with CDR3-VH amino acid residues, 96-NY-97 (coloured in pink and yellow, respectively). **(D)** Epitope-CDR interaction in 3-D for ZME with CDR3-VH of ZV67; 319-HG-320 (coloured in shades of light green) are the amino acids of EDIII epitope interacting with CDR3-VH amino acid residues, 96-NY-97 (coloured in pink and yellow, respectively). **(E)** 2-D interaction map of panel **(C)**; blue arrows indicating 323H forming one carbon-hydrogen and one pi-donor hydrogen bond and 324G forming van der Walls interactions with 96N and 97Y. **(F)** 2-D interaction map of panel **(D)**; blue arrows indicating 319H forming conventional hydrogen bond and 320G forming carbon-hydrogen bond with 96N and 97Y.

epitope 276-AKGRLSSGHLKCRLKMDK-293 (ZIKV_MAH) had interactions with the CDRs of ZV-67, respectively (Figures 10A–D) The 2-D interaction maps of 280-TKGRLSSGHLKCRLKMDK-297 showed that amino acids 288H and 290K had van der Walls interaction with CDR1-VL of the mAb (Figure 10E); whereas epitope 276-AKGRLSSGHLKCRLKMDK-293 had only 284H having van der Walls interaction with CDR3-VL (Figure 10F).

Next, we sought to identify the epitope-CDR interaction of the Indian ZIKV strains with human mAb Z3L1. The novel epitope of EDII 224-PWHAGADTGTPHWNNKE-240, conserved in both ZIKV_RAJ and ZIKV_MAH, was the only epitope showing molecular interactions with the CDRs (Table 4) of Z3L1. However, novel epitopes in other E-domains also interacted with Z3L1, but these interactions were not with CDR regions. On the other hand, this EDII epitope interacted with CDR1-VH of Z3L1 in the case of both the Indian ZIKV strains (Table 4; Figures 11A, B). This displayed different epitope-CDR bonding interactions specific for ZIKV_RAJ (Figures 11A, C) and ZIKV_MAH (Figures 11B, D) which were demonstrated by their 2-D interaction maps. The epitope 224-PWHAGADTGTPHWNNKE-240 had amino acids 231T, 232G and 233A forming 1 H-bond, 1 C-H bond and van der Walls interaction, respectively which were specific to ZIKV_RAJ (Figure 11E) whereas the same epitope (220-PWHAGADTGTPHWNNKE-236) had amino acids 223A forming 1 H-bond and 1 pi-alkyl bond and 224G forming van der Walls interaction in the case of ZIKV_MAH (Figure 11F)

4 Discussion

In the case of emerging Flavivirus infection, such as the Zika virus, E and NS1 proteins play a vital role in eliciting robust humoral immunity (19, 26). The ZIKV-specific B-cell epitopes are crucial in inducing the humoral immune response (15, 42, 43). The advent of immunoinformatics has led to the prediction of various B-cell epitopes in Flaviviruses' antigens. However, a similar broad-spectrum analysis of the ZIKV-specific B-cell epitopes is required and remains poorly understood. In-silico prediction and validation of these epitopes may help in ZIKV therapeutics and vaccine design.

In this study, we identified B cell epitopes of E and NS1 proteins of two co-circulating Indian ZIKV strains. Further, these epitopes were mapped to their modelled 3-D structures leading to the identification of novel epitopes based on the stringent study design criteria. Molecular docking with potent ZIKV-neutralizing mAbs validated our findings via epitope-CDR interactions. The identification of these novel epitopes was specific and unique to Indian ZIKV strains. These epitopes have the potential to induce peptide-specific antibodies, mostly involved in neutralizing response, especially against the ZIKV E protein.

In-silico antigenic characterization of viral proteins is essential for epitope identification. The E protein of ZIKV plays an important role in virus entry, attachment, and fusion. Apart from these functions, the ZIKV E protein is the major target of neutralizing antibodies. It consists of three domains: the central beta-barrel domain (EDI), an elongated finger-like domain (EDII) consisting of

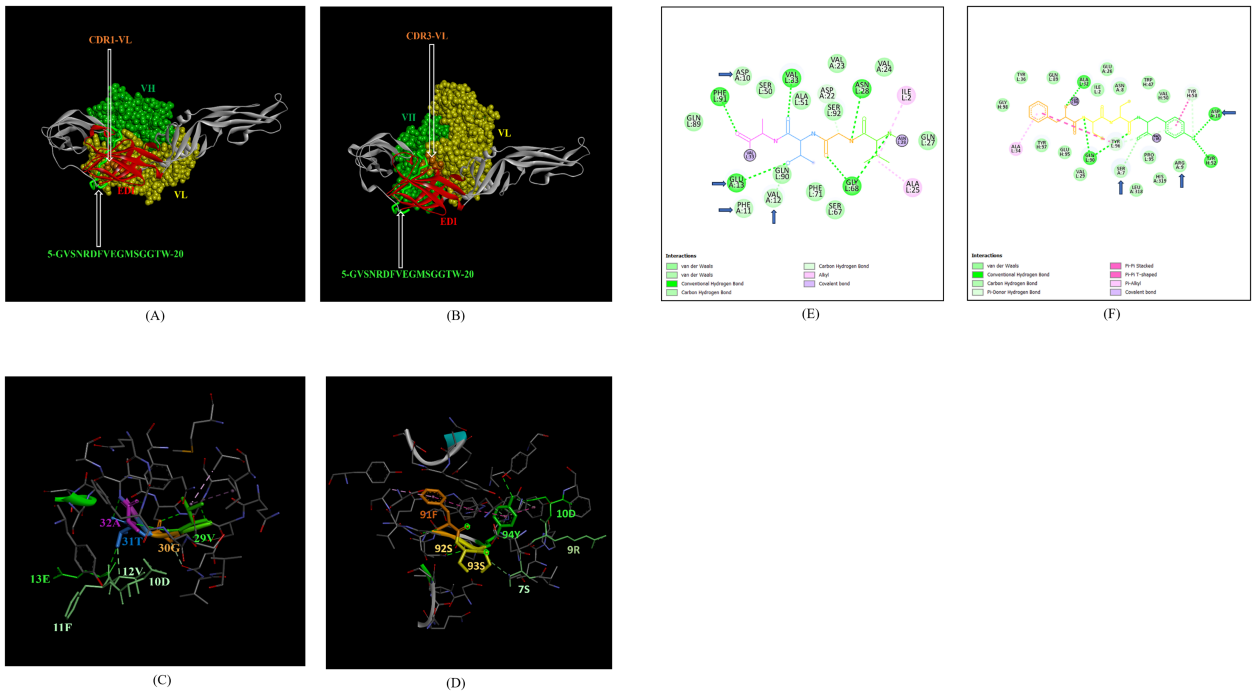


FIGURE 8
Molecular docking of novel linear EDI epitopes with CDRs of ZV-67 mAb: **(A)** 3-D model of docked ZRE and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR-1-VL (coloured in dark orange) interacting with EDI (coloured in red) epitope 5-GVSNRDFVEGMSGGTW-20 (coloured in light green). **(B)** 3-D model of docked ZME and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VL (coloured in dark orange) interacting with EDI (coloured in red) epitope 5-GVSNRDFVEGMSGGTW-0 (coloured in light green). **(C)** Epitope-CDR interaction in 3-D for ZRE with CDR1-VL of ZV67; 10-DFVE-13 (coloured in shades of light green) are the amino acids of EDI epitope interacting with CDR1-VH amino acid residues, 29-VGTA-31 (coloured in light green, orange, purple and light blue, respectively). **(D)** Epitope-CDR interaction in 3-D for ZME with CDR3-VL of ZV67; 7S, 9-RD-10 (coloured in shades of light green) are the amino acids of EDI epitope interacting with CDR3-VL amino acid residues, 91-FSSY (coloured in dark orange, yellow, yellow and light green, respectively). **(E)** 2-D interaction map of panel **(C)**; blue arrows indicating 13E forming one conventional hydrogen bond, 12E forming one carbon-hydrogen bond and 10-DF-11 forming van der Walls interactions with 31T, 32A, 30G and 29Y, respectively. **(F)** 2-D interaction map of panel **(D)**; blue arrows indicating 10D forming one conventional hydrogen bond, 7S forming one carbon-hydrogen bond, and 9R forming van der Walls interactions with 94Y, 93S,92S and 91Y.

a hydrophobic fusion loop (FL) and an IgC-like immunoglobulin domain (EDIII) (21, 26, 40). Moreover, these domains display differential neutralizing potential, with EDIII being the most potent and ZIKV-type specific. Earlier studies reported antigenicity analyses for the whole E protein of the Zika virus (40). However, our study involves domain-wise estimation of antigenicity for ZIKV E protein for different strains, which provides detailed insight into understanding the antigenic characteristics: similarities and differences between the strains. Following domain-specific antigenicity analysis, we predicted linear and conformational B-cell epitopes specific to ZIKV E protein as the overall humoral immune response depends on both the primary and tertiary structures of these epitopes (44, 45). The prediction of the linear and conformational epitopes was carried out using a combination of immunoinformatic tools to minimize false positive results and the amino acid regions predicted as epitopes by all the different tools were considered for further analyses (27, 33, 46). Moreover, the length of linear B-cell epitopes was also considered as it is a significant parameter in the case of designing peptide vaccines capable of inducing substantial humoral immune response (47). The experimentally validated ZIKV B-cell neutralizing epitopes mostly range from 14-22 residues which overlaps with our predicted epitopes. The novel epitopes that displayed epitope-CDR interactions range from 11-18 residues in length which makes them good candidates to be incorporated into peptide vaccines. Apart from considering the antigenicity of the shortlisted epitopes, their non-allergenicity and non-toxicity were also examined which are important factors to consider in the case of designing multi-epitope proteins (12).

Besides epitope mapping and visualization, three-dimensional models of the proteins are essential to study epitope recognition and accessibility by the antibodies (48). The template identified for ZIKV_RAJ was 7YW8.pdb whereas the template for ZIKV_MAH E protein identified as 7YW7.pdb had a deletion of 4 amino acids in the E-glycan loop (N154-157) (49). ZIKV strains with E glycan loop deletions, especially in Asia, need to be studied extensively to understand their effect on virus infectivity, immunity, and pathogenesis. The predicted structure's precision and stoichiometry were determined by the Ramachandran plot, which identified sterically allowed and disallowed regions (50). The most suitable 3-D models had zero per cent amino acids in disallowed regions and around >99% in the favorably allowed regions. Based on our study design we could identify novel linear and conformational epitopes for ZIKV E protein.

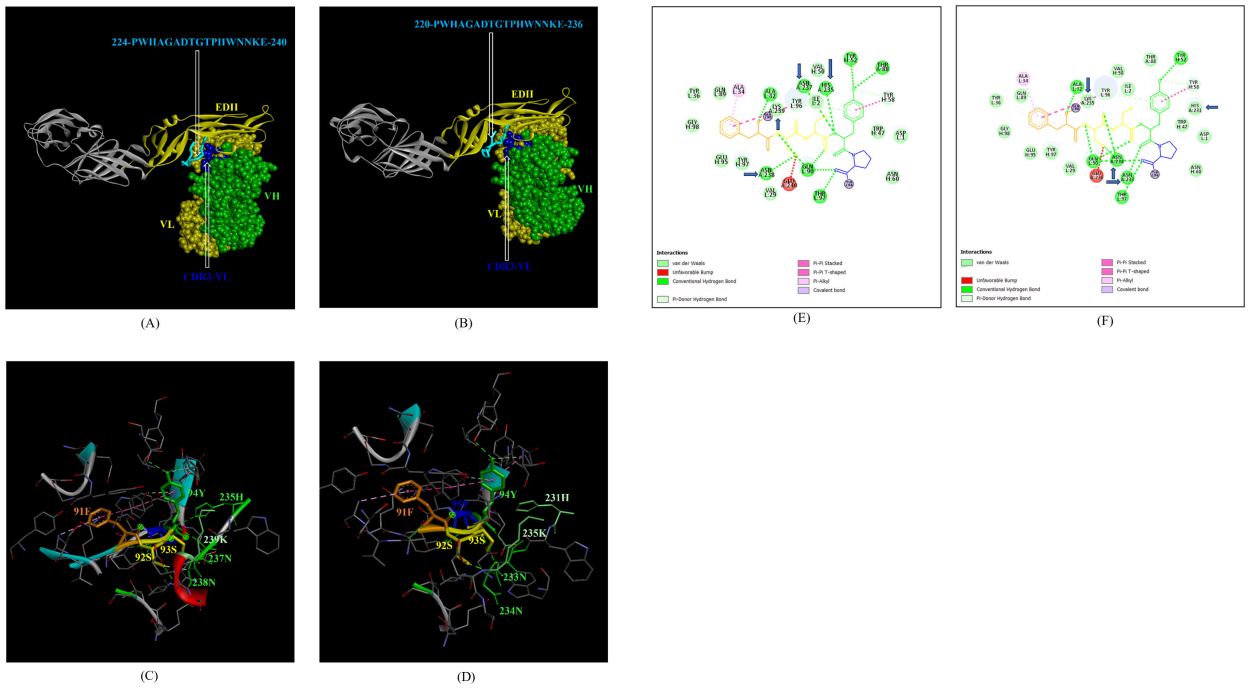


FIGURE 9
Molecular docking of novel linear EDII epitopes with CDRs of ZV-67 mAb. **(A)** 3-D model of docked ZRE and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VL (coloured in blue) interacting with EDII (coloured in yellow) epitope 224-PWHAGADTGTPHWNNKE-240 (coloured in light blue). **(B)** 3-D model of docked ZME and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VL (coloured in blue) interacting with EDII (coloured in yellow) epitope 224-PWHAGADTGTPHWNNKE-240 (coloured in light blue). **(C)** Epitope-CDR interaction in 3-D for ZRE with CDR3-VL of ZV67; 235H, 237-NNK-239 (coloured in shades of light green) are the amino acids of EDI epitope interacting with CDR3-VH amino acid residues, 91-FSSYP-95 (coloured in dark orange, yellow, yellow, light green and blue, respectively). **(D)** Epitope-CDR interaction in 3-D for ZME with CDR3-VL of ZV67; 235H, 237-NNK-239 (coloured in shades of light green) are the amino acids of EDI epitope interacting with CDR3-VH amino acid residues, 91-FSSYP-95 (coloured in dark orange, yellow, yellow, light green and blue, respectively). **(E)** 2-D interaction map of panel **(C)**; blue arrows indicating 235H forming one conventional hydrogen bond, 237-NN-238 forming one carbon-hydrogen bond each and 239K forming van der Walls interaction with 91F, 92S, 93S, 94F and 95P. **(F)** 2-D interaction map of panel **(D)**; blue arrows indicating 233-NN-234 forming one carbon-hydrogen bond each and 231H and 235K forming van der Walls interaction with 91F, 92S, 93S, 94F and 95P.

The stringent study criteria for selecting the novel epitopes were further validated *in silico* by studying their interactions with CDRs of the highly neutralizing ZIKV E-protein specific mouse and human mAbs. Interaction of the epitopes with the CDR regions of the antibody leads to enhanced binding affinity resulting in specific neutralizing response (51). Considering this, we hypothesized to select the CDRs of two highly neutralizing mAbs ZV-67 and Z3L1 to carry out the docking with ZIKV E protein. These two mAbs were selected as they were specific to ZIKV E protein neutralization, possessed high resolution X-ray crystallographic structures, and were effective across both African and Asian ZIKV strains *in vitro* and *in vivo* (21, 41). We identified that the novel linear epitopes across all three domains and DI-DIII hinge regions interacted with CDRs of ZV-67 mAb, which highlights the fact that the Indian ZIKV strains had broad-spectrum epitope-CDR interactions with this mouse mAb (Table 3). However, in the case of the human mAb Z3L1, only the novel epitope of EDII showed CDR interactions (Table 4). 3-D and 2-D interaction maps are essential to study the different bonds that form during epitope-CDR binding. All the novel epitopes, being identical or having point mutations between the ZIKV_RAJ and ZIKV_MAH displayed differences in their 3-D and 2-D receptor-ligand interaction maps, suggesting strain-specific differences in forming bonds with the mAbs' CDRs.

Our predictions also identified the EDIII epitopes 363-PVITESTENSK-373 (ZIKV_RAJ) and 363-PVITESAENSK-373 (ZIKV_MAH), which overlapped with critical residues involved in neutralization by the mAbs ZV-2 and ZV-67 (21). Similarly, epitopes 384-DSYIVIGVGDKKITHHWHRS-403 and 384-DSYIVIGVGEKKITHHWHRS-403 overlapped with key residues required for neutralization with mAbs ZV-48, ZV-64 and ZV-67 (21, 52). As, our prediction spanned across all three domains, including hinge regions, EDII epitope 61-YEASISDMASDSRCPT-76, and DI-DII hinge epitope 118-KFACSKKMTGSIQPE-133 (ZIKV_RAJ and ZIKV_MAH) were also part of key residues required for neutralization by mAb ZIKV-117 (53). Moreover, the mAb Z3L1 also had neutralizing epitopes overlapping with our predicted epitope 126-TGKSIQPENLEYRIMLSV-143 in the DI-DII hinge (41).

The ZIKV NS1 protein is another major target of neutralizing antibodies, following the immunodominant ZIKV E protein (23–26, 54–56). Hence, immunoinformatic analyses of ZIKV NS1 were undertaken. The ZIKV NS1 has three distinct domains: the beta roll

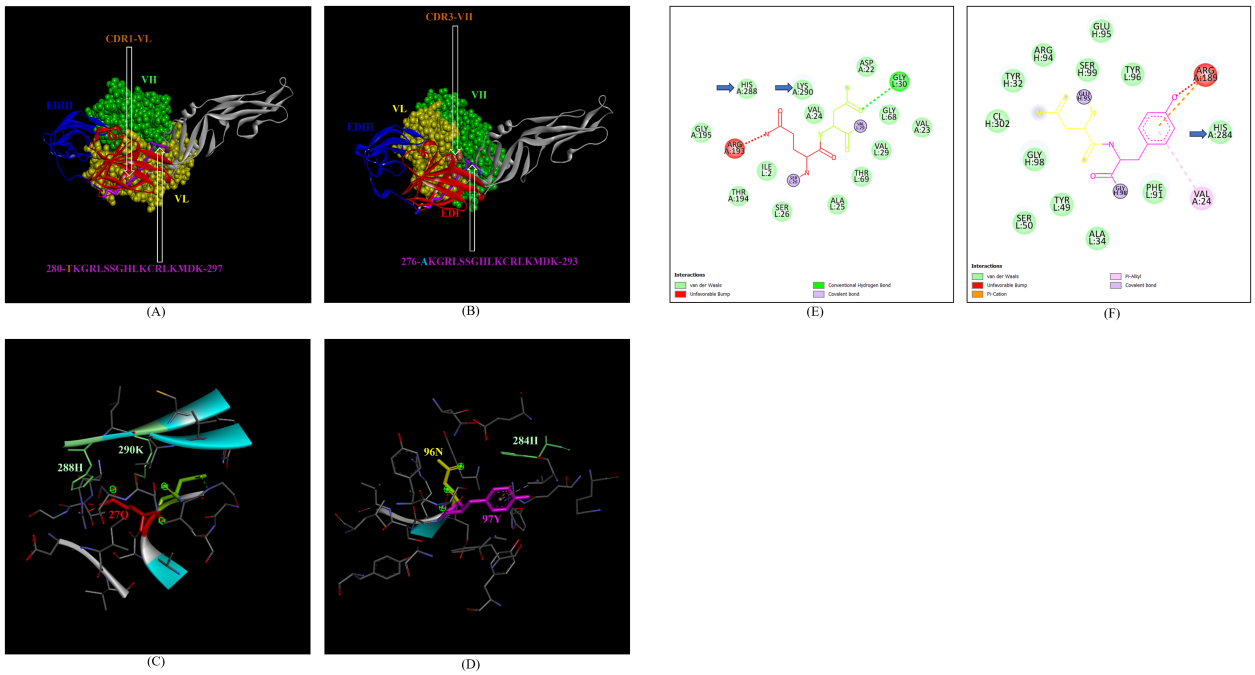


FIGURE 10
Molecular docking of novel linear hinge region epitopes with CDRs of ZV-67 mAb. **(A)** 3-D model of docked ZRE and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR1-VL (coloured in brown) interacting with DI/DIII hinge region (coloured in red/blue) epitope 280-TKGRLSSGHLKCRLKMDK-297 (coloured in purple; except Threonine which is colored in brown). **(B)** 3-D model of docked ZME and ZV67; VL (coloured in yellow), VH (coloured in light green), CDR3-VH (coloured in brown) interacting with DI/DIII hinge region (coloured in red/blue) epitope 276-AKGRLSSGHLKCRLKMDK-293 (coloured in purple; except alanine which is colored in light brown). **(C)** Epitope-CDR interaction in 3-D for ZRE with CDR1-VL of ZV67; 288H and 290K (coloured in shades of light green) are the amino acids of DI/DIII hinge region epitope interacting with CDR1-VL amino acid residues, 27-QN-28 (coloured in red and green, respectively). **(D)** Epitope-CDR interaction in 3-D for ZME with CDR3-VH of ZV67; 284H (coloured in shades of light green) is the amino acid of DI/DIII hinge region epitope interacting with CDR3-VH amino acid residues, 96-NY-97 (coloured in yellow and purple, respectively). **(E)** 2-D interaction map of panel **(C)**; blue arrows indicating 288H and 290K forming with van der Walls interaction with 27Q and 28N. **(F)** 2-D interaction map of panel **(D)**; blue arrows indicating 284H forming van der Walls interaction with 96N and 97Y.

(residues 1-29), the wing domain (residues 30-180) and the beta ladder (residues 181-352). Connector residues (33, 35–38, 41, 57, 58) and (152–180) within the wing domain link it to the beta roll and beta-ladder domain, respectively (54). Domain-wise antigenicity analysis revealed the beta-roll and beta ladder domains to be antigenic while the wing domain to be non-antigenic. The non-antigenic values of the wing domain may be attributed to the connector residues which form a part of the inner hydrophobic surface of NS1 (59). However, barring these residues, the wing domain is targeted by ZIKV NS1-specific mAbs that confer protection (24). These observations conveyed the importance of the wing domain in ZIKV protection and hence it was considered for epitope identification, alongside the antigenic beta roll and beta ladder domains.

Both ZIKV_RAJ and ZIKV_MAH NS1 had identical templates i.e. 5K6K. pdb (25) for 3-D structure predictions. As there was unavailability of linear ZIKV NS1-specific neutralizing epitopes at

TABLE 4 H-Docked based molecular docking analysis of ZIKV-specific Novel Linear Epitopes of the E-protein.

Novel Linear Epitopes	CDR Interactions	Indian ZIKV-specificity	Epitope/ CDR Interactions	2-D Interaction maps	Docking/ Binding scores
323-HGTVTVEVQYA-333 (EDIII)	No	ZIKV_RAJ (Unique)	N.A.	N.A.	N.A.
319-HGTVTVEVQYS-329 (EDIII)	No	ZIKV_MAH (Unique)	N.A	N.A.	N.A.
338-PCKVPAQM-345 (EDIII)	No	ZIKV_RAJ and ZIKV_MAH	N.A.	N.A.	N.A.
5-GVSNRDFVEGMSGGTW-20 (EDI)	No	ZIKV_RAJ and ZIKV_MAH	N.A	N.A.	N.A.
32-TVMAQDKPTVDIELVT-47 (EDI)	No	ZIKV_RAJ and ZIKV_MAH	N.A.	N.A.	N.A.

(Continued)

10.3389/fimmu.2025.1534737

TABLE 4 Continued

Novel Linear Epitopes	CDR Interactions	Indian ZIKV-specificity	Epitope/ CDR Interactions	2-D Interaction maps	Docking/ Binding scores
224-PWHAGADTGTPHWNNKE-240 (EDII);	Yes	ZIKV_RAJ and	231-TGA-233/CDR1-VH (ZIKV_RAJ)	231T: 1 H-bond 232G: 1 C-H bond 233A: van der Walls	-253.17
220-PWHAGADTGTPHWNNKE-236 (EDII)		ZIKV_MAH	223-AG-24/CDR1-VH (ZIKV_MAH)	223A: 1 H-bond and 1 pi-alklyl bond 224G: van der Walls	-242.52
43-IELVTTTVSNMAEVRS-58 (Hinge)	No	ZIKV_RAJ and ZIKV_MAH	N.A	N.A	N.A.
280-TKGRLSSGHLKCRLKMDK-297 (Hinge)	No	ZIKV_RAJ (Unique)	N.A	N.A	N.A.
276-AKGRLSSGHLKCRLKMDK-293 (Hinge)	No	ZIKV_MAH (Unique)	N.A	N.A	N.A.

Molecular Docking interactions of the novel epitopes with potent human monoclonal antibody Z3L1; N.A. (Not Available).
Amino acid mutations are coloured in Red with respect to ZIKV Prototype strain MR766 (Green).

IEDB, we were not able to ascertain the novel linear epitopes, but there was concordance of our predicted epitopes with critical residues of ZIKV NS1 neutralizing mAbs, such as 3G2, 4B8, 2E11, 14G5, AA12, EB9 and GB5 (54, 55, 60, 61). However, there were two epitopes 10-KKETRCGTGVFVYNDVE-26 (beta roll) and 83-GIQLTVVVGSVKNP-96 (wing domain) found exclusively in both Indian ZIKV strains which did not overlap with the critical residues of known ZIKV NS1 mAbs, suggesting that these may be important

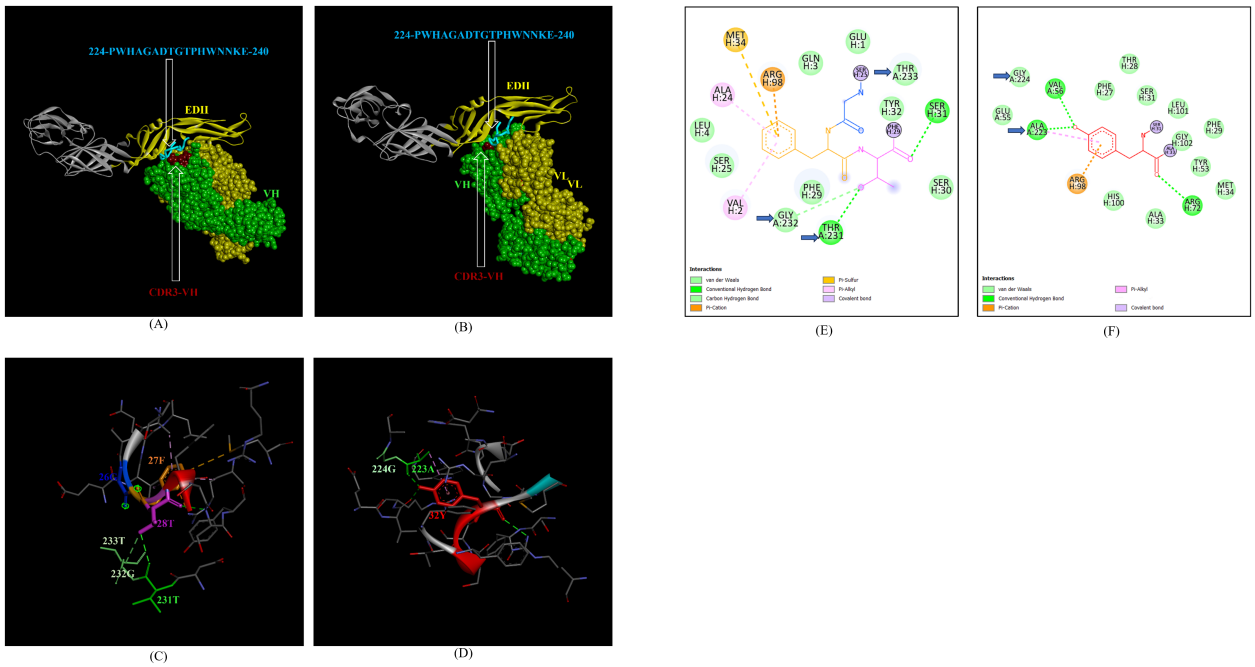


FIGURE 11
Molecular docking of novel linear EDII epitopes with CDRs of Z3L1 mAb. **(A)** 3-D model for docked ZRE and Z3L1; VL (coloured in yellow), VH (coloured in light green), CDR1-VH (coloured in magenta) interacting with EDII (coloured in yellow) epitope 224-PWHAGADTGTPHWNNKE-240 (coloured in light blue). **(B)** 3-D model for docked ZME and Z3L1; VL (coloured in yellow), VH (coloured in light green), CDR1-VH (coloured in magenta) interacting with EDII (coloured in yellow) epitope 224-PWHAGADTGTPHWNNKE-240 (coloured in light blue). **(C)** Epitope-CDR interaction in 3-D for ZRE with CDR1-VH of Z3L1; 231-TGT-233 colored in shades of light green) are the amino acids of EDII epitope interacting with CDR1-VL amino acid residues 26-GFT-28 (coloured in blue, orange and purple, respectively). **(D)** Epitope-CDR interaction in 3-D for ZME with CDR1-VH of Z3L1; 223-AG-224 (coloured in shades of light green) are the amino acids of EDII epitope interacting with CDR1-VL amino acid residue 32Y (coloured in red). **(E)** 2-D interaction map of panel **(C)**; blue arrows indicating 231T forming one conventional hydrogen bond, 232G forming one carbon-hydrogen bond and 233T forming van der Walls interaction with 26G, 27F, and 28T. **(F)** 2-D interaction map of panel **(D)**; blue arrows indicating 223A forming one conventional hydrogen bond and one pi-alkyl bond, 224G forming van der Walls interaction with 32Y. arrows indicating 223A (forming one H bond and one pi-alkyl bond), and 224G (forming van der Walls interaction) with 32Y.

to study as novel ZIKV NS1 epitopes. Among them, epitope 83-GIQLTVVVGSVKNP-96 did not belong to the hydrophobic connector residues of the wing domain, therefore it may be important in studying ZIKV NS1 protection. However, in the case of ZIKV NS1, epitope-CDR interactions were not possible to analyze due to the unavailability of resolved X-ray crystallographic structures of ZIKV NS1 neutralizing mAbs.

5 Conclusion

To summarize, we conducted detailed antigenic characterization of the E and NS1 proteins for co-circulating Indian strains of ZIKV with domain-specific analyses. This helped in the identification of novel epitopes in E and NS1 proteins having zero percent identity with the amino acid compositions of previously reported ZIKV-neutralizing epitopes. Molecular docking studies further revealed that some of the novel epitopes of E protein are being recognized by known ZIKV-neutralizing antibodies. Our studies on *in vitro* and *in vivo* experiments targeting these novel epitopes to understand the key role in humoral immunity are in progress. Therefore, the findings will help in the development of multi-epitope proteins for diagnostics and vaccinology applications in future.

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/s.

Author contributions

RR: Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Methodology. NT: Data curation, Methodology, Writing – review & editing. GD: Data curation, Formal analysis, Writing – review & editing. NA: Methodology, Software, Writing – review & editing. PS: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. GS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The entire research work was funded by ICMR-National Institute of Virology and grants of SRF were supported to RR from the Department of Biotechnology (DBT), Govt. of India.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1534737/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Multiple sequence alignment of E protein in ZIKV strains: ZIKV MR766 (coloured in green), ZIKV NATAL RGN (coloured in red), ZIKV_RAJ (coloured in blue), and ZIKV_MAH (coloured in purple). All the mutations are highlighted in yellow with respect to ZIKV MR766.

SUPPLEMENTARY FIGURE 2
Multiple sequence alignment of NS1 protein in ZIKV strains; ZIKV MR766 (coloured in green), ZIKV NATAL RGN (coloured in red), ZIKV_RAJ (coloured in blue), and ZIKV_MAH (coloured in purple). All the mutations are highlighted in yellow with respect to ZIKV MR766.

SUPPLEMENTARY FIGURE 3
3-D Templates for Indian ZIKV E; **(A)** 7YW8 (ZIKV_RAJ), **(B)** 7YW7 (ZIKV_MAH) where Chain A is highlighted in yellow.

SUPPLEMENTARY FIGURE 4
3-D Template for Indian ZIKV NS1: 5K6K (ZIKV_RAJ and ZIKV_MAH) where Chain A is highlighted in yellow.

SUPPLEMENTARY FIGURE 5
Graphical representation of linear B-cell epitopes by BepiPred 2.0 method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV E (Threshold=0.5). **(A)** ZIKV_RAJ and **(B)** ZIKV_MAH.

SUPPLEMENTARY FIGURE 6
Graphical representation of linear B-cell epitopes by Kolaskar and Tongaonkar method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV E. **(A)** ZIKV_RAJ (Threshold=1.026) and **(B)** ZIKV_MAH (Threshold=1.028).

SUPPLEMENTARY FIGURE 7
Graphical representation of linear B-cell epitopes by Emini surface accessibility method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV E (Threshold=1.00). **(A)** ZIKV_RAJ and **(B)** ZIKV_MAH.

SUPPLEMENTARY FIGURE 8
Graphical representation of linear B-cell epitopes by Parker hydrophilicity method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV E. **(A)** ZIKV_RAJ (Threshold=1.475) and **(B)** ZIKV_MAH (Threshold=1.470).

SUPPLEMENTARY FIGURE 9
Graphical representation of linear B-cell epitopes by BepiPred 2.0 method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV NS1 (Threshold=0.5). **(A)** ZIKV_RAJ and **(B)** ZIKV_MAH.

SUPPLEMENTARY FIGURE 10
Graphical representation of linear B-cell epitopes by Kolaskar and Tongaonkar method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV NS1. **(A)** ZIKV_RAJ (Threshold=1.018) and **(B)** ZIKV_MAH (Threshold=1.023).

SUPPLEMENTARY FIGURE 11
Graphical representation of linear B-cell epitopes by Emini surface accessibility method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV NS1 (Threshold=1.00). **(A)** ZIKV_RAJ and **(B)** ZIKV_MAH.

SUPPLEMENTARY FIGURE 12
Graphical representation of linear B-cell epitopes by Parker hydrophilicity method (yellow peaks = predicted epitopes, and green inverted peaks = non-epitopes) for both Indian ZIKV NS1. **(A)** ZIKV_RAJ (Threshold=1.701) and **(B)** ZIKV_MAH (Threshold=1.726).

References

1. Dick GWA, Kitchen SF, Haddow AJ. Zika Virus (I). Isolations and serological specificity. *Trans R Soc Trop Med Hyg*. (1952) 46:509–20. doi: 10.1016/0035-9203(52)90042-4

2. Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on yap island, federated states of Micronesia. *New Engl J Med*. (2009) 360:2536–43. doi: 10.1056/NEJMoa0805715

3. Yuan L, Huang XY, Liu ZY, Zhang F, Zhu XL, Yu JY, et al. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Sci (1979)*. (2017) 358:933–6. doi: 10.1126/science.aam7120

4. de Oliveira WK, Carmo EH, Henriques CM, Coelho G, Vazquez E, Cortez-Escalante J, et al. Zika virus infection and associated neurologic disorders in Brazil. *New Engl J Med*. (2017) 376:1591–3. doi: 10.1056/NEJMc1608612

5. Cao-Lormeau VM, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet*. (2016) 387:1531–9. doi: 10.1016/S0140-6736(16)00562-6

6. Heukelbach J, Alencar CH, Kelvin AA, De Oliveira WK, Pamplona de Góes Cavalcanti L. Zika virus outbreak in Brazil. *J Infection Developing Countries*. (2016) 10:116–20. doi: 10.3855/jidc.8217

7. Zanluca C, de Melo VCA, Mosimann ALP, dosSantos GIV, dosSantos CND, Luz K. First report of autochthonous transmission of Zika virus in Brazil. *Mem Inst Oswaldo Cruz*. (2015) 110:569–72. doi: 10.1590/0074-02760150192

8. Sapkal GN, Yadav PD, Vegad MM, Viswanathan R, Gupta N, Mourya DT. First laboratory confirmation on the existence of Zika virus disease in India. *J Infection*. (2018) 76:314–7. doi: 10.1016/j.jinf.2017.09.020

9. Yadav PD, Malhotra B, Sapkal G, Nyayanit DA, Deshpande G, Gupta N, et al. Zika virus outbreak in Rajasthan, India in 2018 was caused by a virus endemic to Asia. *Infection Genet Evolution*. (2019) 69:199–202. doi: 10.1016/j.meegid.2019.01.026

10. Malhotra B, Gupta V, Sharma P, Singh R, Sharma H, Vyas M, et al. Clinico-epidemiological and genomic profile of first Zika Virus outbreak in India at Jaipur city of Rajasthan state. *J Infect Public Health*. (2020) 13:1920–6. doi: 10.1016/j.jiph.2020.10.006

11. Gurav YK, Alagarasu K, Yadav PD, Sapkal G, Gokhale M, Parashar D, et al. First case of Zika virus infection during an outbreak of chikungunya in a rural region of Maharashtra state, India. *Trans R Soc Trop Med Hyg*. (2022) 116:974–7. doi: 10.1093/trstmh/trac022

12. Quanquin N, Wang L, Cheng G. Potential for treatment and a Zika virus vaccine. *Curr Opin Pediatr*. (2017) 29:114–21. doi: 10.1097/MOP.0000000000000441

13. Malone RW, Homan J, Callahan MV, Glasspool-Malone J, Damodaran L, Schneider ADB, et al. Zika virus: medical countermeasure development challenges. *PloS Negl Trop Dis*. (2016) 10:e0004530. doi: 10.1371/journal.pntd.0004530

14. Russell K, Oliver SE, Lewis L, Barfield WD, Cragan J, Meaney-Delman D, et al. Update: interim guidance for the evaluation and management of infants with possible congenital zika virus infection — United states, august 2016. *MMWR Morb Mortal Wkly Rep*. (2016) 65:870–8. doi: 10.15585/mmwr.mm6533e2

15. Priyamvada L, Suthar MS, Ahmed R, Wrammert J. Humoral immune responses against zika virus infection and the importance of preexisting flavivirus immunity. *J Infect Dis*. (2017) 216:S906–11. doi: 10.1093/infdis/jix513

16. Pavitrakar DV, Atre NM, Tripathy AS, Shil P. Design of a multi-epitope peptide vaccine candidate against chandipura virus: an immuno-informatics study. *J Biomol Struct Dyn*. (2022) 40:648–59. doi: 10.1080/07391102.2020.1816493

17. Gangwar RS, Shil P, Sapkal GN, Khan SA, Gore MM. Induction of virus-specific neutralizing immune response against West Nile and Japanese encephalitis viruses by chimeric peptides representing T-helper and B-cell epitopes. *Virus Res*. (2012) 163:40–50. doi: 10.1016/j.virusres.2011.08.008

18. Sela-Culang I, Kunik V, Ofran Y. The structural basis of antibody-antigen recognition. *Front Immunol*. (2013) 4. doi: 10.3389/fimmu.2013.00302

19. Yang C, Gong R, de Val N. Development of neutralizing antibodies against zika virus based on its envelope protein structure. *Virol Sin*. (2019) 34:168–74. doi: 10.1007/s12250-019-00093-5

20. Gallichotte EN, Young EF, Baric TJ, Yount BL, Metz SW, Begley MC, et al. Role of zika virus envelope protein domain III as a target of human neutralizing antibodies. *mBio*. (2019) 10. doi: 10.1128/mBio.01485-19

21. Stettler K, Beltramello M, Espinosa DA, Graham V, Cassotta A, Bianchi S, et al. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science*. (2016) 353:823–6. doi: 10.1126/science.aaf8505

22. Hilgenfeld R. Zika virus NS 1, a pathogenicity factor with many faces. *EMBO J*. (2016) 35:2631–3. doi: 10.15252/embj.201695871

23. Bailey MJ, Duehr J, Dulin H, Broecker F, Brown JA, Arumemi FO, et al. Human antibodies targeting Zika virus NS1 provide protection against disease in a mouse model. *Nat Commun*. (2018) 9:4560. doi: 10.1038/s41467-018-07008-0

24. Yu L, Liu X, Ye X, Su W, Zhang X, Deng W, et al. Monoclonal Antibodies against Zika Virus NS1 Protein Confer Protection via Fc γ Receptor-Dependent and -Independent Pathways. *mBio*. (2021) 12.

25. Brown WC, Akey DL, Konwerski JR, Tarrasch JT, Skiniotis G, Kuhn RJ, et al. Extended surface for membrane association in Zika virus NS1 structure. *Nat Struct Mol Biol*. (2016) 23:865–7. doi: 10.1038/nsmb.3268

26. Shi Y, Dai L, Song H, Gao GF. Structures of zika virus E & NS1: relations with virus infection and host immune responses. (2018), 77–87.

27. Antonelli AC, Almeida VP, da Fonseca SG. Immunoinformatics vaccine design for zika virus. (2023), 411–29.

28. 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) 7:226. doi: 10.3390/microorganisms7080226

29. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf*. (2007) 8. doi: 10.1186/1471-2105-8-4

30. 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.

31. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. (1990) 276:172–4. doi: 10.1016/0014-5793(90)80535-Q

32. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure Function Bioinf*. (2006) 65:40–8. doi: 10.1002/prot.21078

33. Tong JC, Ranganathan S. Computational B cell vaccine design. *In: Computer-Aided Vaccine Design. Elsevier;*. (2013), 87–98. doi: 10.1533/9781908818416.87

34. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*. (1985) 55:836–9. doi: 10.1128/jvi.55.3.836-839.1985

35. 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

36. 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:e73957. doi: 10.1371/journal.pone.0073957

37. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. Peptide toxicity prediction. *In*. (2015), 143–57.

38. 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 Bioinf.* (2008) 9:514. doi: 10.1186/1471-2105-9-514

39. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PloS Comput Biol.* (2012) 8:e1002829. doi: 10.1371/journal.pcbi.1002829

40. Zhao H, Fernandez E, Dowd KA, Speer SD, Platt DJ, Gorman MJ, et al. Structural basis of zika virus-specific antibody protection. *Cell.* (2016) 166:1016–27. doi: 10.1016/j.cell.2016.07.020

41. Dai L, Song J, Lu X, Deng YQ, Musyoki AM, Cheng H, et al. Structures of the zika virus envelope protein and its complex with a flavivirus broadly protective antibody. *Cell Host Microbe.* (2016) 19:696–704. doi: 10.1016/j.chom.2016.04.013

42. Wessel AW, Kose N, Bombardi RG, Roy V, Chantima W, Mongkolsapaya J, et al. Antibodies targeting epitopes on the cell-surface form of NS1 protect against Zika virus infection during pregnancy. *Nat Commun.* (2020) 11:5278. doi: 10.1038/s41467-020-19096-y

43. Wang Q, Yang H, Liu X, Dai L, Ma T, Qi J, et al. Molecular determinants of human neutralizing antibodies isolated from a patient infected with Zika virus. *Sci Transl Med.* (2016) 8. doi: 10.1126/scitranslmed.aai8336

44. Collins MH, Tu HA, Gimblet-Ochieng C, Liou GJA, Jadi RS, Metz SW, et al. Human antibody response to Zika targets type-specific quaternary structure epitopes. *JCI Insight.* (2019) 4. doi: 10.1172/jci.insight.124588

45. Aquino VH, Fumagalli MJ, Silva A, de Moura Negrini BV, Rojas A, Guillen Y, et al. Linear epitope mapping in the E and NS1 proteins of dengue and Zika viruses: Prospection of peptides for vaccines and diagnostics. *PloS One.* (2023) 18:e0292451. doi: 10.1371/journal.pone.0292451

46. Gershoni JM, Roitburd-Berman A, Siman-Tov DD, Tarnovitski Freund N, Weiss Y. Epitope mapping. *BioDrugs.* (2007) 21:145–56. doi: 10.2165/00063030-200721030-00002

47. Nevagi RJ, Toth I, Skwarczynski M. Peptide-based vaccines. *Pept Appl Biomedicine Biotechnol Bioengineering.* (2018), 327–58. doi: 10.1016/B978-0-08-100736-5.00012-0

48. Srivastava K, Srivastava V. Prediction of conformational and linear B-cell epitopes on envelop protein of zika virus using immunoinformatics approach. *Int J Pept Res Ther.* (2023) 29:17. doi: 10.1007/s10989-022-10486-y

49. Cheng ML, Yang YX, Liu ZY, Wen D, Yang P, Huang XY, et al. Pathogenicity and structural basis of zika variants with glycan loop deletions in the envelope protein. *J Virol.* (2022) 96. doi: 10.1128/jvi.00879-22

50. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts.* (2010) 1:271–83. doi: 10.1515/bmc.2010.022

51. Mejias-Gomez O, Madsen AV, Skovgaard K, Pedersen LE, Morth JP, Jenkins TP, et al. A window into the human immune system: comprehensive characterization of the complexity of antibody complementary-determining regions in functional antibodies. *MAbs.* (2023) 15. doi: 10.1080/19420862.2023.2268255

52. Almansour I, Alfares R, Aljofi H. Large-scale analysis of B-cell epitopes of envelope: Implications for Zika vaccine and immunotherapeutic development. *F1000Res.* (2019) 7:1624. doi: 10.12688/f1000research

53. Hasan SS, Miller A, Sapparapu G, Fernandez E, Klose T, Long F, et al. A human antibody against Zika virus crosslinks the E protein to prevent infection. *Nat Commun.* (2017) 8:14722. doi: 10.1038/ncomms14722

54. Alvin Chew BL, Pan Q, Hu H, Luo D. Structural biology of flavivirus NS1 protein and its antibody complexes. *Antiviral Res.* (2024) 227:105915. doi: 10.1016/j.antiviral.2024.105915

55. Pan Q, Xing X, Yu J, Chen Q, Jiao H, Zhang W, et al. Structural insights into the distinct protective mechanisms of human antibodies targeting ZIKV NS1. *hLife.* (2024). doi: 10.1101/2023.10.16.562450

56. Cavazzoni CB, Bozza VBT, Lucas TCV, Conde L, Maia B, Mesin L, et al. The immunodominant antibody response to Zika virus NS1 protein is characterized by cross-reactivity to self. *J Exp Med.* (2021) 218. doi: 10.1084/jem.20210580

57. Parker JMR, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry.* (1986) 25:5425–32. doi: 10.1021/bi00367a013

58. Yan Y, Tao H, He J, Huang SY. The HDOCK server for integrated protein–protein docking. *Nat Protoc.* (2020) 15:1829–52. doi: 10.1038/s41596-020-0312-x

59. Akey DL, Brown WC, Dutta S, Konwerski J, Jose J, Jurkiw TJ, et al. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Sci (1979).* (2014) 343:881–5. doi: 10.1126/science.1247749

60. Modhiran N, Watterson D, Muller DA, Panetta AK, Sester DP, Liu L, et al. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. *Sci Transl Med.* (2015) 7. doi: 10.1126/scitranslmed.aaa3863

61. Gao X, Wen Y, Wang J, Hong W, Li C, Zhao L, et al. Delayed and highly specific antibody response to nonstructural protein 1 (NS1) revealed during natural human ZIKV infection by NS1-based capture ELISA. *BMC Infect Dis.* (2018) 18:275. doi: 10.1186/s12879-018-3173-y

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 28 March 2025
DOI 10.3389/fimmu.2025.1546254

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Concordia University, Canada

REVIEWED BY
Satyavani Kaliamurthi,
Saveetha Institute of Medical and Technical Sciences, India
Md. Oliullah Rafi,
Bangladesh Agricultural University, Bangladesh

*CORRESPONDENCE
Bandar Alosaimi
Balosaimi@kfmc.med.sa

RECEIVED 16 December 2024
ACCEPTED 03 March 2025
PUBLISHED 28 March 2025

CITATION
Alnajran H, Awadalla M, Aldakheel FM, Alam I, Momin AA, Alturaiki W and Alosaimi B (2025) Design of a peptide-based vaccine against human respiratory syncytial virus using a reverse vaccinology approach: evaluation of immunogenicity, antigenicity, allergenicity, and toxicity. *Front. Immunol.* 16:1546254. doi: 10.3389/fimmu.2025.1546254

Design of a peptide-based vaccine against human respiratory syncytial virus using a reverse vaccinology approach: evaluation of immunogenicity, antigenicity, allergenicity, and toxicity

Hadeel Alnajran[1,2], Maaweya Awadalla[2], Fahad M. Aldakheel[1], Intikhab Alam[3], Afaque A. Momin[3], Wael Alturaiki[4] and Bandar Alosaimi[2*]

[1]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia, [2]Research Center, King Fahad Medical City, Riyadh Second Health Cluster, Riyadh, Saudi Arabia, [3]Center of Excellence for Smart Health (KCSH), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [4]Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, Majmaah, Saudi Arabia

Background: Attempts to develop an hRSV vaccine have faced safety and efficacy challenges, with only three FDA-approved vaccines (Moderna's Mresvia, Pfizer's Abrysvo, and GSK's Arexvy) available. These vaccines are limited to individuals over 60 years, require boosters, and only reduce disease severity without clearing the infection. Therefore, we employed a reverse vaccinology approach in this study to identify the most promising antigenic epitopes capable of eliciting a robust and protective immune response.

Methodology: This study employed computational techniques to design a novel multi-epitope vaccine targeting hRSV. Using bioinformatics tools, candidate epitopes were identified from conserved viral proteins (F and G glycoproteins), assessing their immunogenicity, antigenicity, and allergenicity. Key tools included ExPASy, ProtParam, VaxiJen v2.0, AllergenFP v1.0, AllerTOP v2.0, NetCTL v1.2, IEDB, and Toxin-Pred. The vaccine construct was assessed for stability and toxicity through *in silico* analyses. We then characterized its kinetic properties, evaluated its structural integrity, and analyzed its interactions with Toll-like receptors (TLRs) using molecular docking, modeling, and refinement with AlphaFold3 and ClusPro.

Results: The designed constructs showed strong antigenicity (0.5996 for F-based and 0.6048 for G-based vaccine), non-allergenicity, and stability (instability index <40). Among these, most amino acids were in the extracellular domain of the construct. Molecular docking and dynamics simulations indicated strong binding interactions with TLR1 and TLR4 and minimal RMSF fluctuations, which ensured structural stability. Strong humoral and cellular responses were suggested by in silico immune simulation demonstrating robust immune activation, with high levels of IgG, IgM, IL-2, and IFN-γ. The physical and

chemical analyses revealed that the majority of amino acids from the F and G proteins were located in the extracellular domain of the construct. The presence of signal peptide cleavage sites in both glycoprotein components further facilitates antigen presentation to the immune system.

Conclusions: This study presents a promising peptide-based vaccine candidate against hRSV that can effectively engage the immune system, showing strong immunogenicity and antigenicity. Future *in vitro* and *in vivo* studies are essential to evaluate the ability of the multi-epitope vaccine candidate to stimulate both humoral and cell-mediated immune responses and to assess its efficacy and safety profile.

KEYWORDS

human respiratory syncytial virus, hRSV, immunoinformatics, CTL epitope, HTL epitope, B-cell epitope, reverse vaccinology

1 Introduction

Human respiratory syncytial virus (hRSV) is a major cause of lower respiratory tract infections in infants, young children, and the elderly (1). It is estimated that hRSV causes 33 million new episodes of acute lower respiratory infections and 3.2 million hospital admissions annually worldwide (2). Despite the significant disease burden, currently, only three hRSV vaccines have been licensed by the FDA (3). Developing an effective hRSV vaccine has been challenging due to the complex immune responses to the virus and the risk of disease enhancement in vaccinated individuals (4).

Human RSV exhibits antigenic variability, with two major antigenic subgroups (A and B) circulating globally, complicating the design of a universal vaccine (4). Furthermore, previous attempts at hRSV vaccine development have been hampered by safety concerns, such as the phenomenon of vaccine-enhanced respiratory disease observed with a formalin-inactivated RSV vaccine candidate (5).

The most recent vaccine design efforts were focused on hRSV envelope proteins embedded in the lipid bilayer, specifically the attachment (G) glycoprotein and/or the fusion (F) glycoprotein. Human RSV G protein can exist in two forms, as complete membrane-bound glycoprotein (mG) that mediates viral attachment to host cells *in vivo* and secreted N-terminally truncated G protein (sG) (6). sG can modulate host immune responses, enabling it to evade, alter, or inactivate both innate defenses and the adaptive immune system, as well as influence the antiviral activity of monoclonal antibodies (mABs) (7). In addition, the extensive antigenic variability of G protein among different hRSV strains has been another significant obstacle to the development of an effective vaccine (8). Although these limitations were recently addressed to some extent through further optimization using the CsA adjuvant, these challenges shifted the focus to F glycoprotein, a more conserved viral surface component (9). The F protein allows hRSV penetration and fusion between adjacent cells to form syncytium. The viral F glycoprotein undergoes dynamic reconfiguration when binding to the target cell's plasma membrane and thus exists in two forms: the prefusion form (pre-F) and the more stable post-fusion form or post-F (10). The unstable pre-F sequence was substantial in developing the two peptide-based vaccines and one mRNA vaccine approved by the FDA for hRSV, as it is highly immunogenic and stimulates the production of RSV-specific neutralizing antibodies (NAbs) (3).

BAFF and APRIL are crucial for B cell survival, differentiation, and antibody production. They interact with specific receptors on B cells, promoting their activation and proliferation, which is essential for generating a robust immune response. Previous studies have demonstrated that BAFF and APRIL can enhance the immunogenicity of vaccines. For instance, research has shown that plasmids expressing multimeric soluble BAFF or APRIL, when co-administered with other immunomodulatory agents, can significantly increase antibody titers and neutralizing antibody responses against HIV-1 (11, 12). Additionally, constructs combining HIV-1 envelope proteins with APRIL have been reported to enhance antibody responses in animal models (11). This has led to interest in their potential as adjuvants to improve vaccine efficacy while modulating immune responses.

The traditional approach of vaccine development relies on virus culturing and its activation which raises several safety concerns. Reverse vaccinology offers a promising approach and a rapid, cost-effective, and reliable methodology for the preliminary selection and design of novel multi-epitope vaccine candidates against hRSV. This approach involves comprehensive *in silico* analysis of the hRSV attachment and fusion proteome to identify the most promising antigenic epitopes capable of eliciting a robust and protective immune response (13, 14). Computational vaccinology techniques, such as epitope prediction, antigenicity and allergenicity analysis, and molecular docking, can be employed to design and

evaluate multi-epitope vaccine candidates *in silico* before experimental validation (13, 14). By targeting multiple conserved epitopes from Glycoprotein and fusion hRSV proteins, a multi-epitope vaccine has the potential to provide broad coverage against both major hRSV subgroups and induce a balanced, long-lasting immune response.

The study aims to investigate the potential of a computationally designed multi-epitope vaccine to initiate a protective immune response, ensure the epitopes have stability and non-allergenicity, and provide broader coverage across the subtypes of hRSV. Furthermore. This study utilized a reverse vaccinology strategy to systematically analyze the hRSV proteome, identify immunogenic epitopes, and design a multi-epitope vaccine candidate against human RSV. The selected epitopes were further evaluated for their antigenicity, immunogenicity, allergenicity, and molecular docking properties to ensure the development of a safer vaccine formulation.

2 Materials and methods

2.1 Protein sequences retrieval

The amino acid sequences of fusion glycoprotein (F) [Human Orthopneumovirus] and attachment G protein [Human Respiratory Syncytial Virus A] with EMBL IDs QID88623.1 and ALB35397.1 respectively were retrieved from UniProt (15) in FASTA format to predict T cell, B cell, and IFN-gamma inducing epitopes.

2.2 Analysis of physicochemical properties

To calculate the chemical and physical properties of the target protein sequence, the ExPASy ProtParam tool (16) was used. This tool enables the calculation of a range of physicochemical parameters for proteins either retrieved from UniProtKB or provided as user-entered sequences In this study, the amino acid sequence of the target protein, represented in one-letter code, was input into the appropriate field, and the compute parameters option was selected. No additional data was required for the analysis. The computed metrics included molecular weight, aliphatic index, theoretical isoelectric point (pI): Determines the pH at which the protein has no net charge, instability index, extinction coefficient, grand average of hydropathicity grand average of hydropathy (GRAVY), and atomic composition.

2.3 Evaluation of antigenic properties

The potential vaccine candidates (PVCs) from the proteome of hRSV were predicted using the VaxiJen v2.0 server (17). VaxiJen is a Perl-based server with an HTML interface that classifies proteins as "Probable Non-Antigen" or "Probable Antigen" based on their antigen probability, which is expressed as a percentage. The default threshold value of 0.4 was used for this analysis. Additionally, two other tools were employed to distinguish between allergens and non-allergens: AllergenFP v.1.0 (18) and AllerTOP v.2.0 server (14). To predict the presence of transmembrane helices and signal peptides, TMHMM v2.0 (19) and SignalP 6 (20) tools were used, respectively. In all these tools the input parameters were protein sequences.

2.4 CTL epitope prediction and binding affinity analysis with MHC I allele

The prediction of cytotoxic T lymphocyte (CTL) epitopes was carried out using the NetCTL v1.2 server. The NetCTL 1.2 server predicts CTL epitopes in protein sequences (21) and an Immune Epitope Database (IEDB) tool (22) The input parameters consisted of peptide sequences, while the expected outputs included predicted CTL epitopes and binding affinities for MHC class I and II. These epitopes were classified, based on their binding to various major histocompatibility complex (MHC) alleles, including HLA-I, HLA-II/H-2-IAb, HLA-II/H-2-IAd, and H-2-Db. The CTL epitopes binding to HLA-I and H-2-Db alleles were retrieved from the NetCTL tool, while the CTL epitopes binding to H-2-IAb and HLA-II/H-2-IAd were obtained from the IEDB tool. Epitopes with a consensus score of less than 2 were considered excellent binders and selected for further analysis. The selected epitopes were then assessed for their antigenicity, immunogenicity, allergenic profile, and toxicity using the VaxiJen v.2.0, IEDB, AllergenFP v.1.0, and ToxinPred servers, respectively. The best epitopes were those with high antigenicity, non-allergenicity, and non-toxicity.

2.5 HTL epitope prediction and binding affinity analysis with MHC2 allele

The prediction of helper T lymphocyte (HTL) epitopes was performed using the NetMHCII pan 3.287 server (23). The HTL epitopes were classified based on their binding to human leukocyte antigen (HLA) class II alleles, specifically HLA-II/H-2-IAb and HLA-II/H-2-IAd. For the NetMHCII tool, the input parameters consisted of peptide sequences, while the expected outputs included predicted HTL epitopes and binding affinities. The antigenicity of each predicted epitope was evaluated using the VaxiJen v.2.0 server, with a threshold value of 0.4. To exclude potential allergenic epitopes, the AllergenFP v.1.089 server was employed. Furthermore, PyMOL was used to visualize the location of the predicted epitopes on the glycoprotein structure. Finally, the ToxinPred server was utilized to assess the toxicity profile of the selected epitopes. ToxinPred is a web server designed to predict whether proteins or peptides are toxic or non-toxic. We used peptide sequences as input parameters, and the expected output is a toxicity score indicating whether the peptide is toxic or non-toxic. The best epitopes were those that demonstrated high antigenicity, non-allergenicity, and non-toxicity after the filtration process.

2.6 Prediction of B-cell epitopes

The Immune Epitope Database (IEDB) (24) was employed to predict B cell epitopes based on the protein sequence. To select the final B cell epitope candidates, several servers were utilized to screen their properties: VaxiJen v.2.0, AllergenFP v.1.0, and ToxinPred. The best B cell epitope candidates were those that demonstrated high antigenicity, non-allergenicity, and non-toxicity after the screening process using these complementary computational tools.

2.7 Prediction of interferon-gamma-inducing epitopes

To predict and design IFN-γ epitopes for vaccine development, an IFN epitope web server was employed. This server enables users to predict and design peptides that induce IFN-gamma, MHC Class II bindings, or T-cell epitopes. We input peptide sequences as parameters, and the anticipated output includes both IFN-inducing and non-inducing epitopes. This server features three primary modules: Predict, Design, and Scan. It utilizes a dataset to classify IFN-γ epitopes into two distinct categories: those capable of producing IFN-γ and those that cannot. The server's predictions are based on three strategies: hybrid, motif-based, and machine-learning approaches, thus offering an accuracy of up to 81.39% (25). For this study, multiple peptide sequences were input into the server, and the IDEB database, an experimentally validated dataset comprising 10,433 T-cell epitopes, was employed. Upon protein input, the hybrid approach combining motifs and support vector machines was selected to perform the predictions. The output was generated as numerical scores, where a positive value indicated the secretion of IFN-γ by the predicted epitopes.

2.8 Population coverage analysis

The population coverage analysis of human MHC alleles (HLA I and II) was carried out using the IEDB population coverage tool and the results were plotted in the form of a bar chart (26). In the study, default settings were used, and population coverage was evaluated for each class of MHC.

2.9 Multi-epitope vaccine design

To construct our vaccine, commonly used linker sequences in multi-epitope vaccine designs were employed to connect different types of epitopes. Linkers are an essential component in the design of multi-epitope vaccines, serving several crucial functions that enhance vaccine efficacy. They facilitate the proper folding of individual epitopes, ensuring that each maintains its correct conformation during protein synthesis, which is vital for effective recognition by the immune system. Additionally, linkers improve the overall immunogenicity of the vaccine by providing flexibility between epitopes, allowing for better presentation to immune cells and thereby enhancing the immune response. They also prevent steric hindrance that could occur if epitopes are positioned too closely together, ensuring effective interaction with T cell receptors and other components of the immune system. Furthermore, the incorporation of linkers contributes to the stability of the vaccine construct, helping to protect the epitopes from degradation. Thus, the strategic use of linkers is fundamental in optimizing the performance of multi-epitope vaccines Cytotoxic T lymphocyte (CTL) epitopes were linked using the AYY linker sequence. The AYY linker is a flexible linker that facilitates the proper folding and presentation of the CTL epitopes (27). Helper T lymphocyte (HTL) epitopes were connected using the GPGPG linker. This linker is commonly used to separate distinct epitopes while maintaining their structures and functions. The selected epitopes targeting B cells were linked using the KK linker. The KK linker, composed of two lysine residues, enhances the immunogenicity of the B cell epitopes by promoting their proper folding and exposure. BAFF and April adjuvant were then incorporated into the vaccine construct. The adjuvants were linked to the N-terminus of the vaccine sequence using the EAAAK linker. The EAAAK linker is a rigid alpha-helical linker that maintains the structural integrity and functionality of the adjuvant. The use of these specific linker sequences aims to optimize the presentation and immunogenicity of the different epitope types (CTL, HTL, and B cell) within the multi-epitope vaccine construct. The BAFF and April adjuvant, when linked to the vaccine, are expected to enhance the overall immune response generated by the vaccine.

2.10 Evaluation of physicochemical properties, antigenicity, and allergenicity of vaccine construct

The allergenic, antigenic, and toxicity profiles of the final multi-epitope vaccine construct were evaluated using the same computational tools identified above for the F and G protein epitopes. The VaxiJen v2.0 server was used to predict the antigenicity of the multi-epitope vaccine construct, the AllergenFP v1.0 server was employed to assess the allergenic potential of the multi-epitope vaccine, and the ToxinPred server was utilized to evaluate the toxicity profile of the multi-epitope vaccine construct.

2.11 Prediction of secondary structure

The three-dimensional structures of Toll-like receptor 2 (TLR2) and Toll-like receptor 4 (TLR4) were retrieved from the RCSB Protein Data Bank (PDB) database (28). These PDB structures served as the structural templates for the computational analysis of the vaccine construct. The secondary structure properties of the multi-epitope vaccine construct were determined using the Self-Optimized Prediction Method with Alignment (SOPMA) server

and the Protein Structure Prediction Server (PSIPRED) v4.0 tool. For the SOPMA analysis, the default parameters were used.

2.12 Protein structural modeling, docking, refinement, and validation

The ERRAT server (29) was used to assess the overall quality of the 3D vaccine model by evaluating the statistics of non-bonded interactions between different atom types. To further refine the modeled structure, a CASP10 web-based approach, the Galaxy Refine tool (30) was utilized. The three-dimensional structures of the vaccine candidates were modeled using the Alphafold3 webserver following the standard settings (31). The TLR1 toll-like receptor sequence for Mus musculus was retrieved using the UniProt database (Uniprot ID-B9EJ46). The structure for the TLR1 receptor was also predicted using the Alphafold3 server. ClusPro molecular docking algorithm with no restraints or modifications in the structure (32) was used to perform molecular docking of the TLR domain with the vaccine candidates from both F and G glycoprotein (33). The dynamics and refinement studies were performed using Cabs-Flex 2.0 standalone (30) in the SS2 mode settings with a minimum distance along the protein chain was set at 3. The minimum length of restraints was set to 3.8 Å and the maximum length to 8.0Å. The number of cycles was increased to 100,000 at a temperature of 310K, while cycles between trajectories were set to 100, due to the large complex formed between TLR and vaccine candidate to have the best quality output per frame. A random seed was generated for every run for better comparison and correct error calculation. The interaction between the TLR receptor and vaccine candidate was analyzed by eye using Discovery Studio 2020 (34) and PyMOL (www.pymol.org) (35). The graphs for the fluctuation were plotted using Prism 10 (www.graphpad.com). To build the three-dimensional (3D) structures of the cytotoxic T-lymphocyte (CTL) and helper T-lymphocyte (HTL) epitopes within the vaccine construct, the PEPFOLD 3.5 web server was utilized (36).

2.13 In silico immune simulation

To model the immune response and assess the immunogenicity of the ALV vaccine in the host, we utilized the C-ImmSim server (https://kraken.iac.rm.cnr.it/C-IMMSIM/). The server can define a set of different models in one software that analyses both humoral and cellular responses including B-cells. The input parameters consisted of random seed, simulation volume, and simulation steps, while the expected outputs included parameters to configure the immune simulation, controlling randomness, size, and duration of the simulation. For this study, we configured the following parameters: Random Seed = 12,345, Simulation Volume = 10, and Simulation Steps = 1000. All other simulation parameters were maintained at their default settings to ensure consistency and reliability in the results.

3 Results

3.1 Protein sequences retrieval

The amino acid sequences of the RSV fusion (F) and attachment (G) glycoproteins were obtained from the UniProt knowledge base (UniProt Consortium, 2021). The F glycoprotein sequence consisted of 574 amino acid residues, while the attachment G glycoprotein sequence was 321 amino acids long. The molecular weights of the F and G glycoproteins were calculated to be 63,751 Daltons and 35,191 Daltons, respectively, based on their amino acid compositions. This information was also retrieved from the UniProt database (Supplementary Table S1).

3.2 Analysis of physicochemical properties of proteins

The physical and chemical characteristics of the RSV F and G glycoproteins were thoroughly analyzed to gain insights into their structural and functional properties. (Supplementary Table S2). The F glycoprotein, consisting of 574 amino acid residues, had a calculated molecular weight of 63,750.57 Daltons. The theoretical isoelectric point (pI) of the F protein was determined to be 9.13, indicating its basic nature. The grand average of hydropathicity (GRAVY) value, which represents the overall hydrophobicity/hydrophilicity of the protein, was -0.038, suggesting a slightly hydrophilic character. The aliphatic index, a measure of the relative volume occupied by aliphatic side chains, was found to be 102.18 for the F protein, indicating a relatively compact structure. The instability index was calculated to be 41.81, however, suggesting that the F protein may be unstable under certain conditions. The estimated coefficient value, a parameter used to predict the expression level of the protein, was determined to be 50,155. In contrast, the RSV G glycoprotein, with 321 amino acids, had a lower molecular weight of 35,190.86 Daltons. The pI value of the G protein was slightly higher than that of the F protein, at 9.77, reinforcing its basic character. Interestingly, the GRAVY value of the G protein was -0.636, indicating a more hydrophilic nature compared to the F protein. The aliphatic index and estimated coefficient values of the G protein were much lower than those of the F protein, at 68.38 and 20,190, respectively. However, the G protein was found to be more stable, with an instability index of 35.70, suggesting it may be less prone to degradation under various conditions.

3.3 Analysis of antigenicity and allergenicity of proteins

The RSV F and G glycoproteins were further analyzed to investigate their antigenic, allergenic, and toxic characteristics. Antigenic potential was assessed using a predictive algorithm, which measured the likelihood of a protein being recognized as

an antigen. The F glycoprotein exhibited an antigenic score of 0.5295, while the G glycoprotein had a score of 0.5771. Both values exceeded the commonly used threshold of 0.4, indicating that these RSV glycoproteins possess significant antigenic properties. To evaluate the potential allergenicity of the F and G proteins, appropriate prediction models were employed. The analysis revealed that neither the F nor the G glycoprotein exhibited characteristics associated with allergenic proteins. This suggests that these RSV proteins are unlikely to elicit allergic responses. The proteins were also evaluated for potential toxic effects. The assessment did not identify any toxicity-related features within the amino acid sequences of the F and G glycoproteins. The findings from these bioinformatics analyses indicate that the RSV fusion (F) and attachment (G) glycoproteins have strong antigenic potential, which may contribute to their ability to stimulate immune responses. The absence of predicted allergenic and toxic properties suggests that these viral proteins are unlikely to cause adverse reactions or toxicity in the host.

3.4 CTL epitope prediction and binding affinity analysis with MHC I allele

The RSV F and G glycoproteins were further analyzed to identify specific epitopes with desired immunological and safety properties. The selected epitopes were evaluated for their antigenic, immunogenic, allergenic, and toxic characteristics. For the F protein, the following HLA-I class epitopes were chosen for detailed analysis: LTLAINALY, LSALRTGWY, and YTSVITIEL. All three epitopes exhibited antigenic and immunogenic properties and were found to be non-toxic (Table 1). However, only the LSALRTGWY epitope was predicted to be non-allergenic. Similarly, the G protein HLA-I class epitopes selected were: LLFISSCLY, SQVHTTSEY, and TTSQSTTIL. These G protein epitopes were identified as potential allergens. The SQVHTTSEY epitope was determined to be both immunogenic and antigenic. Further analysis focused on F protein epitopes for the H-2-IAd MHC class. The selected epitopes were: MELLIHRSSAIFLTL, LLIHRSSAIFLTLAI, and ELLIHRSSAIFLTLA. All three epitopes demonstrated immunogenic, antigenic, and non-toxic properties. Only the ELLIHRSSAIFLTLA epitope was predicted to be an allergen, while the other two were classified as non-allergenic.

The G protein H-2-Iad MHC class epitopes examined were: NLKSIAQITLSILAM, KLNLKSIAQITLSIL, and LYKLNLKSIAQITLS. These epitopes were found to be antigenic and non-toxic, but non-immunogenic. NLKSIAQITLSILAM and LYKLNLKSIAQITLS were identified as potential allergens. For the H-2-Iab MHC class, the selected F protein epitopes were: FYQSTCSAVSRGYLS, GVGSAIASGIAVSKV, and TREFSVNAGVTTPLS. All these epitopes were determined to be non-toxic and non-allergenic. However, only the TREFSVNAGVTTPLS epitope was antigenic, while FYQSTCSAVSRGYLS and GVGSAIASGIAVSKV were non-immunogenic. The G protein H-2-Iab MHC class epitopes analyzed were: IAAIIFIASANHKVT, AIIFIASANHKVTLT, and IIFIASANHKVTLTT. These epitopes exhibited antigenic and non-toxic properties, and IAAIIFIASANHKVT and AIIFIASANHKVTLT were also immunogenic. Lastly, for the H-2-Db MHC class, the F protein epitopes selected were: YMLTNSELL and VSLSNGVSV. Both epitopes were found to be non-toxic and non-immunogenic. However, the YMLTNSELL epitope was classified as non-antigenic and non-allergenic, while the VSLSNGVSV epitope was antigenic but allergenic. The G protein H-2-Db MHC class epitopes examined were LAMIISTSL and AMIISTSLI. These epitopes were determined to be non-toxic and non-immunogenic but were predicted to be allergenic.

3.5 HTL epitope prediction and binding affinity analysis with MHC2 allele

RSV F and G glycoproteins were further analyzed to identify additional epitopes with desirable immunological and safety characteristics, focusing on the H-2-IAb and H-2-Iad MHC class contexts (Table 2). For the H-2-Iab MHC class, the following F protein epitopes were selected for analysis: GVGSAIASGIAVSKV, TREFSVNAGVTTPLS, and EFSVNAGVTTPLSTY. All three epitopes were found to be non-toxic and non-allergenic. However, only TREFSVNAGVTTPLS and EFSVNAGVTTPLSTY were determined to be immunogenic. The G protein epitopes examined for the H-2-Iab MHC class were: IAAIIFIASANHKVT, IIFIASANHKVTLTT, and AIIFIASANHKVTLT. All three of these epitopes exhibited antigenic, non-toxic, and non-allergenic properties. For the H-2-Iad MHC class, the selected F protein epitopes were: MELLIHRSSAIFLTL, GVGSAIASGIAVSKV, and AIASGIAVSKVLHLE. All of these epitopes were found to be antigenic, non-toxic, and non-allergenic. The G protein epitopes analyzed for the H-2-Iad MHC class were NLKSIAQITLSILAM, AAIIFIASANHKVTL, and KLNLKSIAQITLSIL. These epitopes were all identified as antigenic and non-toxic. However, only the NLKSIAQITLSILAM epitope was predicted to be an allergen, while AAIIFIASANHKVTL and KLNLKSIAQITLSIL were classified as non-allergenic.

3.6 Prediction Of B-cell epitopes

The IEDB (Immune Epitope Database) server was utilized to analyze the F and G protein and identify potential epitopes. The epitopes that exceeded the 0.5 threshold were then evaluated for their allergenicity, antigenicity, immunogenicity, and toxicity characteristics. The epitope that exhibited the highest score in the IEDB analysis was deemed the most promising candidate for further study. Based on the comprehensive evaluation using the IEDB server, a subset of F and G protein epitopes was selected for further analysis due to their potential to induce a B-cell response (SF1 and SF2). The selected F protein epitopes were ETKCNGTDT, KCTASNKN, and NTPVTLS. All three were found to be antigenic. However, only the NTPVTLS epitope was determined to be both immunogenic and non-toxic, making it the most promising

10.3389/fimmu.2025.1546254

TABLE 1 List of overall attributes of MHC class I interacting CTL epitopes that were employed for designing a vaccine construct.

MHC-I Allele	Epitope	Protein	Length	Immunogenicity score	Antigenicity Score	Allergenicity	Toxicity
HLA1	LTLAINALY	Fusion Glycoprotein	9	0.18582	0.7147	Allergen	Non-Toxin
	LSALRTGWY	Fusion Glycoprotein	9	0.2465	1.1132	Non-allergen	Non-Toxin
	YTSVITIEL	Fusion Glycoprotein	9	0.3248	0.6842	Allergen	Non-Toxin
	LLFISSCLY	Attachment Glycoprotein	9	-0.19689	0.1194	Allergen	Toxin
	SQVHTTSEY	Attachment Glycoprotein	9	0.03917	0.6339	Allergen	Non-Toxin
	TTSQSTTIL	Attachment Glycoprotein	9	-0.1843	0.3578	Allergen	Non-Toxin
HLA II/H-2-IAd	MELLIHRSSAIFLTL	Fusion Glycoprotein	15	0.16576	0.4136	Non-allergen	Non-Toxin
	LLIHRSSAIFLTLAI	Fusion Glycoprotein	15	0.12489	0.6449	Non-allergen	Non-Toxin
	ELLIHRSSAIFLTLA	Fusion Glycoprotein	15	0.18756	0.4213	Allergen	Non-Toxin
	NLKSIAQITLSILAM	Attachment Glycoprotein	15	-0.04729	0.9619	Allergen	Non-Toxin
	KLNLKSIAQITLSI	Attachment Glycoprotein	14	-0.26934	1.1468	Non-allergen	Non-Toxin
	LYKLNLKSIAQITLS	Attachment Glycoprotein	15	-0.2629	0.9181	Allergen	Non-Toxin
HLA II/H-2-IAb	FYQSTCSAVSRGYLS	Fusion Glycoprotein	15	-0.38997	0.5709	Non-allergen	Non-Toxin
	GVGSAIASGIAVSKV	Fusion Glycoprotein	15	-0.13683	0.6023	Non-allergen	Non-Toxin
	TREFSVNAGVTTPLS	Fusion Glycoprotein	15	0.1929	0.3074	Non-allergen	Non-Toxin
	IAAIIFIASANHKVT	Attachment Glycoprotein	15	0.30119	0.6127	Non-allergen	Non-Toxin
	AIIFIASANHKVTLT	Attachment Glycoprotein	15	0.08604	0.7845	Allergen	Non-Toxin
	IIFIASANHKVTLTT	Attachment Glycoprotein	15	-0.00772	0.6941	Non-allergen	Non-Toxin
H-2-Db	YMLTNSELL	Fusion Glycoprotein	9	-0.04855	0.2930	Non-allergen	Non-Toxin
	VSLSNGVSV	Fusion Glycoprotein	9	-0.20629	0.8926	Allergen	Non-Toxin
	LAMIISTSL	Attachment Glycoprotein	9	-0.01311	0.5518	Allergen	Non-Toxin
	AMIISTSLI	Attachment Glycoprotein	9	-0.09354	0.3295	Allergen	Non-Toxin

candidate from this group for further investigation. Similarly, a set of G protein epitopes was selected for analysis: LSGTTSQST, MSKTKDQRTAKT, and TNQIKNTTPTYLTQN. All three G protein epitopes were identified as antigenic and non-toxic. However, none were predicted to be immunogenic (Table 3).

3.7 Prediction of interferon-gamma-inducing epitopes

In addition to the B-cell response-inducing epitopes, the analysis also identified a set of interferon-gamma-inducing F protein epitopes that were selected for further investigation. These epitopes were LPIGAVSIVAIALLL, IGAVSIVAIALLLRL, and PIGAVSIVAIALLLR. All three were found to be antigenic, immunogenic, and non-toxic, making them promising candidates for inclusion in a multi-epitope vaccine construct. The analysis also identified a single G protein epitope, TNQIKNTTPTYLTQN, that was also selected for further consideration (Table 4).

3.8 Multi-epitope vaccine design

Based on the detailed analysis and evaluation of the F and G protein epitopes, a multi-epitope vaccine construct was designed that met the criteria for antigenicity, allergenicity, toxicity, and

10.3389/fimmu.2025.1546254

TABLE 2 List of overall attributes of MHC class II interacting HTL epitopes used for designing a vaccine construct.

MHC-II Allele	Epitope	Protein	Length	Immunogenicity Score	Antigenicity Score	Allergenicity	Toxicity
HLA II/H-2-IAb	GVGSAIASGIAVSKV	Fusion Glycoprotein	15	-0.13683	0.6023	Non-allergen	Non-Toxin
	TREFSVNAGVTTPLS	Fusion Glycoprotein	15	0.1929	0.3074	Non-allergen	Non-Toxin
	EFSVNAGVTTPLSTY	Fusion Glycoprotein	15	0.03026	0.2190	Non-allergen	Non-Toxin
	IAAIIFIASANHKVT	Attachment Glycoprotein	15	0.30119	0.6127	Non-allergen	Non-Toxin
	IIFIASANHKVTLTT	Attachment Glycoprotein	15	-0.00772	0.6941	Non-allergen	Non-Toxin
	AIIFIASANHKVTLT	Attachment Glycoprotein	15	0.08604	0.7845	Allergen	Non-Toxin
HLA II/H-2-IAd	MELLIHRSSAIFLTL	Fusion Glycoprotein	15	0.16576	0.4136	Non-allergen	Non-Toxin
	GVGSAIASGIAVSKV	Fusion Glycoprotein	15	-0.13683	0.6023	Non-allergen	Non-Toxin
	AIASGIAVSKVLHLE	Fusion Glycoprotein	15	-0.23339	0.8407	Non-allergen	Non-Toxin
	NLKSIAQITLSILAM	Attachment Glycoprotein	15	-0.04729	0.9619	Allergen	Non-Toxin
	AAIIFIASANHKVTL	Attachment Glycoprotein	15	0.21554	0.6395	Non-allergen	Non-Toxin
	KLNLKSIAQITLSIL	Attachment Glycoprotein	15	-0.23436	1.1468	Non-allergen	Non-Toxin

population coverage. The final vaccine construct comprised a sequence of 315 amino acid residues, which incorporated non-overlapping epitopes selected from the F protein. The vaccine design included 11 cytotoxic T lymphocyte (CTL) epitopes, six helper T lymphocyte (HTL) epitopes, and three B-cell-inducing epitopes. To facilitate the appropriate presentation and processing of the different epitope types, specific linker sequences were utilized to connect the epitopes within the multi-epitope vaccine construct. The CTL epitopes were joined using Ala-Ala-Tyr (AAY) linkers, which are known to enhance CD8+ T cell activation and antigen processing. The HTL epitopes were connected by Gly-Pro-Gly-Pro-Gly (GPGPG) linkers, a flexible linker sequence that allows for optimal presentation of the helper T cell epitopes. The B-cell-inducing epitopes were linked using KK Lys-Lys linkers, which have been shown to improve B-cell recognition and antibody production. The strategic arrangement of the F protein's different epitope types, along with the incorporation of the selected linker sequences, is depicted in Figure 1A. Similarly, a 317 amino acid residue long G protein multi-epitope vaccine construct included 11 cytotoxic T lymphocyte (CTL) epitopes, 6 helper T lymphocyte (HTL) epitopes, and 3 B-cell-inducing epitopes. The same linker strategies used for the F protein epitopes were also applied to the G protein epitopes to facilitate appropriate presentation and processing. Specifically, the CTL epitopes were joined using AAY linkers, the HTL epitopes were connected by GPGPG linkers, and the B-cell-inducing epitopes were linked using KK linkers. The visual representation of the multi-epitope vaccine construct, including the arrangement and linkage of the epitopes from G protein, is shown in Figure 1B.

3.9 Evaluation of physical properties of vaccine construct

After constructing the multi-epitope vaccines based on the F and G proteins, their physicochemical properties were determined

TABLE 3 List of overall attributes of B cell epitopes that were employed to design a vaccine construct.

Protein	Epitope	Length	Immunogenicity	Antigenicity	Allergenicity	Toxicity
Fusion Glycoprotein	ETKCNGTDT	9	-0.05293	0.9375	Non-allergen	Toxin
	KCTASNKN	8	-0.29722	1.2880	Allergen	Toxin
	NTPVTLS	7	0.0653	1.1172	Allergen	Non-Toxin
Attachment Glycoprotein	LSGTTSQST	9	-0.26229	0.7385	Non-allergen	Non-Toxin
	MSKTKDQRTAKT	12	-0.33818	0.6219	Non-allergen	Non-Toxin
	TNQIKNTTPTYLTQN	15	-0.06634	0.5845	Allergen	Non-Toxin

10.3389/fimmu.2025.1546254

TABLE 4 List of overall attributes of Interferon-Gamma inducing F and G glycoprotein epitopes that were employed to design a vaccine construct.

Protein	Epitope	Position	Immunogenicity	Antigenicity	Allergenicity	Toxicity
Fusion Glycoprotein	LPIGAVSIVAIALLL	15	0.35299	1.1321	Non-allergen	Non-Toxin
	IGAVSIVAIALLLRL	15	0.26838	1.1321	Non-allergen	Non-Toxin
	PIGAVSIVAIALLLR	15	0.27879	1.0998	Allergen	Non-Toxin
Attachment Glycoprotein	TNQIKNTTPTYLTQN	15	-0.06634	0.5845	Allergen	Non-Toxin

and compared with the original F and G protein values reported in (Supplementary Table S3). The analysis revealed that the number of amino acids in the vaccine constructs was reduced to 315 and 317 for the F and G protein-based vaccines respectively, compared to the original full-length protein sequences. The molecular weight of the F protein-based vaccine construct decreased from 63,750.57 Da to 31,982.84 Da, while the G protein-based vaccine showed a slight decrease from 35,190.86 Da to 33,292.41 Da. The theoretical isoelectric point (pI) values of the F and G protein-based vaccine constructs were calculated to be 9.46 and 10.07, respectively. The grand average of hydropathicity (GRAVY) values for the F and G protein-based vaccine constructs were positive, at 0.448 and 0.598 respectively, indicating that the proteins are generally hydrophobic. The aliphatic index, which provides an estimate of the relative volume occupied by aliphatic side chains, was higher for the vaccine constructs compared to the original proteins, with values of 101.40 for the F protein-based vaccine and 117.22 for the G protein-based vaccine. The instability index, which predicts the stability of a protein, was less than 40 for both the F and G protein-based vaccine constructs, at 24.42 and 20.22, respectively, suggesting that both vaccine constructs are stable. The extinction coefficient values, which indicate the amount of light absorbed by a protein

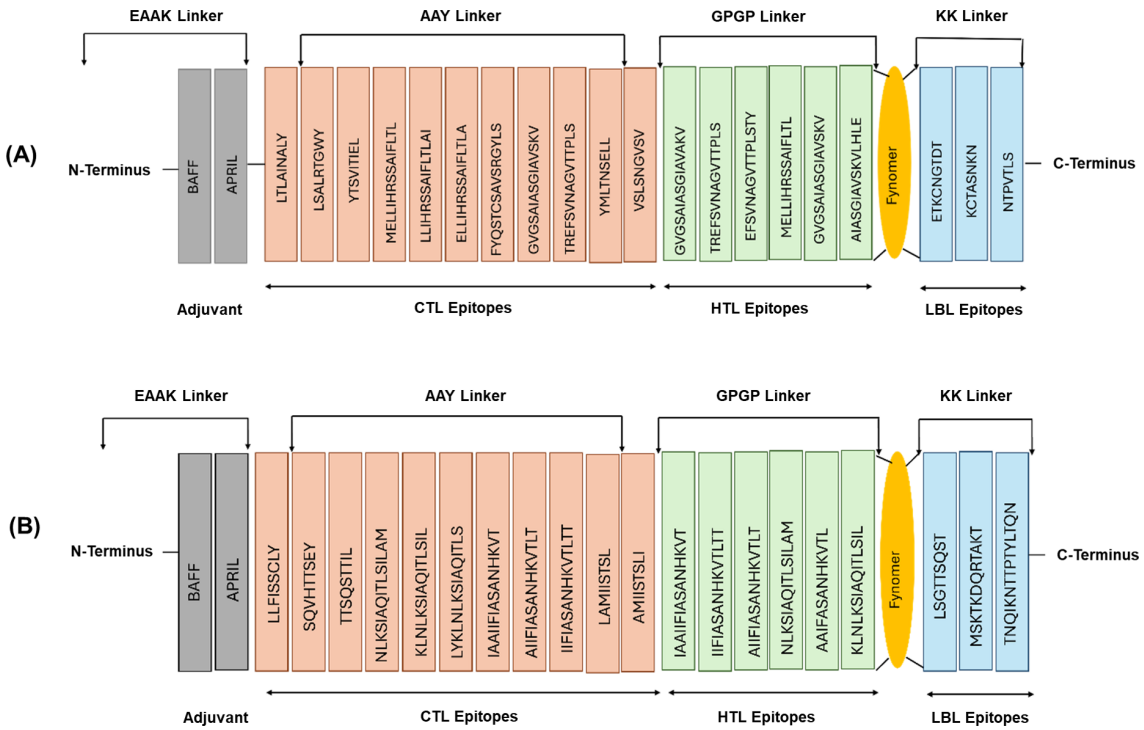


FIGURE 1
Multi-epitope vaccine construct of F glycoprotein **(A)** and G glycoprotein **(B)** with epitopes linked by different linkers. The orange color shows CTL epitopes interconnected by AAY linkers, the green color represents HTL epitopes interconnected by GPGP linkers, and the blue color shows LBL epitopes interconnected by KK linkers. BAFF and APRIL adjuvants are connected to the N-terminus via EAAK linkers.

solution, were calculated to be 30,955 M−1 cm−1 for the F protein-based vaccine and 20,860 M−1 cm−1 for the G protein-based vaccine. The antigenic values for the F and G protein-based vaccine constructs were 0.5996 and 0.6048 respectively, which are higher than the threshold of 0.4, confirming their antigenic potential. Both the F and G protein-based vaccine constructs were also assessed to be non-allergenic. Overall, the physicochemical characterization of the multi-epitope vaccine constructs demonstrated favorable properties, including reduced molecular weight, improved stability, and retained antigenicity, compared to the original F and G proteins, indicating their suitability for further development and evaluation as potential respiratory syncytial virus vaccine candidates.

3.10 Evaluation of antigenic and allergenicity properties of the vaccine constructs

The antigenicity of the vaccine constructs derived from the F (fusion) and G (attachment) proteins was predicted using the default settings in the antigenicity prediction tool, with a threshold value of 0.4. The overall antigenicity prediction score for the F protein-based vaccine construct was 0.5996, while the score for the G protein-based vaccine construct was 0.6048. Both of these scores exceeded the 0.4 thresholds, indicating that the vaccine constructs derived from the F and G proteins were likely to be "Probable ANTIGENS". This antigenicity analysis revealed that both the F and G protein-based vaccine constructs exhibited strong antigenic potential. Allergen and AllerTOP tools revealed that the vaccine against F protein and G protein was "Probable Non-Allergen" (Data not shown).

The TMHMM server was used to analyze the presence and distribution of transmembrane helices in the vaccine constructs, derived from the F and G glycoproteins. For the F glycoprotein vaccine construct, the analysis revealed the presence of two transmembrane helices. The amino acids were distributed as follows: 1) Outside region: Amino acids 1-19 and 78-315; transmembrane helices (purple); Amino acids 20-42 and 55-77. 2) Inside region (between transmembrane helices): Amino acids 43-54. This distribution indicated that the majority of the amino acids in the F protein-based vaccine construct were located in the outside region, which is the extracellular domain of the protein. The SignalP server was used to analyze the signal peptide and cleavage site predictions. For the F glycoprotein vaccine construct, the analysis revealed that the C-score showed a distinct peak at the 23rd amino acid position, indicating the predicted cleavage site. S-score (signal peptide score graph showed the presence of a signal peptide sequence. The Y-score which combines the C-score and S-score, also reached a maximum at the 23rd amino acid position, further confirming the predicted cleavage site. These results suggest that the F glycoprotein vaccine construct is likely to be cleaved at the 23rd amino acid position, resulting in the removal of the signal peptide

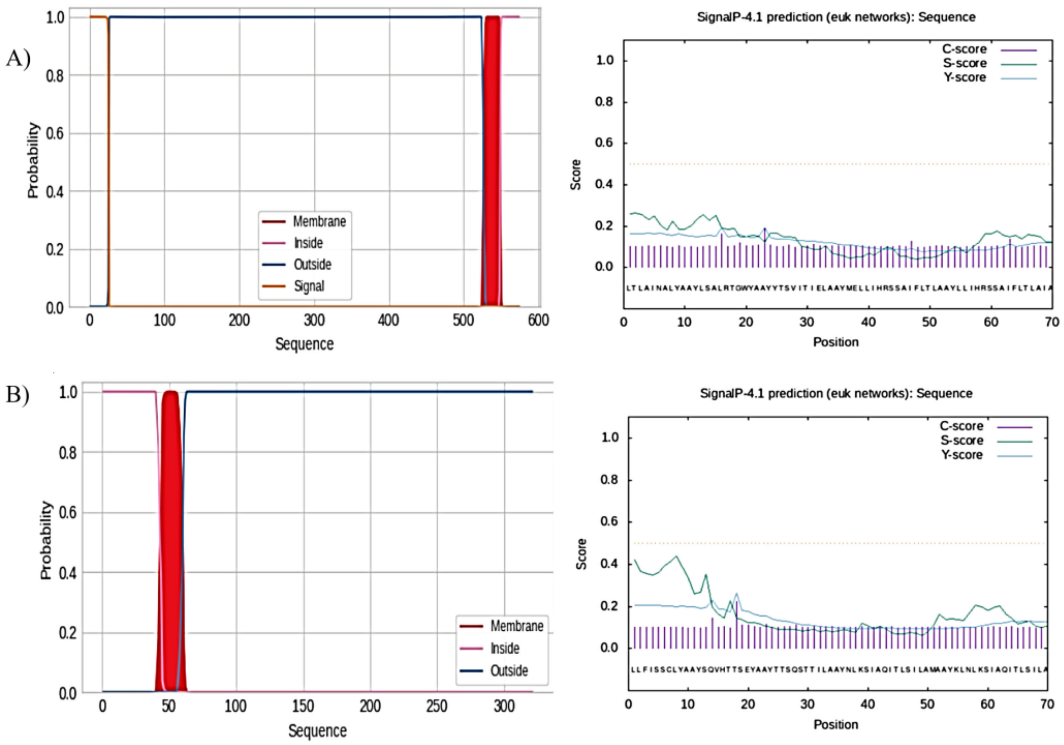


FIGURE 2
TMHMM server for the prediction of the nature of amino acid residues. **(A)** Nature of amino acid residues of F protein and prediction of the presence of signal peptide on F proteins. **(B)** Nature of amino acid residues of G protein and prediction of the presence of signal peptide on G proteins.

and the presentation of the mature, processed form of the antigen (Figure 2A).

Similarly, the analysis of the G glycoprotein vaccine construct revealed the presence of five transmembrane helices. The amino acid distribution was as follows: 3) Outside region: Amino acids 1-31, 99-107, and 166-234; transmembrane helices (purple); amino acids 32-54, 76-98, 108-130, 143-165, and 235-257. 4) Inside region (between transmembrane helices): Amino acids 55-75, 131-142, and 258-317 Again, most of the amino acids in the G protein-based vaccine construct were in the outside region, which is the extracellular domain of the protein. For the G glycoprotein vaccine construct, the SignalP analysis showed that the Cleavage Site Score (C-score) peaked at the 18th amino acid position, indicating the predicted cleavage site. The Signal Peptide Score (S-score) graph suggested the presence of a signal peptide sequence. The Y-score also reached a maximum at the 18th amino acid position, corroborating the predicted cleavage site. These results suggest that the G glycoprotein vaccine construct is predicted to be cleaved at the 18th amino acid position, leading to the removal of the signal peptide and the exposure of the mature antigen (Figure 2B). We identified the signal peptide cleavage sites for both the F and G glycoprotein vaccine constructs. These signal peptide cleavage sites ensure the proper processing and presentation of the antigens to the immune system.

3.11 Prediction of secondary structure

The secondary structure of the F and G glycoprotein vaccine constructs was predicted, using the PSIPRED algorithm. For the F glycoprotein vaccine construct, the analysis revealed that the secondary structure composition contained 23.7% helices, 13.70% strands, and 34.67% coils. The results showed that the F protein-based vaccine construct is predominantly composed of coil regions, with a significant proportion of helical structures and a smaller fraction of beta-strand regions. The analysis of the G glycoprotein vaccine construct showed a similar trend, with the secondary structure dominated by helical elements, followed by coils and strands. A helical structure was observed in both the F and G protein-based vaccine constructs which play a crucial role in maintaining the native-like conformation of proteins and preserving the integrity of important functional epitopes (SF3 and SF4).

3.12 Three-dimensional structural modeling, interaction, and stability

We predicted the three-dimensional structure of the vaccine constructs namely, F1, F2, F3 for F-glycoprotein and G1, G2, G3 for G-glycoprotein. The structure modeling was performed using the latest artificial intelligence (AI) and machine learning (ML) based algorithm Alphafold3 webserver (31). The TLR1 toll-like receptor sequence for Mus musculus was retrieved using the UniProt database (Uniprot ID-B9EJ46). The structure for the TLR1 receptor was also predicted using the Alphafold3 server. All models obtained were of high quality with plDDT scores for all reported to be >70. Further, we wanted to test if the vaccine candidates would bind and, in turn, block the TLR receptors. Hence, the 3D models for all proteins (TLR and vaccine candidates) were then used to predict TLR: vaccine complexes. The protein: protein docking simulation models were performed using the ClusPro docking algorithm (32).

We found that all vaccine candidates could occupy the interaction binding pocket on the TLR receptor. The interactions were mostly charged where critical positive and negative amino acids formed the salt bridges and combined with pi-pi interaction through bulky hydrophobic residues. The interaction is shown for TLR1:F1 (Figure 3A top), TLR1:F2 (Figure 3B top), TLR1:F3 (Figure 3C top), TLR1:G1 (Figure 4A top), TLR1:G2 (Figure 4B top), TLR1:G3 (Figure 4C top). Important interaction residues found are listed in Supplementary Table S4.

We then verified the stability of the TLR:vaccine complex using dynamics and stability analysis using CabsFlex 2.0 standalone (30). We found the TLR: vaccine candidates to be very stable and had the per residue fluctuation (RMSF) within the required limits through time. The RMSF plots are shown for TLR1:F1 (Figures 3B, C, TLR1: F2 (Figures 3B, C), TLR1:F3(Figures 3B, C), TLR1:G1 (Figures 4B, C), TLR1:G2 (Figures 4B, C), TLR1:G3 (Figures 4B, C). Though all vaccine candidates were within the allowed RMSF limit, the best ones found were G1 and G2, with the least residue fluctuations. The higher fluctuations observed at the C-terminal ends of all the vaccine candidates are due to the extended unstructured region on the candidates, which includes the 6xHistag. Overall, we found all predicted vaccine candidates to be well-folded and specifically targeting the TLR domains.

3.13 Immune stimulation

The immune simulation results showed a significant increase in the primary, secondary, and tertiary immune responses, corresponding with a reduction in antigen concentration (Figure 5). The levels of IL-2 were found to align with the measure of diversity, indicating a robust immune activation. Furthermore, an increase in diversity over time is interpreted as a danger signal, particularly in conjunction with the presence of leukocyte growth factor. Thus, a lower measure of diversity value reflects diminished immune diversity, suggesting potential implications for the effectiveness of the immune response.

4 Discussion

Over the past few years, significant resources and efforts have been dedicated to developing a safe and effective vaccine against hRSV, a major respiratory pathogen. Natural RSV infection fails to provide lasting immunity, leading to multiple infections throughout an individual's life. Consequently, designing a vaccine that effectively mimics the immune response generated by natural

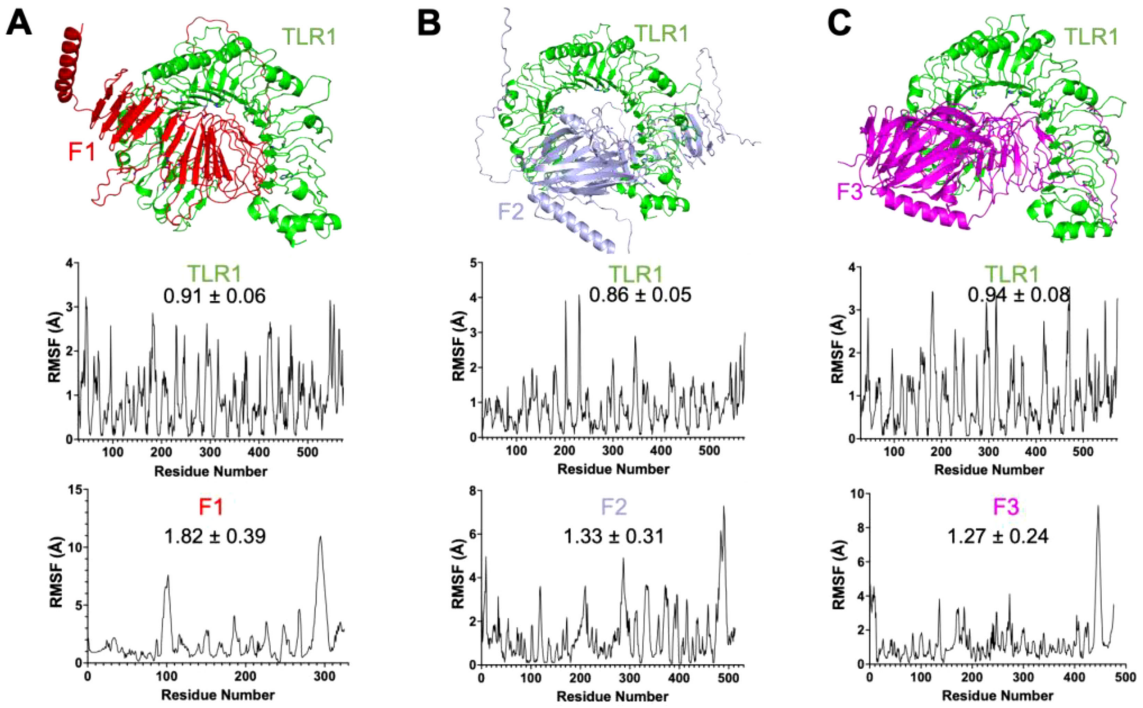


FIGURE 3
Three-dimensional protein complex between [**(A)**: Top] TLR1 domain (green) and F1 vaccine candidate (red), [**(B)**: Top] TLR1 domain (green) and F2 vaccine candidate (slate-blue) and [**(C)**: Top] TLR1 domain (green) and F3 vaccine candidate (magenta) shown in cartoon representation. Root mean square fluctuation (RMSF) plots for the interaction between TLR1 and vaccine candidates are below the respective structural representation arranged accordingly. The RMSF value is also mentioned on the plot with n=3. Interaction residues also mentioned in the Supplementary Table are shown as sticks.

RSV infection, while accounting for the variability among different viral strains remains a substantial challenge for researchers (37).

The emergence of diverse vaccine candidates utilizing various technologies presents an opportunity to tailor immunization strategies to meet the specific needs of vulnerable age groups. Arexvy® (GSK) and Abrysvo® (Pfizer) are significant advancements in this area, being the first vaccines approved to prevent hRSV infections in older adults (3). Notably, Abrysvo® extends its utility by offering passive immunization for infants through maternal administration during pregnancy, thereby providing dual protection for both mothers and their newborns (38). Utilizing the mRNA platform, Moderna received U.S. FDA approval for the RSV vaccine mRESVIA(mRNA-1345). These approaches underscore the importance of developing age-specific vaccine strategies that can effectively address the unique immunological challenges faced by different populations (3, 38). These innovative vaccines pave the way for more personalized and effective immunization programs against hRSV (38).

Recent advancements in computational biology, immunoinformatics, and reverse vaccinology hold promise for accelerating the development of safe and effective vaccines in a more time- and cost-efficient manner (39, 40). By leveraging genomic and proteomic data, we can identify potential epitopes and design vaccines with immunogenic subunits that elicit long-lasting immunity, facilitating the validation of these candidates in preclinical settings (41, 42). In this study, we employ immunoinformatic approaches to identify key B cells, cytotoxic T lymphocyte (CTL), and helper T lymphocyte (HTL) epitopes derived from F and G proteins of hRSV, to develop a highly safe, synthetic multi-component vaccine tailored for the human host.

In this study, we employed several complementary tools to ensure a robust analysis of potential epitopes derived from the F and G glycoproteins of the hRSV. For Cytotoxic T Lymphocyte (CTL) epitope prediction, we utilized both the NetCTL v1.2 and the Immune Epitope Database (IEDB) tools. NetCTL focuses on predicting CTL epitopes based on their binding affinities to MHC class I molecules, providing a quantitative measure of potential immunogenicity. Meanwhile, IEDB offers additional validation through a comprehensive database of experimentally confirmed epitopes, enhancing the reliability of our findings. For Helper T Lymphocyte (HTL) epitope prediction, we employed NetMHCII pan 3.287 alongside IEDB, allowing us to cross-verify predictions and bolster confidence in the selected epitopes for further analysis. This dual approach is critical, as it ensures that our predicted HTL epitopes are not only computationally validated but also supported by empirical data. Additionally, for B-cell epitope prediction, we relied on IEDB for preliminary assessments while further investigating the epitopes through VaxiJen v2.0 and AllergenFP v1.0. This multifaceted approach enabled us to evaluate the antigenicity and allergenic potential of the identified epitopes comprehensively, laying a solid foundation for the design of a safe and effective multi-epitope vaccine against hRSV.

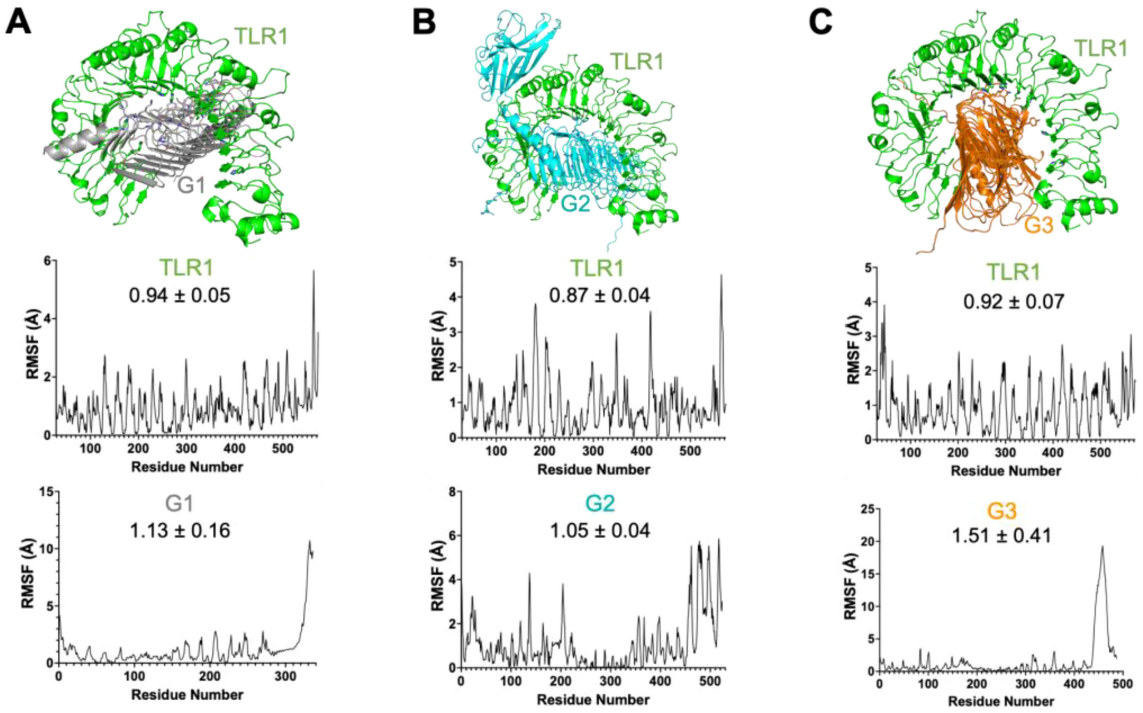


FIGURE 4
Three-dimensional protein complex between [**(A)**: Top] TLR1 domain (green) and G1 vaccine candidate (gray), [**(B)**: Top] TLR1 domain (green) and G2 vaccine candidate (cyan) and [**(C)**: Top] TLR1 domain (green) and G3 vaccine candidate (orange) shown in cartoon representation. Root means square fluctuation (RMSF) plots for the interaction between TLR1 and vaccine candidates are below the respective structural representation arranged accordingly. The RMSF value is also mentioned on the plot with n=3. Interaction residues also mentioned in the Supplementary Table are shown as sticks.

An emerging and important field of research is multi-epitope vaccines. Multi-epitope vaccinations offer the benefit of reducing undesirable effects, such as allergies and antigenic load. This results in a more specific immune response toward conserved epitopes without the reversion of the pathogenesis of the virus. The combined effect of the present epitopes from various antigens exceeds an isolated antigen epitope's ability to stimulate an immune response, including both humoral and cell-mediated responses. Multi-epitope vaccinations have been developed to limit a diverse range of diseases (43). From a pharmacological perspective, multi-epitope vaccinations exhibit advantageous characteristics. Multi-epitope vaccines can be effectively and economically generated due to their focus on chemically well-characterized peptides. The multi-epitope vaccination can protect a broad spectrum of pathogens or different strains of a certain pathogen, particularly for highly adaptable pathogens that undergo many mutations and give rise to new variations (44).

The physicochemical characterization of F and G protein-based multi-epitope vaccines revealed a reduction in the size of 574-amino-acid-long F glycoprotein (molecular weight: 63,750.57 Da) to 315 amino acids (molecular weight: 31,982.84 Da), and 321-amino acid-long G glycoprotein (molecular weight: 35,190.86 Da) to 317 amino acids (molecular weight: 33,292.41 Da). The pI values of 9.46 and positive GRAVY value of 0.448 for F protein-based vaccine constructs, along with the pI value of 10.07, and positive GRAVY values of 0.598 for G protein-based vaccine constructs, indicated the basic and hydrophobic nature of these vaccines.

Additionally, the higher aliphatic index, with values above 100, and the instability index values below 40 for both the F and G protein-based vaccine constructs, highlight the significant relative volume occupied by aliphatic side chains and the stability of these protein-based vaccine constructs. These favorable physicochemical properties, compared to the original F and G proteins indicate their potential suitability for further development and evaluation as hRSV vaccine candidates. The findings of this research align with a previous study, which reported similar physiochemical stability in multi-epitope-based vaccine design against hRSV (45).

The F and G protein-based vaccine constructs demonstrated higher antigenicity, with values of 0.5996 and 0.6048, respectively. AllerTOP analysis further confirmed the non-allergic nature of the selected proteins for the hRSV vaccine. This antigenic and non-allergenic profile suggests that the vaccine constructs have the potential to stimulate an active immune response against the hRSV without triggering allergic reactions in humans. Consequently, these proteins are promising candidates for developing a vaccine against hRSV.

A similar study reported the antigenic and non-allergic properties of the multi-epitope vaccine candidates using the AntigenPro and Vaxijen servers (46). Furthermore, research on RSV on structural proteins, such as MHC II, 3 B-cell epitopes, and

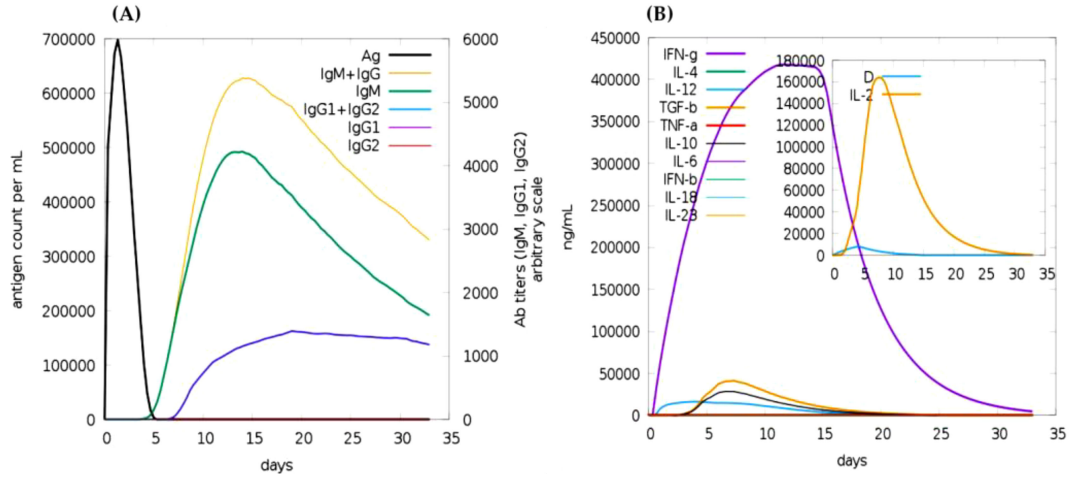


FIGURE 5
Immune simulation of the predicted vaccine following two injections via the c-immsim server. **(A)** Immunoglobulin production in response to antigen injections, with specific subclasses represented in different colors. **(B)** Cytokine secretion induced by the vaccine highlights IL-2 levels and a measure of diversity.

6MHC-I revealed their antigenic and non-allergic nature. These characteristics demonstrated their ability to stimulate immune responses and prevent viral replication (47, 48). The distribution of transmembrane helices in the F and G glycoproteins was analyzed using the TMHMM server, revealing that the proteins consist of 315 and 317 amino acids, respectively. The F protein contains two transmembrane helices, with 256 of its amino acids in the extracellular region. In contrast, the G protein has five transmembrane helices, with 107 residues in its extracellular region. These results suggest that most amino acids in these protein-based vaccine constructs are located in the extracellular domains of the proteins. The presence of transmembrane helices was similarly reported in a previous study involving the TMHMM server, which identified two transmembrane helices in the envelope protein (49).

The secondary structure analysis of the F glycoprotein-based vaccine construct revealed a predominant composition of coil regions (34.67%), followed by a significant proportion of helical structures (23.7%) and a smaller fraction of beta-strand regions (13.7%). A similar trend was observed in the G glycoprotein-based vaccine construct, with its secondary structure dominated by helical elements, followed by coils and strands. These findings are consistent with another study that reported a comparable pattern of secondary structures (45). The G and F glycoproteins play critical roles in the early stages of hRSV infection (6). Historically, determining whether the G protein was of viral or host origin posed challenges due to variations in the cell lines, virus strains, and protein detection technologies, all of which influenced the observed size and presence of the G protein. Notably, inhibiting the cleavage of the G-protein and incorporating it into a live attenuated RSV vaccine candidate could result in a virus with an intact G protein, leading to a 5-fold increase in infectivity for the nasal epithelium – the primary site of vaccine administration (50).

In combination with the secondary structure data, the three-dimensional modeling using the advanced AI-ML-based Alphafold3 method revealed that the vaccine candidates form well-folded proteins with optimal plDDT values. Molecular docking and dynamics refinement demonstrated that all vaccine candidates exhibit strong interactions with the TLR domain through charged and hydrophobic interactions, forming a tight bonds. The observed interaction patterns align with previously studied vaccine-TLR receptor complex models (51).

The F protein exhibits a higher degree of conservation compared to the G protein, making it the primary target for RSV development. The pre-fusion F protein is the main target of antibody neutralization in the sera of individuals who have experienced multiple RSV infections throughout their lifetime (52). Due to its capacity to elicit a higher concentration of neutralizing antibodies, most vaccine research has focused on the F protein. Prior infection and elevated levels of neutralizing antibodies, particularly those passed down from the mother, provide partial protection against the disease. Moreover, the use of a neutralized F protein mAbs in immunological prophylaxis underscores the critical role of the F protein in RSV vaccine development (53).

In this study, we strategically chose Toll-like receptors (TLRs) 2 and 4 due to their pivotal roles in immune recognition and their established potential as adjuvants, supported by robust literature. TLR2 and TLR4 are integral to the innate immune system, recognizing a diverse array of pathogen-associated molecular patterns (PAMPs) and playing a critical role in the initiation of immune responses (54). This foundational function is essential for

the effective development of vaccines, as their activation can significantly enhance the adaptive immune response. Furthermore, existing research underscores the efficacy of targeting these receptors in various vaccine strategies, affirming their relevance and effectiveness in enhancing vaccine efficacy (55). This comprehensive rationale underscores our decision to focus on TLR2 and TLR4, making these vaccine candidates achieve improved immune responses against hRSV.

5 Conclusions

This study highlights the promising potential of multi-epitope vaccines developed through immunoinformatics for combating hRSV. We have designed an F and G proteins-based synthetic vaccine that aims to elicit robust immune responses while minimizing adverse effects. The physicochemical characterization of the vaccine constructs indicates favorable properties, including stability and non-allergenic profiles, enhancing their suitability for further development. Additionally, it was demonstrated to stimulate immune responses in both cells and antibodies without triggering type 2 immunity, which are typically associated with RSV infection. This study highlights the potential of bioinformatics-based methods in developing effective therapies for emerging viruses, particularly under constraints such as restricted time and resources. However, these findings are derived from in silico computational analysis and must be validated through experimental studies with *in vivo* and *in vitro* models in laboratory settings. Overall, this research contributes to the ongoing efforts in vaccine innovation, paving the way for effective and safe immunization strategies against hRSV.

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.

Ethics statement

The Ethics Committee of King Fahad Medical City, Riyadh Second Health Cluster, Saudi Arabia (IRB No. 22-156 on 07.04.2022) approved the study.

Author contributions

HA: Data curation, Formal Analysis, Methodology, Writing – original draft. MA: Data curation, Methodology, Writing – review & editing. FA: Writing – review & editing, Funding acquisition, Resources, Supervision. IA: Formal Analysis, Validation, Writing – original draft. AM: Formal Analysis, Methodology, Writing – review & editing. WA: Formal Analysis, Writing – review & editing, Supervision. BA: Supervision, Conceptualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors extend their appreciation to the Researchers Supporting Project number (RSP2025R506), King Saud University, Riyadh, Saudi Arabia for supporting this work.

Acknowledgments

The authors would like to thank the Research Center at King Fahad Medical City for their valuable technical support.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1546254/full#supplementary-material

References

1. Shi T, McAllister DA, O'Brien KL, Simoes EA, 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

2. 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

3. Wildenbeest JG, Lowe DM, Standing JF, Butler CC. Respiratory syncytial virus infections in adults: a narrative review. *Lancet Respir Med*. (2024) 12:822–36. doi: 10.1016/S2213-2600(24)00255-8

4. Graham BS. Vaccine development for respiratory syncytial virus. *Curr Opin Virol*. (2017) 23:107–12. doi: 10.1016/j.coviro.2017.03.012

5. Sunita, Sajid A, Singh Y, Shukla P. Computational tools for modern vaccine development. *Hum Vaccines immunotherapeutics*. (2020) 16:723–35. doi: 10.1080/21645515.2019.1670035

6. McLellan JS, Ray WC, Peeples ME. Structure and function of respiratory syncytial virus surface glycoproteins. *Curr Topics Microbiol Immunol*. (2013) 372:83–104. doi: 10.1007/978-3-642-38919-1_4

7. Van Royen T, Rossey I, Sedeyn K, Schepens B, Saelens X. How RSV proteins join forces to overcome the host innate immune response. *Viruses*. (2022) 14:419. doi: 10.3390/v14020419

8. Nuñez Castrejon AM, O'Rourke SM, Kauvar LM, DuBois RM. Structure-based design and antigenic validation of respiratory syncytial virus G immunogens. *J Virol*. (2022) 96:e0220121. doi: 10.1128/jvi.02201-21

9. Su C, Zhong Y, Zhao G, Hou J, Zhang S, Wang B. RSV pre-fusion F protein enhances the G protein antibody and anti-infectious responses. *NPJ Vaccines*. (2022) 7:168. doi: 10.1038/s41541-022-00591-w

10. Palomo C, Mas V, Thom M, Vázquez M, Cano O, Terrón MC, et al. Influence of respiratory syncytial virus F glycoprotein conformation on induction of protective immune responses. *J Virol*. (2016) 90:5485–98. doi: 10.1128/JVI.00338-16

11. Liu J, Ostrowski M. Development of targeted adjuvants for HIV-1 vaccines. *AIDS Res Ther*. (2017) 14:43. doi: 10.1186/s12981-017-0165-8

12. Sicard T, Kassardjian A, Julien JP. B cell targeting by molecular adjuvants for enhanced immunogenicity. *Expert Rev Vaccines*. (2020) 19:1023–39. doi: 10.1080/14760584.2020.1857736

13. 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

14. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. *J Mol modeling*. (2014) 20:1–6. doi: 10.1007/s00894-014-2278-5

15. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. (2015) 43:D204–12. doi: 10.1093/nar/gku989

16. 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 N.J.)*. (1999) 112:531–52. doi: 10.1385/1-59259-584-7:531

17. Ong E, Wang H, Wong MU, Seetharaman M, Valdez N, He Y. Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics*. (2020) 36:3185–91. doi: 10.1093/bioinformatics/btaa119

18. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*. (2014) 30:846–51. doi: 10.1093/bioinformatics/btt619

19. 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

20. Nielsen H. Predicting secretory proteins with SignalP. *Protein Funct prediction: Methods Protoc*. (2017) p:59–73. doi: 10.1007/978-1-4939-7015-5_6

21. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M, et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf*. (2007) 8:1–12. doi: 10.1186/1471-2105-8-424

22. Kaabinejadian S, Barra C, Alvarez B, Yari H, Hildebrand WH, Nielsen M. Accurate MHC motif deconvolution of immunopeptidomics data reveals a significant contribution of DRB3, 4 and 5 to the total DR immunopeptidome. *Front Immunol*. (2022) 13:835454. doi: 10.3389/fimmu.2022.835454

23. 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.2018.154.issue-3

24. 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 Bioinf*. (2008) 9:1–8. doi: 10.1186/1471-2105-9-514

25. 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

26. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res*. (2008) 36:W513–8. doi: 10.1093/nar/gkn254

27. Srinivasan S, Selvaraj GF, Gopalan V, Padmanabhan P, Ramesh K, Govindan K, et al. Epitope identification and designing a potent multi-epitope vaccine construct against SARS-coV-2 including the emerging variants. *J Glob Infect Dis*. (2022). 14:24–30. doi: 10.4103/jgid.jgid_96_21

28. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallographica Section D: Biol Crystallogr*. (2002) 58:899–907. doi: 10.1107/S0907444902003451

29. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. (1993) 2:1511–9. doi: 10.1002/pro.5560020916

30. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. *Chem Rev*. (2016) 116:7898–936. doi: 10.1021/acs.chemrev.6b00163

31. Abramson J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. (2024) 630:493–500. doi: 10.1038/s41586-024-07487-w

32. Jones G, Jindal A, Ghani U, Kotelnikov S, Egbert M, Hashemi N, et al. Elucidation of protein function using computational docking and hotspot analysis by ClusPro and FTMap. *Acta Crystallographica Section D*. (2022) 78:690–7. doi: 10.1107/S2059798322002741

33. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. In *Algorithms in Bioinformatics: Second International Workshop, WABI*, Rome, Italy, September 17–21, 2002 Proceedings 2. Springer Berlin Heidelberg. (2002), pp. 185–200.

34. Jejurikar BL, Rohane SH. *Drug designing in discovery studio*. (2021), 135–8.

35. DeLano WL, Bromberg S. PyMOL user's guide. In: *DeLano scientific LLC*. (2004) 629.

36. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res*. (2016) 44:W449–54. doi: 10.1093/nar/gkw329

37. Zhao X, Yang J, He B, Li X, Yan H, Liu S, et al. A safe and effective mucosal RSV vaccine in mice consisting of RSV phosphoprotein and flagellin variant. *J Vaccine Res*. (2021) 45:113–24. doi: 10.1016/j.celrep.2021.109401

38. Topalidou X, Kalergis AM, Papazisis G. Respiratory syncytial virus vaccines: A review of the candidates and the approved vaccines. *Pathogens*. (2023) 12:1259. doi: 10.3390/pathogens12101259

Cherryholmes C, Yang J, He B, Li X, Yan H, Liu S, et al. Current methods of epitope identification for cancer vaccine design. *Vaccine*. (2015) 33:4067–77. doi: 10.1016/j.vaccine.2015.06.116

40. Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathog Global Health*. (2018) 112:181–8. doi: 10.1080/20477724.2018.1446773

41. Kuo TY, Lin MY, Coffman RL, Campbell JD, Traquina P, Lin YJ, Chen C, et al. Development of CpG-adjuvanted stable prefusion SARS-CoV-2 sp*ike antigen as a subunit vaccine against COVID-19*. *Sci Rep*. (2020) 10:20085. doi: 10.1038/s41598-020-77077-z

42. Pumchan A, Krobthong S, Roytrakul S, Sawatdichaikul O, Kondo H, Hirono I, et al. Novel chimeric multiepitope vaccine for streptococcosis disease in nile tilapia (Oreochromis niloticus linn.). *Sci Rep*. (2020) 10:603. doi: 10.1038/s41598-019-57283-0

43. Lu Q, Wu H, Meng J, Wang J, Wu J, Liu S, et al. Multi-epitope vaccine design for hepatitis E virus based on protein ORF2 and ORF3. *Front Microbiol*. (2024) 15. doi: 10.3389/fmicb.2024.1372069

44. Maleki A, Russo G, Parasiliti Palumbo GA, Pappalardo F. In silico design of recombinant multi-epitope vaccine against influenza A virus. *BMC Bioinf*. (2022) 22:617. doi: 10.1186/s12859-022-04581-6

45. 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

46. Dar HA, Almajhdi FN, Aziz S, Waheed Y. Immunoinformatics-aided analysis of RSV fusion and attachment glycoproteins to design a potent multi-epitope vaccine. *Vaccines*. (2022) 10:1381. doi: 10.3390/vaccines10091381

47. Cordeiro PAS, Assone T, Prates G, Tedeschi MRM, Fonseca LAM, Casseb J. The role of IFN-γ production during retroviral infections: an important cytokine involved in chronic inflammation and pathogenesis. *Rev do Instituto Medicina Trop São Paulo*. (2022) 64:e64. doi: 10.1590/s1678-9946202264064

48. Dittmer U, Peterson KE, Messer R, Stromnes IM, Race B, Hasenkrug KJ. Role of interleukin-4 (IL-4), IL-12, and gamma interferon in primary and vaccine-primed immune responses to Friend retrovirus infection. *J Virol*. (2001) 75:654–60. doi: 10.1128/JVI.75.2.654-660.2001

49. Hossain MS, Hossan MI, Mizan S, Moin AT, Yasmin F, Akash AS, et al. Immunoinformatics approach to designing a multi-epitope vaccine against Saint Louis Encephalitis Virus. *Inf Med Unlocked*. (2021) 22:100500. doi: 10.1016/j.imu.2020.100500

50. Corry J, Johnson SM, Cornwell J, Peeples ME. Preventing cleavage of the respiratory syncytial virus attachment protein in vero cells rescues the infectivity of progeny virus for primary human airway cultures. *J Virol*. (2016) 90:1311–20. doi: 10.1128/JVI.02351-15

51. Alturaiki W, Alkadi H, Alamri S, Awadalla ME, Alfaez A, Mubarak A, et al. Association between the expression of toll-like receptors, cytokines, and homeostatic chemokines in SARS-CoV-2 infection and COVID-19 severity. *Heliyon*. (2023) 9:e12653. doi: 10.1016/j.heliyon.2022.e12653

52. Hause AM, Henke DM, Avadhanula V, Shaw CA, Tapia LI, Piedra PA. Sequence variability of the respiratory syncytial virus (RSV) fusion gene among contemporary and historical genotypes of RSV/A and RSV/B. *PloS One*. (2017) 12:e0175792. doi: 10.1371/journal.pone.0175792

53. Anderson LJ, Jadhao SJ, Paden CR, Tong S. Functional features of the respiratory syncytial virus G protein. *Viruses*. (2021) 13:1214. doi: 10.3390/v13071214

54. Dowling JK, Mansell A. Toll-like receptors: the Swiss army knife of immunity and vaccine development. *Clin Transl Immunol*. (2016) 5:e85. doi: 10.1038/cti.2016.22

55. Martin WR, Cheng F. A rational design of a multi-epitope vaccine against SARS-CoV-2 which accounts for the glycan shield of the sp*ike glycoprotein*. *J Biomolecular Structure Dynamics*. (2022) 40:7099–113. doi: 10.1080/07391102.2021.1894986

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 14 April 2025
DOI 10.3389/fimmu.2025.1555741

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College & Hospital, India

REVIEWED BY
Maaweya Awadalla,
King Fahd Medical City, Saudi Arabia
Jyotirmayee Dey,
KIIT University, India

*CORRESPONDENCE
Xu Songxiao
xusx@zjcc.org.cn
Sher Muhammad
shermuhammad@superior.edu.pk

†These authors have contributed equally to this work and share authorship

RECEIVED 05 January 2025
ACCEPTED 25 March 2025
PUBLISHED 14 April 2025

CITATION
Dingding H, Muhammad S, Manzoor I, Ghaffar SA, Alodaini HA, Moubayed NMS, Hatamleh AA and Songxiao X (2025) Subtractive proteomics and reverse-vaccinology approaches for novel drug targets and designing a chimeric vaccine against *Ruminococcus gnavus* strain RJX1120. *Front. Immunol.* 16:1555741. doi: 10.3389/fimmu.2025.1555741

Subtractive proteomics and reverse-vaccinology approaches for novel drug targets and designing a chimeric vaccine against *Ruminococcus gnavus* strain RJX1120

Hou Dingding[1,2†], Sher Muhammad[3*†], Irfan Manzoor[4], Sana Abdul Ghaffar[4], Hissah Abdulrahman Alodaini[5], Nadine MS. Moubayed[5], Ashraf Atef Hatamleh[5] and Xu Songxiao[1*]

[1]Department of Clinical Laboratory, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, China, [2]Postgraduate Training Base Alliance of Wenzhou Medical University (Zhejiang Cancer Hospital), Hangzhou, Zhejiang, China, [3]Faculty of Agriculture and Veterinary Sciences, Superior University Lahore, Lahore, Pakistan, [4]Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan, [5]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia

Mediterraneibacter gnavus, also known as *Ruminococcus gnavus*, is a Gram-positive anaerobic bacterium that resides in the human gut microbiota. Notably, this bacterium plays dual roles in health and disease. On one side it supports nutrient metabolism essential for bodily functions and on the other it contributes to the development of Inflammatory Bowel Disease (IBD) and other gastrointestinal disorders. *R. gnavus* strain RJX1120 is an encapsulated strain and has been linked to develop IBD. Despite the advances made on its role in gut homeostasis, limited information is available on strain-specific virulence factors, metabolic pathways, and regulatory mechanisms. The study of such aspects is crucial to make microbiota-targeted therapy and understand its implications in host health. A multi-epitope vaccine against *R. gnavus* strain RJX1120 was designed using reverse vaccinology-based subtractive proteomics approach. Among the 3,219 proteins identified in the *R. gnavus* strain RJX1120, two critical virulent and antigenic proteins, a Single-stranded DNA-binding protein SSB (A0A2N5PT08) and Cell division ATP-binding protein FtsE (A0A2N5NK05) were screened and identified as potential targets. The predicted B-cell and T-cell epitopes from these proteins were screened for essential immunological properties such as antigenicity, allergenicity, solubility, MHC binding affinity, and toxicity. Epitopes chosen were cross-linked using suitable spacers and an adjuvant to develop a multi-epitope vaccine. Structural refinement of the construct revealed that 95.7% of the amino acid residues were located in favored regions, indicating a high-quality structural model. Molecular docking analysis demonstrated a robust interaction between the vaccine construct and the human Toll-like receptor 4 (TLR4), with a binding energy of –1277.0 kcal/mol. The results of molecular dynamics simulations further confirmed the stability of the vaccine-receptor complex under physiological conditions. *In silico* cloning of

the vaccine construct yielded a GC content of 48% and a Codon Adaptation Index (CAI) value of 1.0, indicating optimal expression in the host system. These results indicate the possibility of the designed vaccine construct as a candidate for the prevention of *R. gnavus*-associated diseases. However, experimental validation is required to confirm its immunogenicity and protective efficacy.

KEYWORDS

***Ruminococcus gnavus*, multi-epitope vaccine, reverse vaccinology, inflammatory bowel disease (IBD), subtractive proteomics, immunoinformatics**

Introduction

R. gnavus is a gram-positive anaerobic bacterium that is a key component of the human gut microbiota, playing significant roles in both health and disease (1). This bacterium is currently of interest due to its association with IBD and its ability to produce pro-inflammatory polysaccharides that modulate host immune responses (2). Among these strains, RJX1120 stands out for its role in gut inflammation and distinctive genetic characteristics (3). *R. gnavus* is mostly considered a commensal organism but often becomes a pathobiont in dysbiotic conditions and can be associated with diseases like Crohn's disease and ulcerative colitis (4). Its interactions with mucosal surfaces, the production of mucin-degrading enzymes, and its ability to produce immunomodulatory metabolites have been implicated in disease pathogenesis (5). The pathogenic potential of *R. gnavus* underscores the understanding of the virulence determinants and the molecular mechanisms (6). Currently, no licensed vaccine exists for *R. gnavus*, despite its association with inflammatory bowel disease (IBD) and other gastrointestinal disorders. This highlights an urgent need for novel vaccine strategies. Traditional vaccine development relies on culturing and isolating antigens, which is labor-intensive and time-consuming. In contrast, computational vaccine design offers a more efficient and targeted approach by identifying immunogenic proteins through reverse vaccinology and subtractive proteomics. Proteomic studies are very relevant for revealing the specializations of RJX1120 to unveil its potential therapeutic goals and thus help in generating preventive mechanisms (7).

R. gnavus exhibits significant adaptability and resilience in the human gut, which may contribute to its potential resistance to therapeutic interventions (2). This bacterium is known for its ability to degrade complex carbohydrates and mucins, producing metabolites such as short-chain fatty acids (e.g., propionate) that enhance its competitiveness and survival in the gut microbiota (8). In addition, the capacity to immunomodulate through immunogenic polysaccharides could enable this microbe to evade host immunity and promote long-term persistence in the gut (3). Although direct evidence for the presence of antimicrobial resistance mechanisms in *R. gnavus* is scanty, the effects of horizontal gene transfer of resistance genes are favored by its metabolic versatility (9). Understanding the mechanism of its resistance, such as its biofilm formation ability and possibly resistance-determining factors, is important to define suitable therapeutic strategies against pathogenic strains without disrupting the homeostasis of gut microbiota (10).

Subtractive proteomics represents a strategic approach for identifying pathogen-specific proteins important for survival but missing from the host proteome (11). In comparison, the unique set of proteins characteristic to *R. gnavus* points out the drug target candidates and vaccine targets (12). In subtractive proteomics, the subtraction of homologous proteins in the host background removes the critical proteins involved in pathogenicity in the bacterium but avoids the unwanted effects that would otherwise result in drug and vaccine therapies (11). A bioinformatics-driven approach called reverse vaccinology complements subtractive proteomics, antigenic proteins to be used in vaccines are analyzed by reverse vaccinology (11). Such proteins are likely to be surface-exposed, conserved among strains, and capable of inducing potent immune responses (12). Reverse vaccinology does not require isolation and culture of the pathogen, making vaccine discovery for complex organisms like *R. gnavus* significantly faster. Subtractive proteomics combined with reverse vaccinology offers an integrative approach to designing new vaccines and drugs against *R. gnavus*. The membrane and secreted proteins are the most relevant for vaccine development because these are exposed to the host immune system (13). Cytoplasmic proteins are the most suitable drug targets, as they are often critical to bacterial metabolism and survival (14). Using these approaches, multiepitope vaccine constructs can be designed by selecting epitopes from membrane-bound proteins that are non-allergenic, antigenic, and non-toxic (15). These vaccines can induce targeted immunity to pathogenic strains of *R. gnavus* and maintain the commensal balance of the gut microbiota.

In order to create a targeted vaccination against *R. gnavus* strain RJX1120, this study explores the pathogenic role of this bacteria in inflammatory bowel disease (IBD). The study finds possible vaccine candidates by searching the entire proteome of *R. gnavus* for essential, antigenic, and non-homologous proteins using a mix of subtractive proteomics and reverse vaccinology. Single-stranded DNA-binding protein (SSB) and cell division ATP-binding

protein FtsE are two important antigenic proteins that have been discovered as potential targets for vaccine development. B-cell and T-cell epitopes are included in a multi-epitope vaccination construct to produce a potent and defense-enhancing immune response. To assess the construct's stability, immunogenicity, and possible effectiveness, it is subjected to molecular docking with human Toll-like receptor 4 (TLR4), molecular dynamics simulations, and in silico immune response simulations. Codon optimization also guarantees effective expression in *Escherichia. coli*, which makes subsequent experimental validation easier. This study's main premise is that the multi-epitope vaccination will produce a strong immune response against *R. gnavus*, possibly acting as a prophylactic against the diseases it causes while maintaining the balance of the gut microbiota. To verify its immunogenicity and protective effectiveness, more *in vitro* and *in vivo* validation is necessary.

Materials and methods

Retrieval of proteome

The complete proteome of *R. gnavus* strain RJX1120 (Proteome ID: UP000234812) was retrieved in FASTA format from the UniProt database (16). A BLASTp search was performed against the Database of Essential Genes (DEG) to identify essential proteins in *R. gnavus* by comparing its proteome with known essential proteins. These proteins are vital for the bacterium's survival, growth, and key biological processes (17). Being integral components, these proteins are therefore crucial for an organism's survival in a particular environment (18). For greater refinement, Cello tool, which predicts the subcellular localization of proteins, was used to identify membrane-associated proteins from the list of essential proteins (19). Due to their accessibility to the host immune system, these membrane-associated proteins are highly promising targets for vaccine development (20). Then the screened proteins were analyzed for the presence of antigenicity based on a threshold of 0.5 using Vaxijen server because the proteins with high values are known to induce immense immune response upon exposure in the host (21, 22). The TMHMM v-2.0 server was used to predict potential transmembrane helices in the target proteins (23, 24).

Selection and assessment of CTL epitope

CTL epitopes for the target molecule were predicted using the MHC-I binding tool on the Immune Epitope Database (IEDB) (25). The consensus method was applied in the MHC-I binding tool to predict CTL epitopes. Epitopes with a consensus score of less than 2 were selected for further study (26). Subsequently, the potential immunogenicity of epitopes selected for CTL usage was assessed by the use of the IEDB immunogenicity tool (27). To confirm that the selected epitopes have strong potential to elicit an effective immune response, their antigenicity was assessed using the VaxiJen v2.0 server with a threshold of 0.5 (28). Only those epitopes that were considered antigenic were selected for incorporation in the vaccine construct (15). It is crucial that the vaccine candidate does not induce allergic or toxic reactions. The allergenic potential of the predicted epitopes was evaluated using the AllerTOP v2.0 server, and their toxicity was evaluated using the ToxinPred server (22, 29, 30). This broad approach ensured that safe and immunogenic CTL epitopes were identified for potential inclusion in a vaccine design.

HTL epitopes selection and analysis

Helper T lymphocytes (HTLs) are essential players in the adaptive immune system, where they orchestrate both cell-mediated and humoral responses against foreign pathogens (31). HTL epitopes from the target protein were identified by using the MHC Class II binding tool available through the Immune Epitope Database (IEDB) (32). The search was confined to 15-mer HTL epitopes for their binding affinity to a wide range of HLA-DR alleles to comprehensively cover the immunological coverage and were chosen for their optimal binding affinity and immunogenic potential (26). All selected epitopes were evaluated for antigenicity, allergenicity, and toxicity using the VaxiJen, AllerTOP v2.0, and ToxinPred servers, respectively (27).

LBL epitope identification and analysis

Linear B-cell epitopes are sequences of amino acids on the surface of proteins that can be recognized by antibodies (33). They are recognized by either naturally occurring antibodies or receptors on B cells, therefore able to stimulate cellular and humoral immunity (34). Vaccine development involves such epitopes critically because they enhance adaptive immunity through amplification of defence mechanisms in the immune response (35). The ABCPred server was used to predict linear B-cell epitopes from the target protein, with the threshold set at a minimum of 0.5 for prediction (36). After predicting epitopes, their antigenicity, allergenicity, and toxicity were assessed using the VaxiJen v2.0, AllerTOP v2.0, and ToxinPred servers, respectively (22, 27, 30). This way, only immunogenic, safe, and non-toxic epitopes were selected to be included in vaccine design (15).

Designing of vaccine construct

The multi-epitope vaccine (MEV) construct was designed by appropriately linking B-cell and T-cell epitopes with an adjuvant (37). Adjuvants are crucial for enhancing the immunogenicity of vaccine constructs, eliciting a stronger immune response in recipients (38). In this study, cholera enterotoxin subunit B (Accession No: P01556) was chosen due to its established ability to enhance the immunogenic potential of vaccine constructs (39). For linking components, EAAAK linkers were used to attach the CTL epitopes to the adjuvant, providing structural stability and

maintaining the functional integrity of the epitopes. GPGPG and AAY linkers were used to connect CTL and HTL epitopes, allowing their efficient presentation and enhancing their immune responses (40). Bi-lysine (KK) linkers were employed to preserve the specific immunogenic activity of linear B-cell (LBL) epitopes (41). Systematic arrangement of the construction is essential for immunogenic efficiency and structural stability; thus, the MEV construct represents a promising candidate for vaccine development (42).

Structural analysis

The structural properties of the MEV construct were analyzed using various bioinformatics tools (43). First, the physiochemical characteristics of the construct, such as theoretical isoelectric point (pI), molecular weight (MW), instability index (II), aliphatic index (AI), Grand Average of Hydropathicity (GRAVY), and *in vivo*/*in vitro* half-life, were evaluated using the ProtParam server (44, 45). Immunological efficacy was ensured by checking the antigenic and immunogenic profiles of the MEV construct using the IEDB immunogenicity tools and the Vaxijen v2.0 server (46). Allergenic potential was computed using the AllerTOP v2.0 server to ensure safety for the construct in its potential application in humans (42). The SOPMA tool was used to predict secondary structural features of the MEV construct, assessing the proportions of random coils, alpha-helices, beta-turns, and extended chains (44, 47). These detailed structural analyses allow researchers to understand stability, functionality, and applicability in the development of vaccines (48).

Refinement, confirmation, and prediction of tertiary structure

The prediction of the tertiary structure of the MEV construct is crucial for evaluating its structural and functional efficacy (37). For this purpose, the 3D structure of the MEV construct was predicted using the Alphafold server, which is a state-of-the-art tool for accurate protein structure modeling (49). The Galaxy Refine server was then used to refine and optimize the 3D structure for better stereochemical quality, and to minimize any structural errors that may occur (50). After refining the structure, the quality of the model was analyzed using the RAMPAGE server, where it evaluates the quality based on Ramachandran plot statistics (51). ERRAT server was used to check for possible errors and to assess the overall quality and reliability of the 3D structure of the MEV construct (52).

B-cell epitopes screening

The B-cell epitopes of the MEV construct were screened through the ABCPred online server and Ellipro tool within the IEDB-AR v2.22 suite (53, 54). For the prediction of linear epitopes, the amino acid sequence of the MEV was fed into the ABCPred server with a threshold set at 0.5 and an amino acid length fixed at 15 residues (55). For conformational epitope prediction, default parameters were used with Ellipro to analyze the 3D structure of MEV construct (15). The conformational approach would serve as a complement to these studies to identify linear epitopes, while identifying putative regions of immunogenicity in the MEV construct (56).

Binding analysis of TLR4 receptor with the designed vaccine

The effective recognition of the designed vaccine by the host's immune system is crucial for initiating a robust immune response (56). Molecular docking studies were carried out to evaluate the designed vaccine's binding capability with the immune receptors. For assessing the role of these immune receptors in stimulating the antimicrobial and adaptive immune responses, TLR4, MHC Class I and II were selected (57). TLR4 was selected due to its critical role in recognizing bacterial antigens and initiating an innate immune response (58). Protein-protein interactions between the vaccine construct and these receptors were modeled using the ClusPro server, which is a reliable tool for molecular docking (59). The resulting docked complexes were visualized using Chimera, a visualization tool for 3D molecular structures (60). The interactions within the docked complexes were analyzed by using the PDBsum server that provides detailed insights into interface residues and binding interactions (61).

Molecular dynamic simulation

Molecular dynamics simulations are computational methodologies used to study the dynamic behavior and stability of molecular systems such as protein-protein complexes (62). Interaction of the designed MEV construct with the selected receptor was analyzed by the iMODS server that proves to be a fast and efficient tool for molecular dynamics studies (15, 63). iMODS facilitates the exploration of dynamic properties and transition pathways between molecular entities to gain actionable insights into conformational changes (63). The stability of the docked complexes was evaluated through key parameters, such as the main-chain deformability plot, covariance matrix, eigenvalue analysis, B-factor values, and the elastic network model (64). These analyses provided a detailed understanding of the mechanical and dynamic stability of the protein-protein interaction (64).

Immune simulation

The immune response to the predicted vaccine construct was evaluated using the C-ImmSim 10.1 server (53). This platform is designed to simulate the interactions of the immune system, focusing on key functional components such as the bone marrow,

lymph nodes, and thymus (65). The simulation was performed using the following input parameters: Human Leukocyte Antigen (HLA) alleles (DRB1 0101, B0702, A0101), a random seed (12,345), a simulation volume of 10, one injection, and 100 steps (66). These alleles were selected for their broad population coverage and their relevance in antigen presentation, ensuring a diverse immune response (67). Other parameters were set to their default values to ensure accurate simulation of the immune response. This comprehensive immune simulation helped assess the potential efficacy and immunogenicity of the vaccine construct in a simulated mammalian immune system environment (67).

Dry lab cloning and codon optimization

Codon usage is species-specific, and the presence of non-adapted codons in a gene sequence can result in suboptimal expression levels in the host organism (68). To enhance gene expression, it is essential to optimize the codon usage to match the host's translational machinery (69). In this study, the Java Codon Adaptation Tool (JCAT) was utilized for the optimization and reverse translation of the MEV construct (70). During this process, prokaryotic ribosome binding sites, Rho-independent transcription termination signals, and appropriate restriction enzyme cleavage sites were selected to facilitate efficient expression and cloning (69). Subsequently, the optimized vaccine construct was inserted into the pET30a(+) vector using SnapGene software, ensuring a seamless cloning process for subsequent experimental validation (71).

Result

Proteome analysis

The complete proteome of the pathogenic strain *R. gnavus* RJX1120 was extracted from the UniProt database (Proteome ID: UP000234812), and a subtractive genomics approach was applied for the identification of potential vaccine targets against infections caused by *R. gnavus*. The total number of proteins in the proteome of the strain was found to be 3,219, and a comprehensive filtering pipeline was applied in order to find key target proteins. Initially, the database DEG identified 848 proteins that are essential in the proteome of a pathogen for survival and proliferation. These essential proteins were screened using BLASTp for non-homologous proteins, thus narrowing them down to 245. These 245 were again screened for localization of their subcellular positions and antigenicity. From there, the further studies have been conducted to further investigate the remaining 15 membrane-associated proteins that would possibly become vaccine candidates. These proteins were screened for their antigenicity, allergenicity, and stability. Of these, seven had the highest antigenicity, were non-allergenic, and very stable. The transmembrane helices of these were further tested. The two best vaccine candidates were the Single-stranded DNA-binding protein and the Cell division ATP-binding protein FtsE. These proteins exhibited high antigenicity, were non-allergenic and stable and lacked transmembrane helices (Table 1).

Epitope selection phase

Cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and linear B lymphocyte (LBL) epitopes of specific antigenic proteins were predicted during the epitope selection phase. Among the forecasted epitopes, the top seven CTL epitopes were selected for vaccine formulation based on their non-toxic, immunogenic, antigenic, and non-allergenic properties (Table 2). Similarly, the top four HTL epitopes exhibiting non-allergenicity, antigenicity, immunogenicity, IFN-gamma induction capability, and non-toxicity were identified for vaccine design (Table 3). Additionally, the top two LBL epitopes, characterized by their antigenicity, immunogenicity, non-allergenicity, and non-toxicity, were chosen for incorporation into the vaccine construct (Table 4).

Construction of multi epitope vaccine

The vaccine construct was designed by integrating 7 CTL epitopes, 4 HTL epitopes, and 2 B-cell epitopes with a suitable adjuvant and linkers. Cholera enterotoxin subunit B, consisting of 236 amino acids, was incorporated at the N-terminal of the vaccine using an EAAAK linker to enhance immunogenicity. The CTL, HTL, and B-cell epitopes were linked using AAY, GPGPG, and KK linkers, respectively, to maintain their individual immunological properties. The finalized vaccine construct comprised 347 amino acids (Figure 1).

Population coverage analysis

A comprehensive population coverage analysis was performed on the selected CTL and HTL epitopes utilized in the development of the multi-epitope vaccine (MEV). The analysis revealed that the chosen epitopes collectively covered approximately 71% of the global population. Notably, the highest population coverage was observed in Sweden, with an impressive 87%. Other countries also

TABLE 1 Comprehensive details regarding the antigenic vaccine protein derived from *Ruminococcus gnavus*.

Accession no	Protein	Antigenicity	Allergenicity	Toxicity
A0A2N5PT08	Single-stranded DNA-binding protein	0.7402	Non -allergen	Non -toxin
A0A2N5NK05	Cell division ATP-binding protein FtsE	0.5909	Non -allergen	Non -toxin

TABLE 2 Selected CTL epitopes finalized for vaccine construction targeting *Ruminococcus gnavus*.

Epitope	Protein	Allele	Position	Antigenicity	Immunogenicity
NLKRMKHRNIAK	Cell division ATP-binding protein FtsE	HLA-A*03:01	65-76	0.5296	-0.18793
VARYTVAVDRRF	Single-stranded DNA-binding protein	HLA-A*23:01 HLA-A*24:02	27-38	0.5194	0.26418
SVSGRIQTGSYT	Single-stranded DNA-binding protein	HLA-A*26:01 HLA-A*25:01	72-83	1.376	-0.02448
FRQGMRISVSGR	Single-stranded DNA-binding protein	HLA-A*31:01	65-76	0.9521	-0.23656
VNEMNERVITMK	Cell division ATP-binding protein FtsE	HLA-B*18:01	200-211	0.7641	0.09304
KRMKHRNIAKYR	Cell division ATP-binding protein FtsE	HLA-A*31:01 HLA-B*27:05	67-78	0.7392	-0.21613
NEMNERVITMKQ	Cell division ATP-binding protein FtsE	HLA-B*18:01	201-212	0.6243	0.01653

TABLE 3 Finalized HTL epitopes for vaccine construction targeting *Ruminococcus gnavus*.

Epitope	Protein	Allele	Position	Antigenicity	Immunogenicity
FAEKYFRQGMRISVS	Single-stranded DNA-binding protein	HLA-DRB1*15:02	60-74	0.6184	-0.21432
QGMRISVSGRIQTGS	Single-stranded DNA-binding protein	HLA-DRB1*07:03 HLA-DRB1*13:02	67-81	1.3073	-0.0008
RQGMRISVSGRIQTG	Single-stranded DNA-binding protein	HLA-DRB1*07:03 HLA-DRB1*13:02	66-80	1.2518	0.1341
SATAVARYTVAVDRR	Cell division ATP-binding protein FtsE	HLA-DRB1*08:06	23-37	0.6282	0.38099

exhibited significant coverage, including the Philippines (86%), Japan (80%), and Finland (76%). These findings substantiate the potential of the filtered epitopes as promising candidates for constructing an effective MEV targeting diverse populations globally (Figure 2).

Post-analysis of vaccine structure

The stereochemical properties of the constructed vaccine were analyzed using the ProtParam tool. The vaccine structure exhibited a molecular weight of 38,154.87 Da and an isoelectric point (pI) of 10.30, indicating its basic nature. It contained 55 positively charged amino acids (arginine and lysine) and 23 negatively charged amino acids (glutamic acid and aspartic acid). The instability index of the structure was calculated as 27.14, classifying it as stable. Furthermore, an aliphatic index of 68.44 confirmed its thermostability, while the GRAVY (Grand Average of Hydropathicity) value of -0.464 indicated a hydrophilic nature. The half-life of the vaccine was predicted to be 30 hours in mammals (*in vivo*), over 20 hours in yeast (*in vivo*), and over 10 hours in *E. coli* (*in vivo*). Additionally, the vaccine was confirmed to be non-allergenic, non-toxic, and antigenic.

Structural analysis of vaccine

Secondary structure analysis using SOPMA revealed that the 347-amino acid sequence comprises 162 residues forming α-helices (46.69%), 71 residues forming extended strands (20.46%), and 114 residues involved in random coils (32.85%), indicating a well-organized structural profile. The three-dimensional structure of the vaccine construct was predicted using the Alphafold server, followed by refinement through the Galaxy Refine server to optimize structural quality. Validation of the refined model was performed using a Ramachandran plot, which indicated that 95.7% of amino acid residues were located in the most favorable regions, 3.0% in the allowed regions, and 0.0% in the disallowed regions (Figure 3). Further evaluation demonstrated that the vaccine structure achieved a high-quality factor of 85.246 and a Z-score of -5.06, confirming the absence of poor rotamers (Figure 4).

Selection of B-cell epitopes

B-lymphocytes play a pivotal role in humoral immunity by producing antibodies. Therefore, an effective vaccine must include optimal B-cell epitope domains to elicit a robust antibody response.

TABLE 4 Finalized B-cell epitope selected for vaccine construction against *Ruminococcus gnavus*.

Epitope	Protein	Score	Position	Antigenicity	Immunogenicity
TRAANNKAANNKMEDG	Cell division ATP-binding protein FtsE	0.59	227	1.6355	0.38477
TAANPTMEDGNSINGL	Single-stranded DNA-binding protein	0.7	7	1.0945	-0.01349

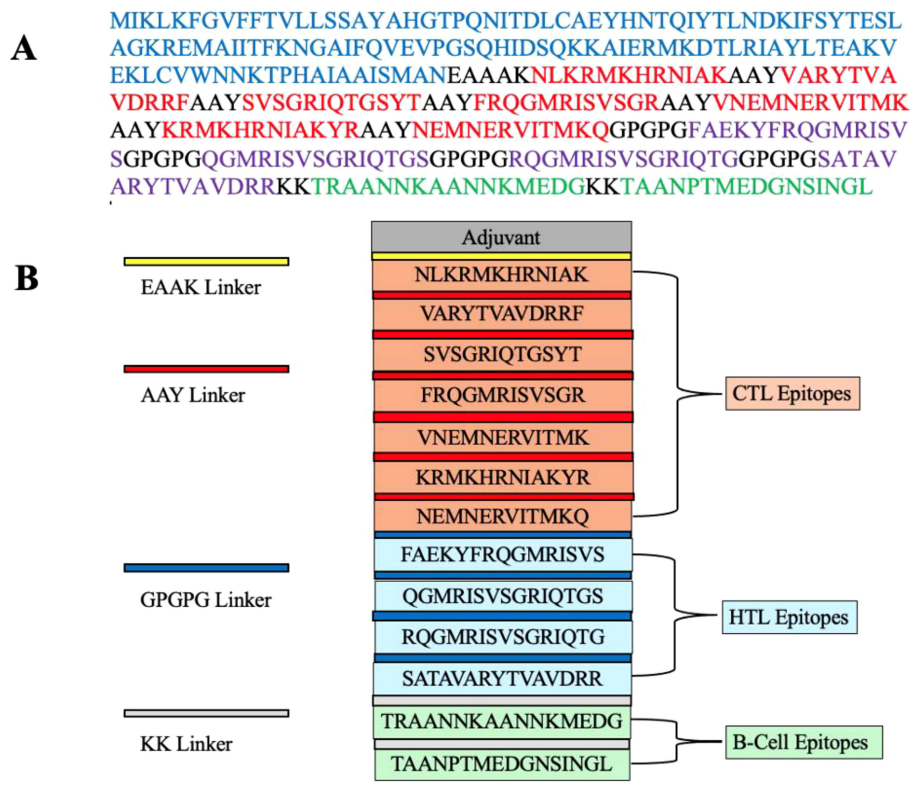


FIGURE 1
(A) A schematic representation of the MEV construct highlights the color-coded elements: the adjuvant (blue), CTL epitopes (red), HTL epitopes (purple), B-cell epitopes (green), and linkers (EAAAK, AAY, GPGPG, KK; all depicted in black). **(B)** The final multi-epitope vaccine (MEV) construct is composed of 347 amino acids. It includes an adjuvant (blue) linked via an EAAAK linker (black) and is connected to CTL epitopes (red) using an AAY linker (black). HTL epitopes (purple) are joined by GPGPG linkers (black), while KK linkers (black) connect B-cell epitopes (green).

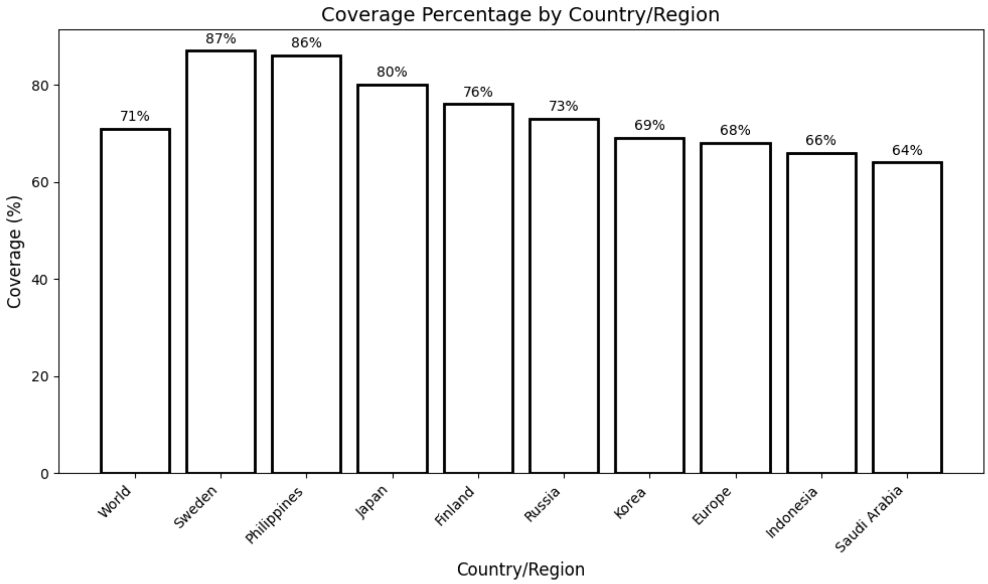


FIGURE 2
Population coverage analysis of selected T-cell epitopes across different countries/regions. The bar graph depicts the percentage coverage in the global population (71%) and specific regions, including Sweden (87%), the Philippines (86%), Japan (80%), Finland (76%), Russia (73%), Korea (69%), Europe (68%), Indonesia (66%), and Saudi Arabia (64%). This analysis highlights the broad applicability and potential impact of the designed multi-epitope vaccine in diverse populations.

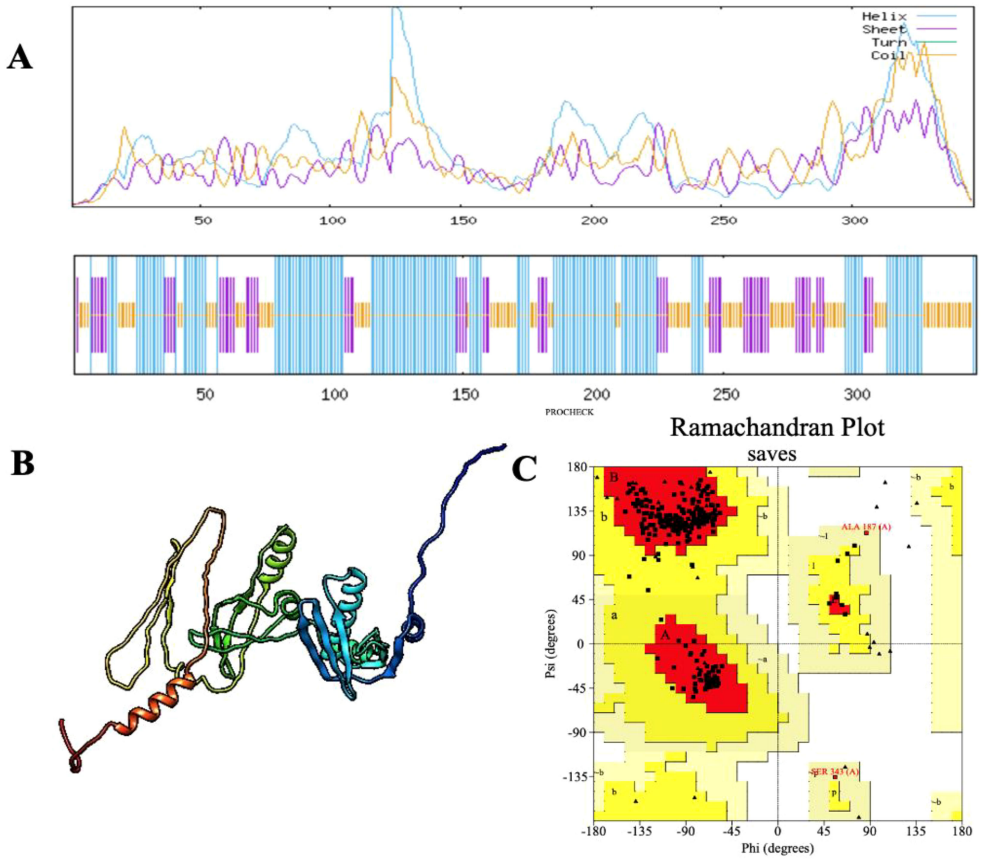


FIGURE 3
(A) Secondary structure prediction of the final multi-epitope vaccine construct using the SOPMA tool. The diagram illustrates the distribution of helices (blue), sheets (red), coils (purple), and turns (green). The horizontal black bar at the bottom represents the full length of the protein. **(B)** The refined 3D structure of the vaccine construct, displaying its spatial conformation. **(C)** Ramachandran plot analysis of the vaccine construct, demonstrating structural quality with 95.7% of amino acid residues positioned in favored regions.

In this study, 14 conformational B-cell epitopes, ranging from 3 to 53 residues in length, were identified with scores between 0.518 and 0.988. Additionally, 8 linear B-cell epitopes were predicted using the ElliPro server with default parameters. The conformational B-cell epitopes were visualized using PyMOL v1.3, a molecular graphics system, during the vaccine design process (Figure 5).

Molecular docking with host immune receptor

Molecular docking is a critical technique for elucidating the binding interactions between vaccine constructs and immune receptor proteins. In this study, the molecular docking of the designed multi-epitope vaccine (MEV) with the human Toll-like receptor 4 (TLR4) was performed using the ClusPro server. ClusPro is a highly reliable protein-protein docking platform that integrates a hybrid docking algorithm, combining experimental substrate binding site data with small-angle X-ray scattering for docking analyses. The refined 3D structure of the vaccine construct (ligand) was docked against the TLR4 receptor (PDB ID: 3FE8), generating 10 docking models. The top-ranked docking model, with 230 members in its cluster and an interaction energy of -1277.0 kcal/mol, demonstrated high stability of the vaccine-TLR4 complex. Molecular interactions within the docking complex were analyzed using the PDBsum server, which revealed that the MEV construct exhibited favorable binding with chain A of the TLR4 receptor, forming 13 hydrogen bonds (Figure 6). Thermodynamic parameters for the binding energy of the docking complex were computed using the PRODIGY tool. The equilibrium dissociation constant (Kd) was determined to be 4.1×10−8 at 37°C, with a Gibbs free energy change (ΔG) of -10.1 kcal/mol. These results confirm the stability and strong binding affinity of the MEV construct to the TLR4 receptor.

Normal mode analysis

Normal mode analysis (NMA) was performed to evaluate the molecular stability and functional motions of the MEV-TLR4 complex. The deformability plot revealed peak regions corresponding to main-chain residues exhibiting flexibility in

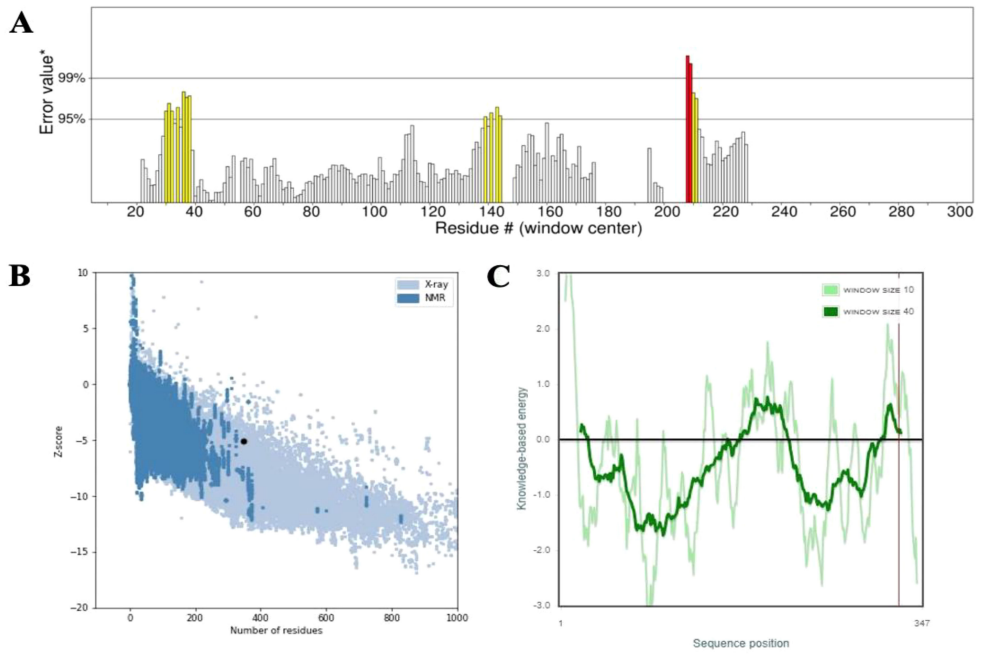


FIGURE 4
(A) Structural validation of the refined 3D vaccine model using the ERRAT tool. Regions of the structure rejected at the 99% confidence level are highlighted in red, while those rejected at the 95% confidence level are shown in yellow. **(B, C)** The Z-score plot of the refined 3D model, generated by ProSA-web, provides an assessment of the overall quality and reliability of the predicted vaccine structure.

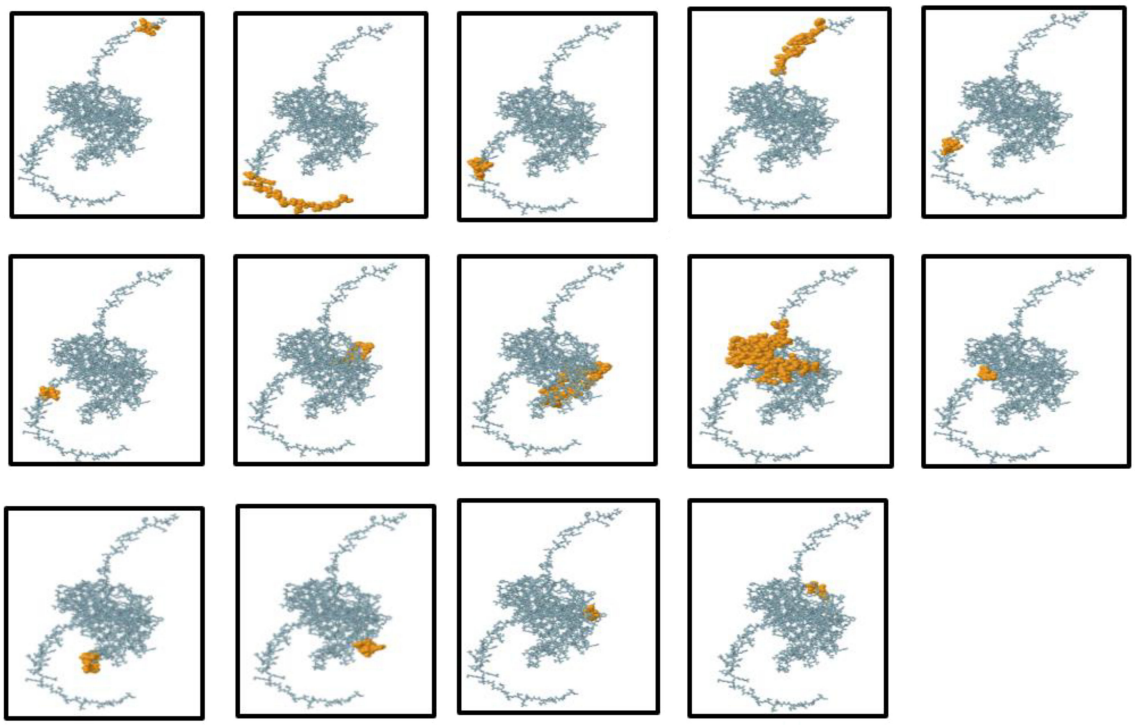


FIGURE 5
Three-dimensional representation of the conformational or discontinuous B-cell epitopes in the designed multi-epitope vaccine. The conformational B-cell epitopes are highlighted as an orange surface, while the remaining bulk of the polyprotein is depicted using grey stick representation.

the MEV-TLR4 complex. These highly deformable regions are indicative of "hinges" or "linkers" within the main chain. The experimental B-factor plot demonstrated the relationship between the NMA-predicted mobility and the MEV-TLR4 complex, showcasing the average RMSD values of the docked complex. The computed eigenvalue of the complex was 1.945512×10−7, reflecting the stiffness associated with each normal mode of motion. The variance bar illustrated individual (purple) and cumulative (green) contributions of each normal mode, indicating a negative correlation between variance and eigenvalue. Furthermore, a covariance map was generated to depict interatomic motions within the MEV-TLR4 complex. The map identified correlated (red), uncorrelated (white), and anti-correlated (blue) motions between different residue pairs. Additionally, a specialized elastic network model was constructed, representing the interatomic connections within the complex. The spring-like assembly between corresponding atoms and their stiffness were indicated by colored dots, with darker greys signifying more rigid interactions. Collectively, the NMA results demonstrated stable interactions and coordinated motions within the MEV-TLR4 complex, supporting its structural integrity and functionality (Figure 7).

Immune simulations

The immune simulation results demonstrated a robust enhancement of both primary and secondary immune responses to the top-ranked vaccine construct. Administration of the vaccine led to elevated levels of immunoglobulins, including IgG1 + IgG2, IgM, and IgM + IgG, indicative of a strong antibody-mediated immune response. The B-cell population showed significant expansion upon repeated exposure to the vaccine antigen, highlighting the formation of humoral immune memory. The simulations also revealed a marked increase in cytotoxic T cells (CTLs) and helper T cells (HTLs), coupled with a substantial reduction in antigen levels during secondary and tertiary immune responses, underscoring the vaccine's ability to enhance adaptive immunity. Additionally, the proliferation of natural killer cells, dendritic cells, and macrophages was predicted following each immunization cycle, reinforcing the construct's capacity to stimulate innate immune responses. The vaccine also elicited cytokine and interleukin release, particularly IFN-γ, TGF-β, IL-23, IL-10, and IFN-β, which are crucial for mounting an effective immune response against infection. Notably, continuous antigen exposure during the immunization period resulted in significantly elevated levels of IFN-γ and TGF-β, while other cytokines were detected at lower concentrations. The calculated Simpson's Index (D) confirmed a balanced immune response, reflecting the construct's comprehensive impact on immune diversity. These findings suggest that the proposed vaccine construct can effectively activate T and B lymphocytes, inducing robust antibody production and establishing long-lasting memory cells upon repeated antigen exposure. The immune simulation results further support the potential of the vaccine construct to elicit strong innate and adaptive immune responses, demonstrating its efficacy in combating leishmaniasis (Figure 8).

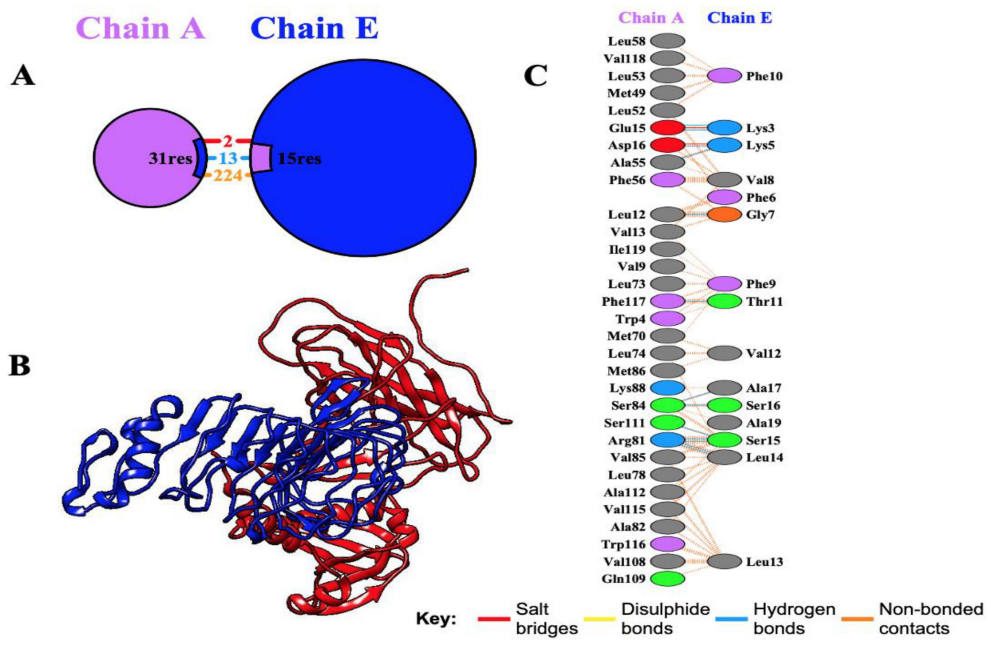


FIGURE 6
Docking analysis of the interaction between the receptor TLR4 and the multi-epitope vaccine (MEV). **(A)** Chain A and Chain E of the receptor and the MEV are depicted in purple and blue, respectively, highlighting the interacting residues. **(B)** Docking visualization showing Chain A of the receptor (blue) and the MEV (red), illustrating the optimal binding affinity. **(C)** Detailed illustration of interacting residues between the receptor and the vaccine construct, showing the formation of 13 hydrogen bonds between the receptor residues and the vaccine molecule.

Codon optimization and *in silico* restriction cloning

The expression potential of the proposed vaccine constructs was evaluated through codon optimization. Results obtained from the JCAT server revealed that all vaccine constructs achieved a Codon Adaptation Index (CAI) value of 1.0, indicating optimal codon usage. Furthermore, the GC content of the optimized cDNA sequences was 48%, which lies within the ideal range for efficient expression in the *E. coli* K12 vector. The optimized gene sequence of the prioritized vaccine construct was successfully integrated into the widely utilized pET30a(+) plasmid vector through *in silico* cloning. The total length of the recombinant plasmid was determined to be 5211 bp, confirming the feasibility of the construct for downstream applications (Figure 9).

Discussion

R. gnavus is an important member of the human gut microbiota that plays both commensal and pathogenic roles (72). The involvement of this bacterium in diseases like inflammatory bowel disease (IBD) highlights its clinical significance, positioning it as a potential target for therapeutic interventions (4). Conventional treatments are challenging to implement because the bacterium is resilient and can evade immune responses, making the development of innovative solutions such as vaccines an urgent necessity (73). Vaccination remains one of the most effective strategies for reducing morbidity and mortality associated with microbial infections, particularly against emerging pathogens (74). Advances in immunoinformatics and reverse vaccinology offer a modern, cost-effective framework for rapid vaccine development, overcoming the limitations of traditional methods (75). These methodologies have successfully applied to propose vaccines for pathogens as diverse as *Mycoplasma pneumoniae*, *Salmonella Typhimurium*, and *Campylobacter jejuni*, among others (76, 77).

Here, a multi-epitope vaccine (MEV) construct against R. gnavus was designed using subtractive proteomics combined with immunoinformatics, molecular docking, and simulation techniques. Core proteome analysis identified essential proteins that are non-homologous to human proteins while exhibiting antigenic properties (42, 78). Such core proteins are crucial because they give the host a broad-spectrum protection against various strains of the pathogen (76). Among the identified proteins, the single-stranded DNA-binding protein (SSB) and FtsE are crucial for bacterial survival and virulence (74, 79). SSB is essential for maintaining genomic stability during DNA

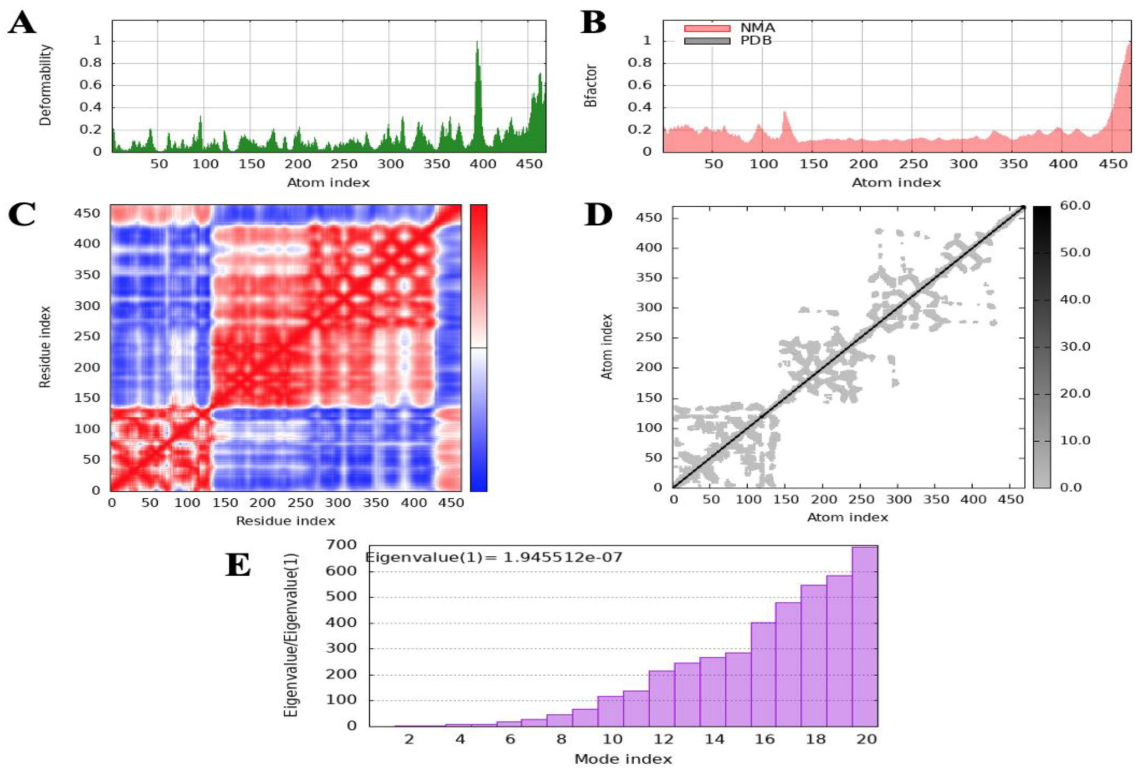


FIGURE 7
Molecular dynamics (MD) simulation analysis of the docked complex of the multi-epitope vaccine (MEV) with the receptor. **(A)** Deformability plot illustrating the flexibility of different regions in the docked complex. **(B)** B-factor analysis indicating the atomic fluctuations within the complex. **(C)** Covariance index depicting the correlated motions of residues. **(D)** Elastic network analysis demonstrating the connectivity and motion of residues within the complex. **(E)** Eigenvalue analysis representing the stiffness of the docked structure and its associated energy requirements.

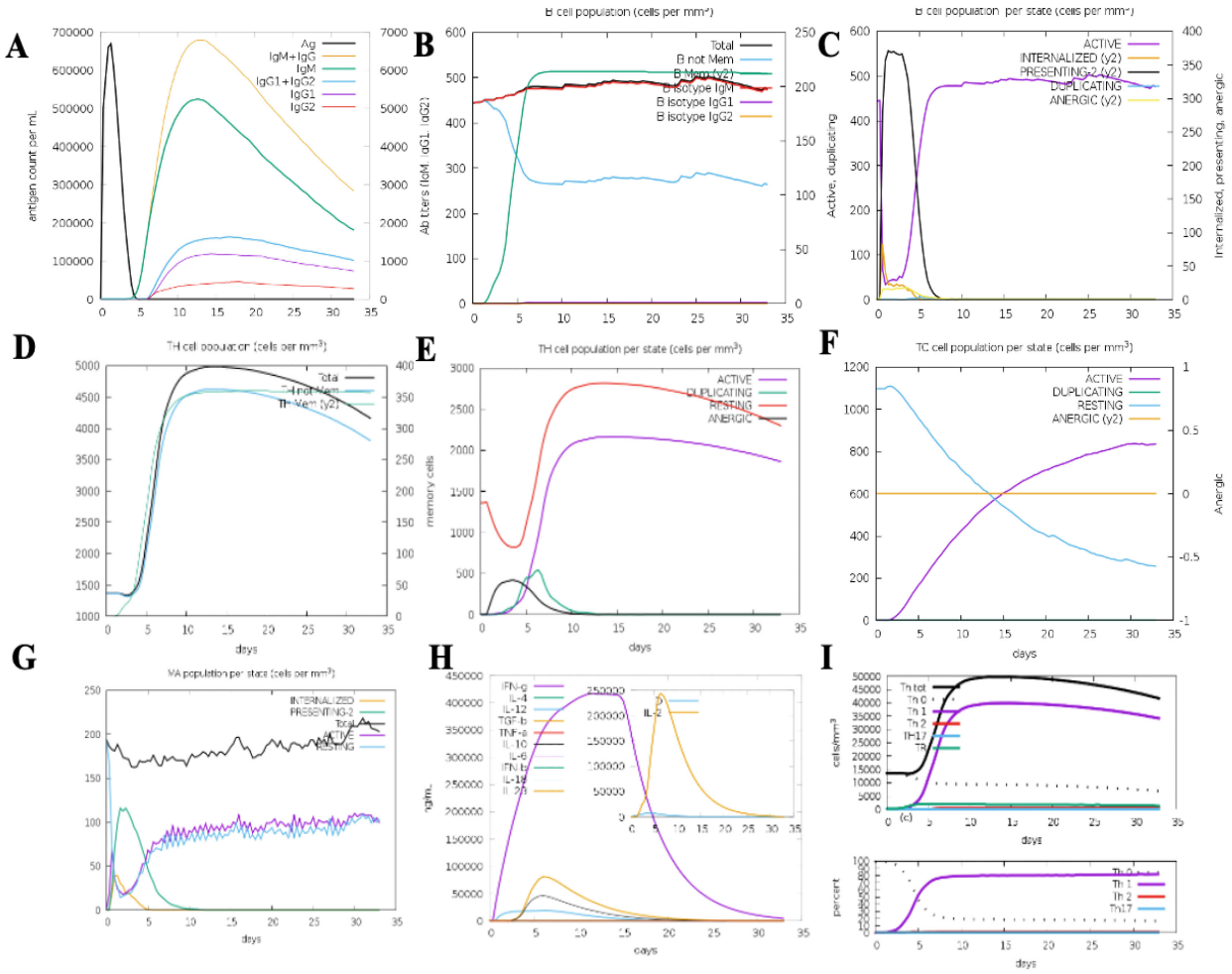


FIGURE 8
C-ImmSim immunization simulation results for the multi-epitope vaccine construct: **(A)** Immunoglobulin production depicted through color-coded peaks. **(B)** B-cell population showing increased types and class-switching potential. **(C)** Population distribution per state of B cells. **(D)** Evolution of T-helper cells over time. **(E)** Population distribution per state of T-helper cells. **(F)** Generation and dynamics of cytotoxic T cells. **(G)** Macrophage population distribution per state. **(H)** Cytokine and interleukin induction, showing elevated levels of IFN-γ and IL-2 post-vaccination. **(I)** Th1-mediated immune response activation.

replication and repair, particularly under stress conditions, ensuring the resilience of *R. gnavus* (80). Its conservation across bacterial species underscores its importance in safeguarding replication fidelity, which can contribute to the persistence of *R. gnavus* in the gut, even during inflammatory states such as IBD (81, 82). Similarly, FtsE, a component of the FtsEX complex, is integral to bacterial cell division and peptidoglycan remodeling (83). In *R. gnavus*, FtsE likely supports robust cell wall integrity, enhancing survival and adaptability in competitive gut environments (83, 84). These proteins underscore the bacterium's ability to endure host defenses and environmental stresses, making them potential targets for future therapeutic interventions.

Strict selection criteria were applied to identify CTL, HTL, and B-cell epitopes with high antigenicity while ensuring they were non-allergenic and non-toxic for potential use (85). Of extreme importance, the epitopes showed a very good worldwide population coverage, meaning an important potential for inducing immunity across different populations (28).

To increase the immunogenicity and stability of the vaccine, various linkers like AAY, KK, and GPGPG were used for joining the epitopes (86). These linkers have been reported to facilitate effective epitope processing, minimize junctional immunogenicity, and stimulate a robust immune response (87). The use of adjuvant cholera toxin subunit B coupled with the EAAAK linker ensured further stimulation of the immunity. This concept is very similar to a previous studies where these pairs were designed to improve stability and antigenicity of vaccines (76, 77).

Structural analysis of the vaccine construct showed that it was nontoxic, non-allergenic, and antigenic (75). Solubility predictions indicated that the vaccine would be easily expressible and

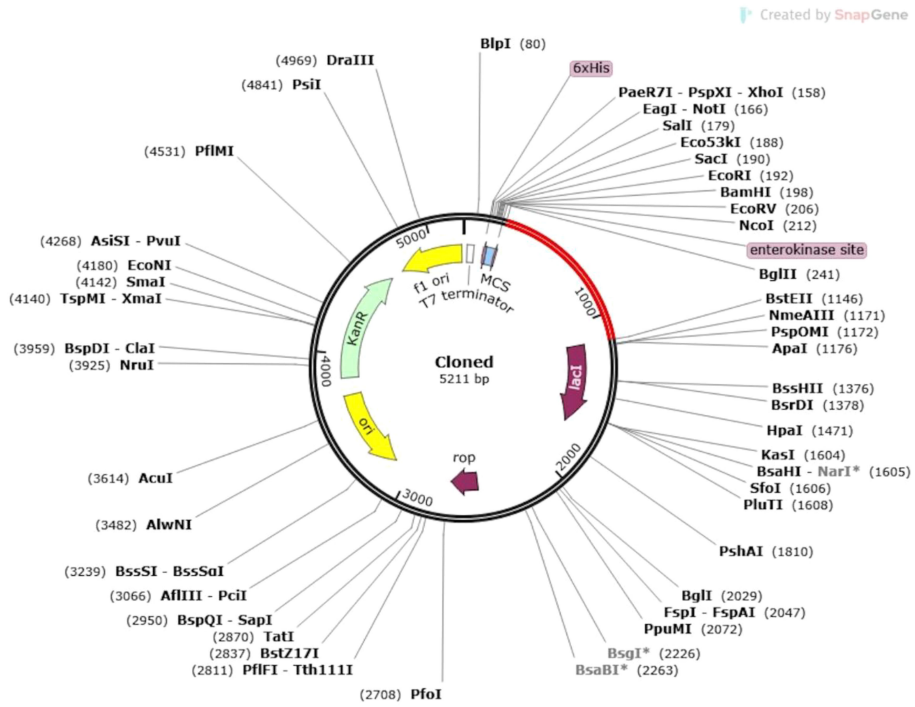


FIGURE 9
In silico cloning of the vaccine construct into the *E. coli* K12 host expression system. The plasmid backbone is represented in black, while the inserted nucleotide sequence of the vaccine construct is highlighted in red.

bioavailable in the host system. This is important because solubility plays a crucial role in determining the effectiveness of subunit vaccines in producing strong immune responses (38). Docking studies showed strong binding interactions between the vaccine construct and the TLR4 receptor, which is a central component of the innate immune system (15, 64). Molecular dynamics simulations further confirmed that the vaccine-TLR4 complex is stable, highlighting its potential to mediate innate immune responses (42, 63, 64).

Codon optimization enabled the construct to be used for the expression of the vaccine in *E. coli* K12, with a codon adaptation index of 1.0 and a GC content of 48%, both indicating high efficiency of transcription and translation (69, 70, 76). Predictions from the immune simulation showed that it would trigger strong cellular and humoral immunity, which would also include strong T cell and B cell activation and the formation of memory cells, thus implying that the vaccine would offer long-term immunity against diseases caused by *R. gnavus*.

Although the findings are encouraging, this research has limitations. Predictions based on immunoinformatics are highly dependent on computational algorithms, which could not perfectly mimic biological outcomes. Hence, *in vitro* and *in vivo* studies that experimentally validate the safety and efficacy of the vaccine are needed. Further information on *R. gnavus* pathogenesis and host immune system interaction could narrow down vaccine targets and produce better outcomes. This is a rationally designed construct of a multi-epitope vaccine that would potentially activate robust immune responses against *R. gnavus*. Further experimental validations are required, but such a vaccine would be ideal for overcoming the challenges related to this opportunistic pathogen while maintaining gut microbiota balance.

Conclusion

This study applied subtractive proteomics and reverse vaccinology to find vaccine candidates and design a multi-epitope vaccine against *R. gnavus* strain RJX1120. Pathogenic strain-specific antigenic proteins were selected to minimize off-target effects on beneficial gut microbiota. The identified antigens included Single-stranded DNA-binding protein and Cell division ATP-binding protein FtsE, promising as vaccine candidates. Epitopes predicted for B and T cells would generate both humoral and cell-mediated immunity. Adjuvants and linkers have been incorporated to increase their immunogenicity and stability. The proposed vaccine showed favorable structural and physicochemical properties, including strong binding affinity with TLR4 receptors, confirmed by molecular docking and simulation studies. Immune simulations predicted robust *in vivo* immunogenicity. Codon optimization and reverse translation ensured efficient expression in *E. coli*. Experimental validation in animal models is essential to confirm the efficacy and safety of the designed vaccine.

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

HD: Investigation, Methodology, Resources, Software, Validation, Writing – review & editing. SM: Conceptualization, Data curation, Formal Analysis, Supervision, Visualization, Writing – review & editing. SG: Data curation, Software, Writing – original draft, Writing – review & editing. HA: Data curation, Formal Analysis, Funding acquisition, Resources, Visualization, Writing – original draft. NM: Formal Analysis, Funding acquisition, Investigation, Resources, Writing – review & editing. AH: Data curation, Investigation, Project administration, Writing – review & editing. IM: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. XS: Conceptualization, Data curation, Methodology, Resources, Software, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors extend their appreciation to the Researchers supporting project number (RSP2025R479) King Saud University, Riyadh, Saudi Arabia.

Acknowledgments

The authors extend their appreciation to the Researchers supporting project number (RSP2025R479) 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Juge N. Microbe Profile: Ruminococcus gnavus: the yin and yang of human gut symbionts: This article is part of the Microbe Profiles collection. *Microbiology*. (2023) 169:001383. doi: 10.1099/mic.0.001383

2. Crost EH, Coletto E, Bell A, Juge N. *Ruminococcus gnavus* : friend or foe for human health. *FEMS Microbiol Rev*. (2023) 47:fuad014. doi: 10.1093/femsre/fuad014

3. Henke MT, Brown EM, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. Capsular polysaccharide correlates with immune response to the human gut microbe *Ruminococcus gnavus*. *Proc Natl Acad Sci*. (2021) 118:e2007595118. doi: 10.1073/pnas.2007595118

4. Gilliland A, Chan JJ, De Wolfe TJ, Yang H, Vallance BA. Pathobionts in inflammatory bowel disease: origins, underlying mechanisms, and implications for clinical care. *Gastroenterology*. (2024) 166:44–58. doi: 10.1053/j.gastro.2023.09.019

5. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci*. (2019) 116:12672–7. doi: 10.1073/pnas.1904099116

6. Zhai L, Huang C, Ning Z, Zhang Y, Zhuang M, Yang W, et al. Ruminococcus gnavus plays a pathogenic role in diarrhea-predominant irritable bowel syndrome by increasing serotonin biosynthesis. *Cell Host Microbe*. (2023) 31:33–44.e5. doi: 10.1016/j.chom.2022.11.006

7. Al-Amrani S, Al-Jabri Z, Al-Zaabi A, Alshekaili J, Al-Khabori M. Proteomics: Concepts and applications in human medicine. *World J Biol Chem*. (2021) 12:57–69. doi: 10.4331/wjbc.v12.i5.57

8. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N. Utilisation of mucin glycans by the human gut symbiont ruminococcus gnavus is strain-dependent. *PloS One*. (2013) 8:e76341. doi: 10.1371/journal.pone.0076341

9. Reygaert WC, Department of Biomedical Sciences. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol*. (2018) 4:482–501. doi: 10.3934/microbiol.2018.3.482. Oakland University William Beaumont School of Medicine, Rochester, MI, USA.

10. Uruén C, Chopo-Escuin G, Tommassen J, Mainar-Jaime RC, Arenas J. Biofilms as promoters of bacterial antibiotic resistance and tolerance. *Antibiotics*. (2020) 10:3. doi: 10.3390/antibiotics10010003

11. Rahman S, Chiou C-C, Ahmad S, Islam ZU, Tanaka T, Alouffi A, et al. Subtractive Proteomics and Reverse-Vaccinology Approaches for Novel Drug Target Identification and Chimeric Vaccine Development against Bartonella henselae Strain Houston-1. *Bioengineering*. (2024) 11:505. doi: 10.3390/bioengineering11050505

12. Hisham Y, Ashhab Y, Hwang S-H, Kim D-E. Identification of highly conserved SARS-coV-2 antigenic epitopes with wide coverage using reverse vaccinology approach. *Viruses*. (2021) 13:787. doi: 10.3390/v13050787

13. Eyayu T, Zeleke AJ, Worku L. Current status and future prospects of protein vaccine candidates against Schistosoma mansoni infection. *Parasite Epidemiol Control*. (2020) 11:e00176. doi: 10.1016/j.parepi.2020.e00176

14. Qureshi NA, Bakhtiar SM, Faheem M, Shah M, Bari A, Mahmood HM, et al. Genome-based drug target identification in human pathogen streptococcus gallolyticus. *Front Genet*. (2021) 12:564056. doi: 10.3389/fgene.2021.564056

15. Ghaffar SA, Tahir H, Muhammad S, Shahid M, Naqqash T, Faisal M, et al. Designing of a multi-epitopes based vaccine against Haemophilius parainfluenzae and its validation through integrated computational approaches. *Front Immunol*. (2024) 15:1380732. doi: 10.3389/fimmu.2024.1380732

16. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. (2015) 43:D204–12. doi: 10.1093/nar/gku989

17. Zhang R. DEG: a database of essential genes. *Nucleic Acids Res*. (2004) 32:271D–272. doi: 10.1093/nar/gkh024

18. Luo H, Lin Y, Liu T, Lai F-L, Zhang C-T, Gao F, et al. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Res*. (2021) 49:D677–86. doi: 10.1093/nar/gkaa917

19. Yu C-S, Cheng C-W, Su W-C, Chang K-C, Huang S-W, Hwang J-K, et al. CELLO2GO: A web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. *PloS One*. (2014) 9:e99368. doi: 10.1371/journal.pone.0099368

20. Hasan A, Ibrahim M, Alonazi WB, Shen J. Application of immunoinformatics to develop a novel and effective multiepitope chimeric vaccine against Variovorax durovernensis. *Comput Biol Chem*. (2024) 113:108266. doi: 10.1016/j.compbiolchem.2024.108266

21. Malik M, Khan S, Ullah A, Hassan M, Haq MU, Ahmad S, et al. Proteome-Wide Screening of Potential Vaccine Targets against Brucella melitensis. *Vaccines*. (2023) 11:263. doi: 10.3390/vaccines11020263

22. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf*. (2007) 8:4. doi: 10.1186/1471-2105-8-4

23. Bianchi F, Textor J, Van Den Bogaart G. Transmembrane helices are an overlooked source of major histocompatibility complex class I epitopes. *Front Immunol*. (2017) 8:1118. doi: 10.3389/fimmu.2017.01118

24. Ganapathiraju M, Balakrishnan N, Reddy R, Klein-Seetharaman J. Transmembrane helix prediction using amino acid property features and latent semantic analysis. *BMC Bioinf*. (2008) 9:S4. doi: 10.1186/1471-2105-9-S1-S4

25. Rencilin CF, Rosy JC, Mohan M, Coico R, Sundar K. Identification of SARS-CoV-2 CTL epitopes for development of a multivalent subunit vaccine for COVID-19. *Infect Genet Evol*. (2021) 89:104712. doi: 10.1016/j.meegid.2021.104712

26. Jakhar R, Gakhar SK. An immunoinformatics study to predict epitopes in the envelope protein of SARS-coV-2. *Can J Infect Dis Med Microbiol*. (2020) 2020:1–14. doi: 10.1155/2020/7079356

27. Mir SA, Alaidarous M, Alshehri B, Bin Dukhyil AA, Banawas S, Madkhali Y, et al. Immunoinformatics-based identification of B and T cell epitopes in RNA-dependent RNA polymerase of SARS-coV-2. *Vaccines*. (2022) 10:1660. doi: 10.3390/vaccines10101660

28. 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) 10:10895. doi: 10.1038/s41598-020-67749-1

29. Chao P, Zhang X, Zhang L, Yang A, Wang Y, Chen X. Proteomics-based vaccine targets annotation and design of multi-epitope vaccine against antibiotic-resistant Streptococcus gallolyticus. *Sci Rep*. (2024) 14:4836. doi: 10.1038/s41598-024-55372-3

30. 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

31. Alexander J, Fikes J, Hoffman S, Franke E, Sacci J, Appella E, et al. The optimization of helper T lymphocyte (HTL) function in vaccine development. *Immunol Res*. (1998) 18:79–92. doi: 10.1007/BF02788751

32. 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

33. Lollier V, Denery-Papini S, Larré C, Tessier D. A generic approach to evaluate how B-cell epitopes are surface-exposed on protein structures. *Mol Immunol*. (2011) 48:577–85. doi: 10.1016/j.molimm.2010.10.011

34. Hoffman W, Lakkis FG, Chalasani G. B cells, antibodies, and more. *Clin J Am Soc Nephrol*. (2016) 11:137–54. doi: 10.2215/CJN.09430915

35. Da Silva BM, Ascher DB, Pires DEV. epitope1D: accurate taxonomy-aware B-cell linear epitope prediction. *Brief Bioinform*. (2023) 24:bbad114. doi: 10.1093/bib/bbad114

36. Singh J, Malik D, Raina A. Immuno-informatics approach for B-cell and T-cell epitope based peptide vaccine design against novel COVID-19 virus. *Vaccine*. (2021) 39:1087–95. doi: 10.1016/j.vaccine.2021.01.011

37. 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:e0244176. doi: 10.1371/journal.pone.0244176

38. Facciolà A, Visalli G, Laganà A, Di Pietro A. An overview of vaccine adjuvants: current evidence and future perspectives. *Vaccines*. (2022) 10:819. doi: 10.3390/vaccines10050819

39. Aslam S, Ashfaq UA, Zia T, Aslam N, Alrumaihi F, Shahid F, et al. Proteome based mapping and reverse vaccinology techniques to contrive multi-epitope based subunit vaccine (MEBSV) against Streptococcus pyogenes. *Infect Genet Evol*. (2022) 100:105259. doi: 10.1016/j.meegid.2022.105259

40. Chakkyarath V, Shanmugam A, Natarajan J. Prioritization of potential drug targets and antigenic vaccine candidates against Klebsiella aerogenes using the computational subtractive proteome-driven approach. *J Proteins Proteomics*. (2021) 12:201–11. doi: 10.1007/s42485-021-00068-9

41. Hammed-Akanmu M, Mim M, Osman AY, Sheikh AM, Behmard E, Rabaan AA, et al. Designing a Multi-Epitope Vaccine against Toxoplasma gondii: An Immunoinformatics Approach. *Vaccines*. (2022) 10:1389. doi: 10.3390/vaccines10091389

42. Rasool D, Jan SA, Khan SU, Nahid N, Ashfaq UA, Umar A, et al. Subtractive proteomics-based vaccine targets annotation and reverse vaccinology approaches to identify multiepitope vaccine against Plesiomonas shigelloides. *Heliyon*. (2024) 10:e31304. doi: 10.1016/j.heliyon.2024.e31304

43. Rastogi A, Gautam S, Kumar M. Bioinformatic elucidation of conserved epitopes to design a potential vaccine candidate against existing and emerging SARS-CoV-2 variants of concern. *Heliyon*. (2024) 10:e35129. doi: 10.1016/j.heliyon.2024.e35129

44. Shams M, Nourmohammadi H, Asghari A, Basati G, Majidiani H, Naserifar R, et al. Construction of a multi-epitope protein for human Toxocara canis detection: Immunoinformatics approach multi-epitope construct for T. *canis serodiagnosis Inform Med Unlocked*. (2021) 26:100732. doi: 10.1016/j.imu.2021.100732

45. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PWR, Kayastha AM, et al. MFPPI – multi FASTA protParam interface. *Bioinformation*. (2016) 12:74–7. doi: 10.6026/97320630012074

46. Zhuang L, Ali A, Yang L, Ye Z, Li L, Ni R, et al. Leveraging computer-aided design and artificial intelligence to develop a next-generation multi-epitope tuberculosis vaccine candidate. *Infect Med*. (2024) 3(4):100148. doi: 10.1016/j.imj.2024.100148

47. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. (1995) 11:681–4. doi: 10.1093/bioinformatics/11.6.681

48. Negahdari B, Sarkoohi P, Ghasemi Nezhad F, Shahbazi B, Ahmadi K. Design of multi-epitope vaccine candidate based on OmpA, CarO and ZnuD proteins against multi-drug resistant Acinetobacter baumannii. *Heliyon*. (2024) 10:e34690. doi: 10.1016/j.heliyon.2024.e34690

49. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK. Before and after AlphaFold2: An overview of protein structure prediction. *Front Bioinforma*. (2023) 3:1120370. doi: 10.3389/fbinf.2023.1120370

50. 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

51. Wang W, Xia M, Chen J, Deng F, Yuan R, Zhang X, et al. Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in Gossypium raimondii and G. arboreum. *Data Brief*. (2016) 9:345–8. doi: 10.1016/j.dib.2016.05.025

52. Al-Khayyat MZS, Al-Dabbagh AGA. In silico Prediction and Docking of Tertiary Structure of LuxI, an Inducer Synthase of Vibrio fischeri. *Rep Biochem Mol Biol*. (2016) 4:66–75.

53. Khan T, Islam J, Parihar A, Islam R, Jerin TJ, Dhote R, et al. Immunoinformatics and molecular modeling approach to design universal multi-epitope vaccine for SARS-CoV-2. *Inform Med Unlocked*. (2021) 24:100578. doi: 10.1016/j.imu.2021.100578

54. 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) 9:514. doi: 10.1186/1471-2105-9-514

55. EL-Manzalawy Y, Dobbs D, Honavar VG. In silico prediction of linear B-cell epitopes on proteins. In: Zhou Y, Kloczkowski A, Faraggi E, Yang Y, editors. *Prediction of protein secondary structure. Methods in molecular biology*. Springer New York, New York, NY (2017). p. 255–64. doi: 10.1007/978-1-4939-6406-2_17

56. Naz S, Aroosh A, Caner A, Şahar EA, Toz S, Ozbel Y, et al. Immunoinformatics approach to design a multi-epitope vaccine against cutaneous leishmaniasis. *Vaccines*. (2023) 11:339. doi: 10.3390/vaccines11020339

57. Frei R, Steinle J, Birchler T, Loeliger S, Roduit C, Steinhoff D, et al. MHC class II molecules enhance toll-like receptor mediated innate immune responses. *PloS One*. (2010) 5:e8808. doi: 10.1371/journal.pone.0008808

58. Ain QU, Batool M, Choi S. TLR4-targeting therapeutics: structural basis and computer-aided drug discovery approaches. *Molecules*. (2020) 25:627. doi: 10.3390/molecules25030627

59. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein–protein docking. *Nat Protoc*. (2017) 12:255–78. doi: 10.1038/nprot.2016.169

60. Goddard TD, Huang CC, Ferrin TE. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure*. (2005) 13:473–82. doi: 10.1016/j.str.2005.01.006

61. 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

62. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. (2018) 99:1129–43. doi: 10.1016/j.neuron.2018.08.011

63. 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

64. Khan SY, Rather MA, Shah A, Ahmad I, Ahmad I, Saba K, et al. Exploring 3D structure of gonadotropin hormone receptor using homology modeling, molecular dynamic simulation and docking studies in rainbow trout, Oncorhynchus mykiss. *Endocr Metab Sci*. (2024) 15:100171. doi: 10.1016/j.endmts.2024.100171

65. Waqas M, Aziz S, Bushra A, Halim SA, Ali A, Ullah S, et al. Employing an immunoinformatics approach revealed potent multi-epitope based subunit vaccine for lymphocytic choriomeningitis virus. *J Infect Public Health*. (2023) 16:214–32. doi: 10.1016/j.jiph.2022.12.023

66. AlChalabi R, Al-Rahim A, Omer D, Suleiman AA. Immunoinformatics design of multi-epitope peptide-based vaccine against Haemophilus influenzae strain using cell division protein. *Netw Model Anal Health Inform Bioinforma*. (2022) 12:1. doi: 10.1007/s13721-022-00395-x

67. Choudhury A, Sen Gupta PS, Panda SK, Rana MK, Mukherjee S. Designing AbhiSCoVac - A single potential vaccine for all 'corona culprits': Immunoinformatics and immune simulation approaches. *J Mol Liq*. (2022) 351:118633. doi: 10.1016/j.molliq.2022.118633

68. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep*. (2022) 49:539–65. doi: 10.1007/s11033-021-06749-4

69. Paremskaia AI, Kogan AA, Murashkina A, Naumova DA, Satish A, Abramov IS, et al. Codon-optimization in gene therapy: promises, prospects and challenges. *Front Bioeng Biotechnol*. (2024) 12:1371596. doi: 10.3389/fbioe.2024.1371596

70. 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:W526–31. doi: 10.1093/nar/gki376

71. Jamil F, Aslam L, Laraib, Ali H, Shoukat K, Rasheed MA, et al. An In silico study of derivative of Newcastle disease virus epitopes based vaccine against Hemagglutunin neuraminidase protein. *J Anim Sci*. (2022) 101:skac375. doi: 10.1093/jas/skac375

72. Laplanche V, Armiento S, Speciale I, Šuligoj T, Crost EH, Lamprinaki D, et al. The human gut symbiont Ruminococcus gnavus displays strain-specific exopolysaccharides modulating the host immune response. *Carbohydr Polym*. (2025) 347:122754. doi: 10.1016/j.carbpol.2024.122754

73. Chinemerem Nwobodo D, Ugwu MC, Oliseloke Anie C, Al-Ouqaili MTS, Chinedu Ikem J, Victor Chigozie U, et al. Antibiotic resistance: The challenges and some emerging strategies for tackling a global menace. *J Clin Lab Anal*. (2022) 36:e24655. doi: 10.1002/jcla.24655

74. McGlynn SE, Boyd ES, Shepard EM, Lange RK, Gerlach R, Broderick JB, et al. Identification and characterization of a novel member of the radical adoMet enzyme superfamily and implications for the biosynthesis of the hmd hydrogenase active site cofactor. *J Bacteriol*. (2010) 192:595–8. doi: 10.1128/JB.01125-09

75. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PWR, Kayastha AM, et al. Immunoinformatics and reverse vaccinomic approaches for effective design. In: *Computational approaches for novel therapeutic and diagnostic designing to mitigate SARS-coV-2 infection*. Amsterdam, Netherlands: Elsevier (2022). p. 357–78. doi: 10.1016/B978-0-323-91172-6.00004-2

76. O'Ryan M, Vidal R, Del Canto F, Carlos Salazar J, Montero D. Vaccines for viral and bacterial pathogens causing acute gastroenteritis: Part II: Vaccines for *Shigella*, *Salmonella*, enterotoxigenic E. coli (ETEC) enterohemorragic *E. coli* (EHEC) and *Campylobacter jejuni*. *Hum Vaccines Immunother*. (2015) 11:601–19. doi: 10.1080/21645515.2015.1011578

77. Vilela Rodrigues TC, Jaiswal AK, Lemes MR, Da Silva MV, Sales-Campos H, Alcântara LCJ, et al. An immunoinformatics-based designed multi-epitope candidate vaccine (mpme-VAC/STV-1) against Mycoplasma pneumoniae. *Comput Biol Med*. (2022) 142:105194. doi: 10.1016/j.compbiomed.2021.105194

78. Osterloh A. Vaccination against bacterial infections: challenges, progress, and new approaches with a focus on intracellular bacteria. *Vaccines*. (2022) 10:751. doi: 10.3390/vaccines10050751

79. Xu L, Halma MTJ, Wuite GJL. Unravelling how single-stranded DNA binding protein coordinates DNA metabolism using single-molecule approaches. *Int J Mol Sci*. (2023) 24:2806. doi: 10.3390/ijms24032806

80. Ashton NW, Bolderson E, Cubeddu L, O'Byrne KJ, Richard DJ. Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol Biol*. (2013) 14:9. doi: 10.1186/1471-2199-14-9

81. Zaitsev EN, Kowalczykowski SC. Binding of double-stranded DNA by Escherichia coli RecA protein monitored by a fluorescent dye displacement assay. *Nucleic Acids Res*. (1998) 26:650–4. doi: 10.1093/nar/26.2.650

82. Kester JC, Kandror O, Akopian T, Chase MR, Zhu J, Rubin EJ, et al. ClpX is essential and activated by single-strand DNA binding protein in mycobacteria. *J Bacteriol*. (2021) 203(4):e00608-20. doi: 10.1128/JB.00608-20

83. Alcorlo M, Straume D, Lutkenhaus J, Håvarstein LS, Hermoso JA. Structural characterization of the essential cell division protein ftsE and its interaction with ftsX in streptococcus pneumoniae. *mBio*. (2020) 11:e01488–20. doi: 10.1128/mBio.01488-20

84. Bisson-Filho AW, Hsu Y-P, Squyres GR, Kuru E, Wu F, Jukes C, et al. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science*. (2017) 355:739–43. doi: 10.1126/science.aak9973

85. Lin L, Ting S, Yufei H, Wendong L, Yubo F, Jing Z. Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Res*. (2020) 288:198082. doi: 10.1016/j.virusres.2020.198082

86. Srinivasan S, Selvaraj GF, Gopalan V, Padmanabhan P, Ramesh K, Govindan K, et al. Epitope identification and designing a potent multi-epitope vaccine construct against SARS-coV-2 including the emerging variants. *J Glob Infect Dis*. (2022) 14:24–30. doi: 10.4103/jgid.jgid_96_21

87. De Groot AS, Moise L, McMurry JA, Martin W. Epitope-based immunome-derived vaccines: A strategy for improved design and safety. In: Falus A, editor. *Clinical applications of immunomics*. Springer US, New York, NY (2009). p. 39–69. doi: 10.1007/978-0-387-79208-8_3

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 May 2025
DOI 10.3389/fimmu.2025.1576557

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College & Hospital, India

REVIEWED BY
Jhon Carlos Castaño,
University of Quindío, Colombia
Viol Dhea Kharisma,
Airlangga University, Indonesia
Pradeep Darshana Pushpakumara,
University of Missouri, United States
Leah Kashiri,
University of Zimbabwe, Zimbabwe

*CORRESPONDENCE
Sivakumar Arumugam
siva_kumar.a@vit.ac.in

†PRESENT ADDRESS
Janaki Ramaiah Mekala,
School of Biosciences and Technology, Vellore Institute of Technology, Vellore, TamilNadu, India

‡These authors have contributed equally to this work and share first authorship

RECEIVED 14 February 2025
ACCEPTED 23 April 2025
PUBLISHED 19 May 2025

CITATION
Ramalingam PS, Aranganathan M, Hussain MS, Elangovan S, Chellasamy G, Balakrishnan P, Mekala JR, Yun K and Arumugam S (2025) Unveiling reverse vaccinology and immunoinformatics toward Saint Louis encephalitis virus: a ray of hope for vaccine development.
Front. Immunol. 16:1576557.
doi: 10.3389/fimmu.2025.1576557

Unveiling reverse vaccinology and immunoinformatics toward Saint Louis encephalitis virus: a ray of hope for vaccine development

Prasanna Srinivasan Ramalingam [1‡], Mahalakshmi Aranganathan [1,2‡], Md Sadique Hussain [3], Sujatha Elangovan [1], Gayathri Chellasamy [4], Purushothaman Balakrishnan [5], Janaki Ramaiah Mekala [6†], Kyusik Yun [4] and Sivakumar Arumugam [1*]

[1]Protein Engineering Lab, School of Biosciences and Technology, Vellore Institute of Technology, Vellore, TamilNadu, India, [2]School of Natural Sciences and Mathematics, The University of Texas at Dallas, Richardson, TX, United States, [3]Uttaranchal Institute of Pharmaceutical Sciences, Uttaranchal University, Dehradun, Uttarakhand, India, [4]Department of Bionanotechnology, Gachon University, Seongnam-si, Gyeonggi-do, Republic of Korea, [5]Department of Biomaterials, Saveetha Dental College and Hospitals, SIMATS, Saveetha University, Chennai, India, [6]Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur, Andhra Pradesh, India

Introduction: Infectious diseases continue to challenge human health with high incidence and mortality rates worldwide. Notably, the adaptability of RNA viruses, highlighted by outbreaks of SARS, MERS, and COVID-19, emphasizes the timely need for effective therapeutics. Saint Louis encephalitis virus (SLEV) belonging to the Flaviviridae family is an RNA virus that mostly affects the central nervous system (CNS) of humans. Although supportive care treatments such as antiemetics and painkillers are being used against SLEV infection, it still lacks potential therapeutics for the effective treatment.

Methods: Reverse vaccinology and immunoinformatics approaches help in the identification of suitable epitopes to design a vaccine construct that will activate both B- and T-cell-mediated responses. Previous studies used only the envelope protein E for the vaccine design, but we have used multiple protein targets to enhance the vaccine efficacy. Thus, in the present study, we have designed a multi-epitope subunit vaccine that specifically targets the membrane glycoprotein M, envelope protein E, and anchored capsid protein anchC of SLEV.

Results: Our results indicated that the vaccine construct is structurally stable, antigenic, non–allergic, non–toxic, and soluble. Additionally, the vaccine construct was structurally refined and indicated significant binding affinity toward the Toll-like receptor 4 (TLR-4) supported by molecular docking and molecular dynamics simulations. Furthermore, it also indicated that it has the potential to induce an immune response.

Conclusion: In addition, it has been cloned in the pET-28a (+) vector-6xHis-TEV-ORF9c expression vector for further experimental validation. We also recommend to evaluate the designed vaccine's therapeutic efficacy through *in vitro* and *in vivo* studies in the near future.

KEYWORDS

Saint Louis encephalitis virus, vaccine, epitope, antigen, immune response

1 Introduction

Infectious diseases caused by pathogenic microorganisms possess a significant challenge and health burden to humans with widespread morbidity and mortality worldwide (1–3). A recent report indicated that the frequency of SLEV varied from 80% of the cases in patients <20 years of age to 95% in those >60 years of age, and of the 47 confirmed human cases, 45 patients were hospitalized and among them 9 died at a younger age (4). Despite the advancements of several strategies to combat these pathogen-induced diseases, they adapt to extreme environments and even result in antimicrobial resistance (AMR) in the case of bacteria and antigenic shifts and drifts in the case of viruses (5–8). Additionally, urbanization (notably in low- to middle-income countries), globalization (rapid dissemination via travel), and sudden climate changes (high risk for outbreaks) also accelerated the wide spread of these infectious diseases causing localized outbreaks, widespread epidemics, and even global pandemics (9, 10). To note, the COVID-19 pandemic due to the SARS-CoV-2 outbreak resulted in high mortality and incidence rates and indicated the risk of these infectious diseases to human health (11–13). Furthermore, addressing these challenges requires multidisciplinary approaches such as a one-health approach, public health interventions, intensive medical research, systemic and bioinformatics approaches, and global collaborations to mitigate their impact on human health (14–16).

Saint Louis encephalitis virus (SLEV) belonging to the Flaviviridae family is a mosquito-borne flavivirus that has a single-stranded RNA in its genome (17, 18). SLEV is a zoonotic disease that is mainly transmitted from the bite of infected Culex mosquitoes, particularly *Culex pipiens*, *Culex quinquefasciatus*, and *Culex nigripalpus* and its first large endemic outbreak was observed in 1933 in the United States, and also observed in Central American and South American regions (18–21). However, Culex mosquitoes are not only restricted to the Americas; they have a global distribution and are commonly found in tropical and temperate regions worldwide, where they serve as vectors for multiple arboviruses including SLEV (20, 22). Primarily, birds are the reservoir hosts of SLEV and humans are the incidental and disease-obtaining hosts (17). SLEV is closely related to other flavivirus such as Japanese encephalitis (JEV) and West Nile virus (WNV), which often show asymptomatic conditions characterized by fatigue, headaches, nausea, vomiting, and body aches among the infected individuals (23–25). The cases of fatality rate for encephalitis caused by SLEV ranges from 5% to 15%, which mostly infects adults and could be diagnosed by neutralizing antibody testing and IgM ELISA kits (26–28). Unfortunately, there is no specific treatment available for SLEV-infected patients and potent prophylactic vaccines to combat SLEV infections; however, supportive care such as antiemetics and painkillers are being provided (26). Alongside, several therapeutic strategies are being developed and studied for the better treatment of SLEV in preclinical and clinical settings. Notably, two previous efforts were made to produce an SLEV vaccine. Hossain et al. have also employed the immunoinformatics approach to design a vaccine against SLEV and showed that it has potential against the envelope protein E SLEV, and Blaney Jr et al. have developed a live attenuated virus vaccine by employing SLE/DEN4-436,437 clone 41 and SLE/DEN4-654,655 clone 46 viruses (29, 30). To note, there is no vaccine currently available for the effective treatment of SLEV (28).

Reverse vaccinology and immunoinformatics approaches help in the identification of suitable epitopes to design a vaccine construct that will activate both B- and T-cell-mediated response using bioinformatics approaches (31). This approach has been extended toward the development of vaccines for various infectious diseases including SARS-CoV-2 and also extended to the development of cancer vaccines (32). In the present study, we employed reverse vaccinology and immunoinformatics approaches to design a multi-epitope subunit vaccine that specifically targets membrane glycoprotein M, envelope protein E, and anchored capsid protein anchC of SLEV.

2 Materials and methods

2.1 Data retrieval

Initially, Saint Louis encephalitis virus was provided as the query, and the FASTA sequences of the proteins, membrane glycoprotein M (NCBI Reference Sequence: YP_009329948.1), envelope protein E (NCBI Reference Sequence: YP_009329949.1), and the anchored capsid protein anchC (NCBI Reference Sequence: YP_009329944.1) of SLEV were retrieved from the NCBI-Protein database (https://www.ncbi.nlm.nih.gov/) (32, 33). The three-

dimensional structures of the HLA-A*02:01 (PDB ID: 1DUZ), HLA-DRB1*01:01 (PDB ID: 1AQD), and Toll-like receptor 4 (TLR4) (PDB ID: 4G8A) were also retrieved from the Protein Data Bank (34, 35).

2.2 CTL and HTL epitope identification and selection

Since the cytotoxic T lymphocytes (CTL) (9-mer) and helper T lymphocytes (HTL) (15-mer) are involved in the induction of immune response in humans, the CTL and HTL epitopes were predicted using the NetCTL 1.2 web server (https://services.healthtech.dtu.dk/services/NetCTL-1.2/) and NetMHCII 2.3 web server (https://services.healthtech.dtu.dk/services/NetMHCII-2.3/), respectively (36, 37). For CTL epitopes, they were identified against the 12 types of MHC-I with 0.75 as the default threshold, and for the HTL epitopes, they were identified against all the alleles of HLA-DR, HLA-DQ, and HLA-DP, respectively. The robustness of the predictions was validated by ANN 4.0 and MHC Flurry 2.0 for MHC I epitopes and validated by Combinatorial library & Tepitope for MHC II epitopes in the IEBD tool, respectively (https://www.iedb.org/) (38). Following this, all the predicted epitopes were subjected to antigenicity (model set as tumor), allergenicity, and toxicity (SVM-based method) analysis using the VaxiJen v2.0 web server (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (39), AllerTOP v.2 web server (https://www.ddg-pharmfac.net/allertop_test/) (40), and ToxinPred web server (https://webs.iiitd.edu.in/raghava/toxinpred/index.html) (41), respectively. Furthermore, the IFN-γ induction potential of HTL epitopes was also predicted with a hybrid approach (motif+SVM model) and the IFN-γ vs. non-IFN-γ model was used using the IFNepitope web server (https://webs.iiitd.edu.in/raghava/ifnepitope/application.php) (42). Alongside, the sequence conservation analysis of the predicted epitopes of SLEV was analyzed using protein-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), toward Dengue virus 1 (taxid:11053), Zika virus (taxid:64320), Yellow fever virus (taxid:11089), West Nile virus (taxid:11082), and Japanese encephalitis virus (taxid:11072), which are closely related to the Flavivirus family.

2.3 Docking of T-cell epitopes with HLA alleles

The three-dimensional structures of the selected CTL and HTL epitopes were modeled using the PEP-FOLD 3.5 web server (https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) (43). Furthermore, the CTL and HTL epitopes were docked against the HLA-A*02:01 and HLA-DRB1*01:01 alleles to evaluate their binding potential and molecular interactions against these more common alleles in the world population using the HPEPDOCK 2.0 web server (http://huanglab.phys.hust.edu.cn/hpepdock/) (44).

2.4 Vaccine construct design

The selected CTL epitopes, HTL epitopes, linkers, and adjuvant were used to design the multi-epitope vaccine construct. CTL epitopes were linked with the AAY linker and HTL epitopes with the CPGPG linker, whereas the adjuvant was connected with an EAAAK linker. The TLR4 agonist, 50s ribosomal L7/L12 protein of *Mycobacterium tuberculosis*, was used as an adjuvant in the vaccine construct to elucidate the strong immune response (45, 46).

2.5 Analysis of physicochemical characteristics, antigenicity, and allergenicity

The physicochemical properties such as the molecular weight, theoretical PI, amino acid composition and length, total number of negatively charged and positively charged residues, instability index, aliphatic index, and GRAVY of the designed multi-epitope vaccine were predicted using the Expasy ProtParam web server (https://web.expasy.org/protparam/) (47). In addition, the antigenicity, allergenicity, and solubility of the designed multi-epitope vaccine were predicted using the VaxiJen v2.0 web server (39), AllerTOP v.2 web server (40), and SOLpro web server (48), respectively. Furthermore, the antigenic nature of the adjuvant was predicted by evaluating the antigenicity of the designed multi-epitope vaccine with and without the presence of adjuvant using the ANTIGENpro web server (https://scratch.proteomics.ics.uci.edu/) (48).

2.6 Structural analysis and molecular docking of the designed vaccine construct

Initially, the 2D structure of the designed multi-epitope vaccine construct was predicted using the PDBsum database (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) (49). Then, the 3D structure of the designed multi-epitope vaccine construct was predicted using the I-TASSER web server (https://zhanggroup.org/I-TASSER/) (50) and further refined by the GalaxyRefine web server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) (51). In addition, the refined 3D model of the designed multi-epitope vaccine construct was validated by Ramachandran plot and Z-score plot by employing the PDBsum database and ProSA-web web server (https://prosa.services.came.sbg.ac.at/prosa.php), respectively (52). Then, the perfectly refined model was docked against the Toll-like receptor-4 (TLR4) protein using the ClusPro 2.0 web server (https://cluspro.bu.edu/login.php?redir=/home.php) (53).

2.7 Molecular dynamics simulations

The molecular dynamics simulations of the TLR4–vaccine complexes were performed using GROMACS 2020, and the protein topology files were generated using the GROMOS 42a1

force field (54, 55). The systems were solvated in an orthorhombic box using the simple point charge water model, and the neutralization was achieved by adding Na+ counter ions. Then, the energy minimization was carried out by the steepest descent algorithm with 50,000 steps, and the system was equilibrated under the NVT ensemble for 500 ps at 300 K, followed by NPT equilibration for 1,000 ps. Furthermore, the cutoff distance of 1.2 nm was applied for short-range non-bonded interactions, including Coulombic and van der Waals potentials, and the system was subjected to a 100-ns molecular dynamics simulation analysis. Finally, the resulting trajectories were analyzed to assess the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and solvent-accessible surface area (SASA) using standard GROMACS tools and visualized using the ggplot2 package (56, 57).

2.8 Normal mode analysis

The protein deformation analysis of the TLR4-vaccine docked complex was analyzed using the internal coordinates normal mode analysis (NMA) by employing the iMODS web server (https://imods.iqf.csic.es/) (58). The NMA analysis was conducted using the CA atomic model to evaluate their B-factor/mobility, eigenvalue, variance, covariance map, and elastic network for the TLR4-vaccine docked complex (59).

2.9 Immune response simulation

The immune response induction is a crucial factor in vaccination, and thus the immune response simulation of the designed multi-epitope vaccine construct was evaluated using the C-ImmSim web server (https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php) that employs the position-specific score matrix (PSSM) and the Celada–Seiden model (60). The simulation parameters were configured with a random seed of 12,345, a simulation volume of 10 μL, and 1,095 simulation steps, representing a time span of 1 year (365 days). The vaccine was administered in three doses on days 0, 28, and 56, corresponding to time steps 1, 84, and 168, respectively. Injection modes were performed without LPS, and all other parameters were set to their default values.

2.10 Codon optimization and *in silico* cloning analysis

The vaccine construct's protein sequence was reverse-translated, and its cDNA sequence was optimized for codon usage by employing the Java Codon Adaptation Tool (JCat) (https://www.jcat.de/), and the *E. coli* K12 was employed as the expression host (61). Then, the optimized sequence was inserted and cloned in the pET-28a (+) vector-6xHis-TEV-ORF9c (5,554 bp) using the SnapGene software (https://www.snapgene.com/). The complete schematic representation of the workflow of the study is shown in Figure 1.

3 Results

3.1 Selected T–cell epitopes showed potential interaction toward HLA alleles

The CTL epitopes (9-mer) and the HTL epitopes (15-mer) were predicted against the 12 types of MHC-I molecules, and all alleles of HLA-DR, HLA-DQ, and HLA-DP. The CTL epitopes predicted against the membrane glycoprotein M, envelope protein E, and anchored capsid protein anchC of SLEV along with their specific biding MHC-I allele are provided in Supplementary Tables S1, S2, and S3, respectively. Similarly, their predicted HTL epitopes of proteins of SLEV along with their specific biding MHC-II allele are provided in Supplementary Tables S4, S5, and S6, respectively. The antigenicity, allergenicity, and toxicity properties of the predicted CTL and HTL epitopes were evaluated, and IFN-γ induction potential was also predicted for the HTL epitopes (15mer). The epitopes were screened with these criteria such as antigenic, non-allergen, non-toxic, and IFN-γ induction (only for 15mer), and the shortlisted CTL and HTL epitopes are provided in Supplementary Tables S7 and S8, respectively. The final CTL and HTL epitopes selected for the vaccine construct along with their epitope names are provided in Supplementary Table S9. Additionally, the sequence conservation analysis was performed toward Dengue virus 1 (taxid:11053), Zika virus (taxid:64320), Yellow fever virus (taxid:11089), West Nile virus (taxid:11082), and Japanese encephalitis virus (taxid:11072), which are closely related to the Flavivirus family, and the results are provided as similarity percentage in Supplementary Table S9. Notably, the West Nile virus and Japanese encephalitis virus shared a similarity percentage of most predicted epitopes, and Dengue virus 1 and Zika virus shared a similarity percentage with one CTL and one HTL epitope, respectively. This similarity-conserved epitopes have the potential to induce cross-reactive T-cell responses and broaden protection toward other species such as West Nile virus and Japanese encephalitis virus, indicating that the developed vaccine construct was broad-spectrum.

Furthermore, the selected CTL and HTL epitopes were docked against the HLA-A*02:01 and DRB1*01:01 alleles, which are the most frequent alleles among the world population, and their binding energies (kcal/mol) and docked pose are shown in Table 1, Supplementary Figures S1 and S2. Totally, five CTL epitopes were docked against HLA-A*02:01 and 17 HTL epitopes were docked against HLA-DRB1*01:01 molecules. From the docking analysis, we observed that the CTL epitope (RVVFVIMLM) and the HTL epitope (TTQINYHWHKEGSSI) showcased high binding affinities toward their respective allele with binding energies of −240.708 and −234.422 kcal/mol, respectively, and the CTL epitope (TISPQAPSF) and HTL epitope (MKMEATELATVREYC) showcased comparatively less binding affinities toward their respective allele with binding energies of −210.309 and −182.066 kcal/mol, respectively. For CTL epitopes,

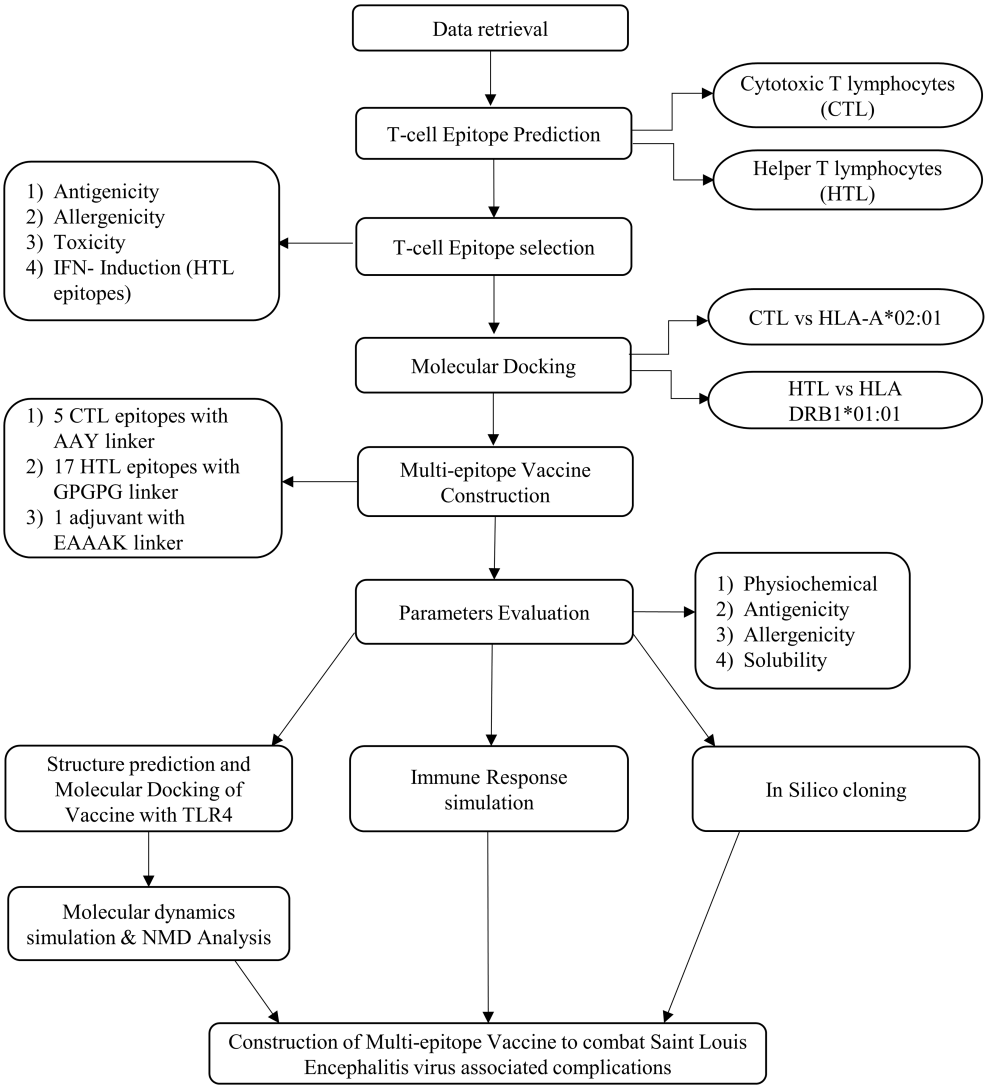


FIGURE 1
Schematic representation of the workflow of the study.

the binding affinities range from −210.309 to −240.708 kcal/mol, and for HTL epitopes, they range from −182.066 to −234.422 kcal/mol. Moreover, all the CTL and HTL epitopes indicated their potential binding affinities and thus they were selected in the construction of a multi-epitope vaccine.

3.2 Designed multi-epitope vaccine showed desired physiochemical properties

Generally, adjuvants are used in multi-epitope peptide vaccines to induce strong immune responses when injected into humans. In our study, we have used the C-terminal region of the large ribosomal subunit protein bL12 of *Mycobacterium tuberculosis* as the adjuvant (MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVK), which highly prevents the autoimmune reactions. AAY linkers were used to link the CTL epitopes, GPGPG linkers were used to link the HTL epitopes, and the EAAAK linker was used to link the adjuvant in the vaccine construct. The total vaccine construct contains 532 amino acids, comprising 5 CTL epitopes, 17 HTL epitopes, 1 adjuvant, 4 CTL linkers, 16 HTL linkers, and 1 adjuvant linker, as shown in Figure 2. Following this, the physiochemical properties of the designed multi-epitope vaccine construct were evaluated and are tabulated in Table 2. We observed that alanine (A) is more frequent with 11.8% followed by Arg (R) with 2.6%, as shown in Figure 3. The SOL-pro web server indicated the soluble nature of the designed multi-epitope vaccine construct with a probability of 0.902, and the instability index of 23.84 (less than 40) indicates the stability of the vaccine. The antigenic score of the designed multi-epitope vaccine construct was observed to be 0.787234 (without adjuvant) and 0.898972 (with adjuvant), indicating the increase in antigenic response when adjuvant is added to the vaccine construct. The

10.3389/fimmu.2025.1576557

TABLE 1 Binding energies of selected CTL and HTL epitopes against HLA molecules.

Epitope type	Epitope	HLA molecule	Binding energy (kcal/mol)
CTL	NIKYEVAIF	HLA-A*02:01	−225.608
	TISPQAPSF	HLA-A*02:01	−210.309
	RDRSISLTL	HLA-A*02:01	−226.722
	QRVVFVIML	HLA-A*02:01	−213.502
	RVVFVIMLM	HLA-A*02:01	−240.708
HTL	ALAIGWMLGSNNTQR	HLA-DRB1*01:01	−216.101
	DFGSIGGVFNSIGKA	HLA-DRB1*01:01	−205.397
	GASGATWIDLVLEGG	HLA-DRB1*01:01	−194.994
	KMEATELATVREYCY	HLA-DRB1*01:01	−215.410
	LFGGMSWITQGLLGA	HLA-DRB1*01:01	−233.383
	LGALLLWMGLQARDR	HLA-DRB1*01:01	−227.230
	LVTVNPFISTGGANN	HLA-DRB1*01:01	−223.634
	MKMEATELATVREYC	HLA-DRB1*01:01	−182.066
	MSWITQGLLGALLLW	HLA-DRB1*01:01	−220.584
	NLPWTSPATTDWRNR	HLA-DRB1*01:01	−232.190
	PQAPSFTANMGEYGT	HLA-DRB1*01:01	−207.506
	PTLDFKVMKMEATEL	HLA-DRB1*01:01	−227.681
	REYCYEATLDTLSTV	HLA-DRB1*01:01	−206.449
	SGINTEDYYVFTVKE	HLA-DRB1*01:01	−240.228
	TKQTVVALGSQEGAL	HLA-DRB1*01:01	−201.193
	TTQINYHWHKEGSSI	HLA-DRB1*01:01	−234.422
	TIDCEARSGINTEDY	HLA-DRB1*01:01	−188.060

Grand Average of Hydropathy (GRAVY) is used to determine the hydrophobic nature of the protein and is generally calculated by summing up the hydropathy values of all the amino acids and dividing it by the total number of amino acids of the protein. The positive value indicates the hydrophobic nature and the negative value indicates the hydrophilic nature of the given protein. In our study, the vaccine construct showed a GRAVY score of −0.040 that indicates its hydrophilic nature, as shown in in Table 2.

TABLE 2 Physiochemical properties of the designed vaccine construct.

Parameter	Value/range
Number of amino acids	532
Molecular formula	$C_{2440}H_{3808}N_{636}O_{741}S_{19}$
Molecular weight	54518.03 Da
Theoretical pI	4.80
Total number of positive charge residues (Arg + Lys)	55
Total number of negative charge residues (Asp + Glu)	39
Instability index	23.84
Aliphatic index	78.21
GRAVY	−0.040
Estimated half life	30 h (mammalian reticulocytes, *in vitro*). >20 h (yeast, *in vivo*). >10 h (Escherichia coli, *in vivo*).

3.3 Structural modeling and refinement of the multi-epitope vaccine

The 2D structure of the designed multi-epitope vaccine construct consisting of 532 amino acids was predicted and observed to have 8 sheets, 5 beta hairpins, 2 beta bulges, 19 strands, 6 helices, 181 beta turns, and 36 gamma turns, as shown in Figure 4. Then, the 3D structure was modeled by the I-TASSER web server, which resulted in five best models with C-scores of −3.15, −3.58, −3.63, −3.79, and −3.68, respectively. Generally, the

10.3389/fimmu.2025.1576557

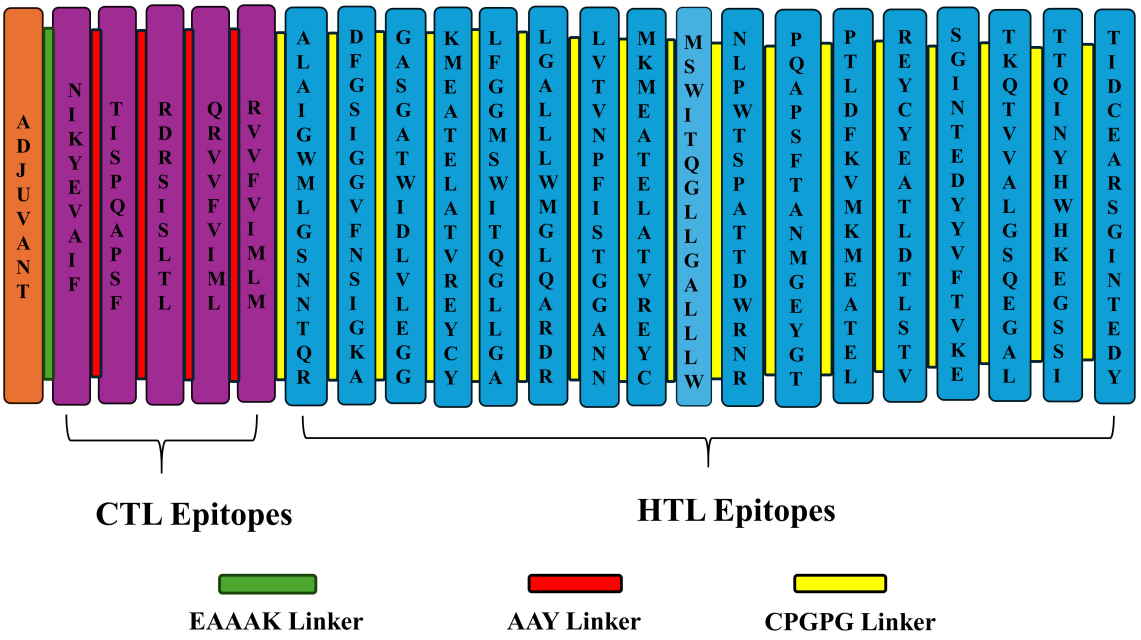


FIGURE 2
Pictorial representation of the designed multi-epitope vaccine construct.

high C-score represents the high confidence of the predicted model, and thus model 1 with a c-score of −3.15 was selected for further refinement acknowledging the crucial role of accurate 3D structural prediction in understanding the vaccine's potential efficacy and stability. Likewise, the GalaxyRefine web server resulted in the best five refined models, in which model 2 was chosen based on a comprehensive evaluation of several structural parameters: a high GDT-HA score of 0.8459, a low RMSD of 0.707, a favorable MolProbity score of 3.467, a clash score of 78.9, a low percentage of poor rotamers at 2.1%, and a significant proportion of Ramachandran favored regions at 67.9% as shown in Figure 5A. These metrics collectively suggest a highly refined and accurate model, crucial for ensuring the vaccine's effectiveness and structural refinement. Furthermore, the refined model models were validated by Ramachandran plot and Z-score analysis. The most favored regions on a Ramachandran plot are important because they help to identify the validity of a vaccine construct's 3D structure and indicate which Phi/Psi angles are possible for an amino acid; thus, high % of most favored regions indicates better structural enhancement whereas the less % shows poor enhancement. Notably, in our findings, the Ramachandran plot analysis demonstrated an increase in the most favored regions from 43.2%

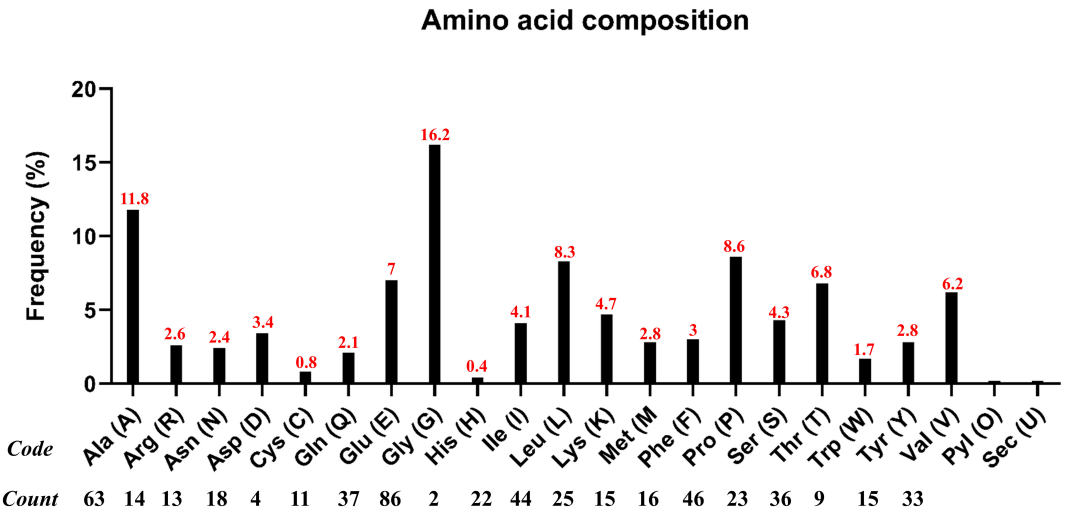


FIGURE 3
Amino acid composition and frequency of the designed multi-epitope vaccine construct.

10.3389/fimmu.2025.1576557

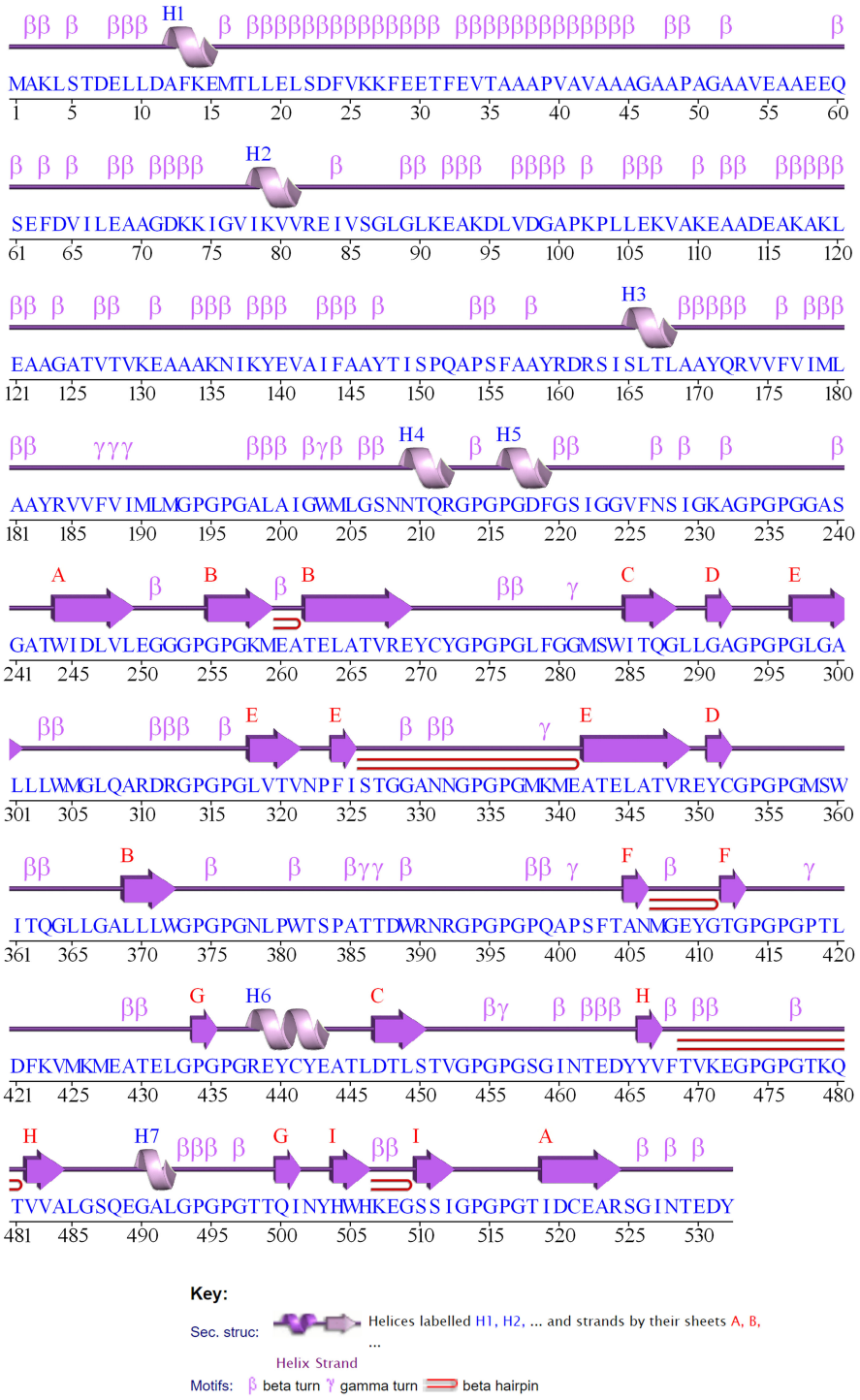


FIGURE 4
2D structure of the designed multi-epitope vaccine construct.

in the unrefined model to 56.8% in the refined model, indicating improved structural quality and reduced steric clashes. Also, the G-factor, measuring the overall structural unusualness, improved significantly from −2.02 (unrefined) to −1.24 (refined), highlighting the enhanced accuracy and reliability of the refined model, as depicted in Figures 5B and C, respectively. In the Z-score plot, the higher negative value indicates the high confidence of the modeled structure of the vaccine construct whereas the lesser negative value indicates less confidence and the positive value indicates very poor confidence of the vaccine structure. In our study, we have observed that the Z-score was improved from −2.36 (unrefined) to −2.42 (refined), further confirming the high structural refinement and enhanced stability of the vaccine construct, as shown in Figures 5D and E, respectively.

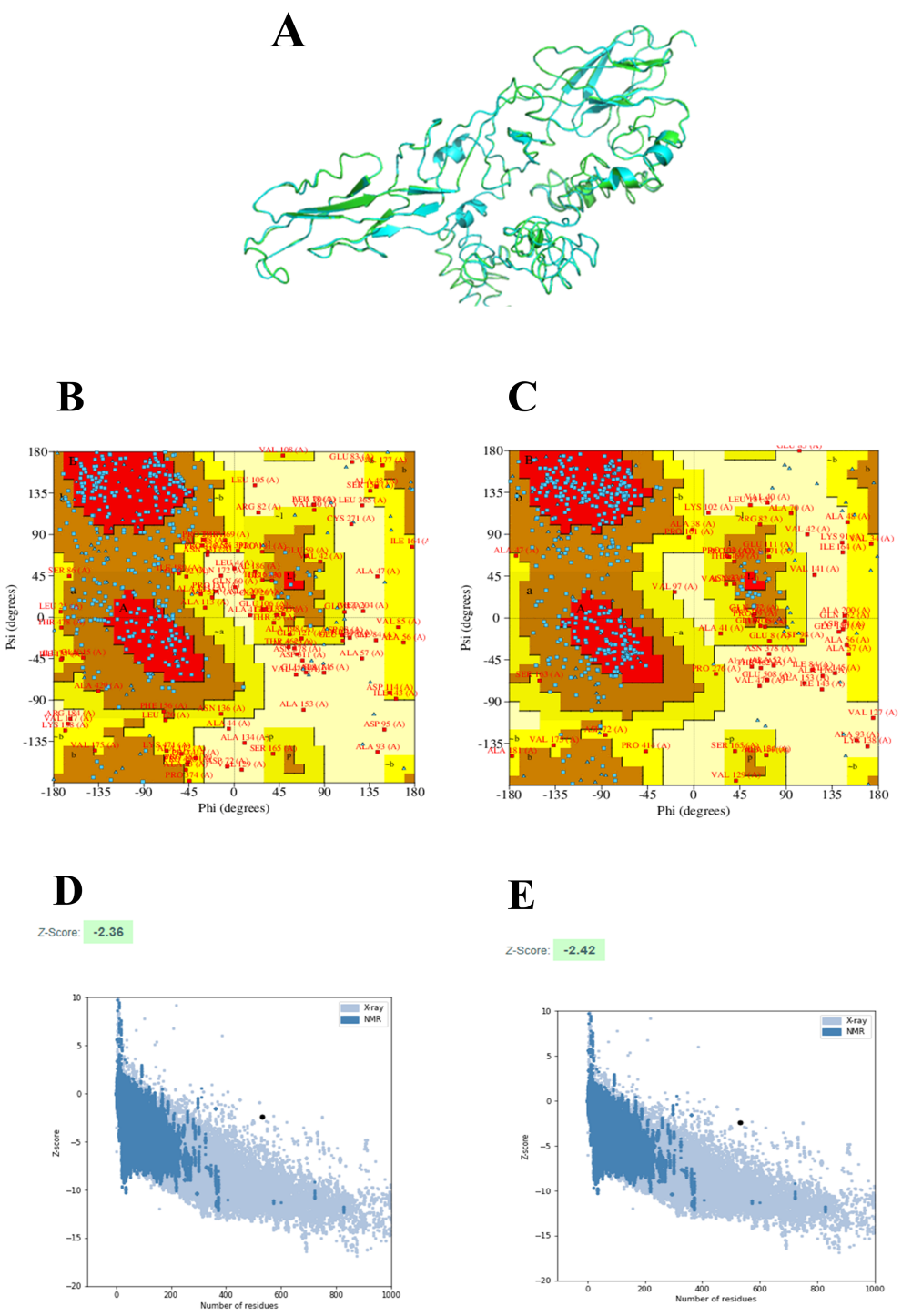


FIGURE 5
3D structure-refinement and validation: 3D structure of the designed multi-epitope vaccine construct in which the refined and unrefined models are shown in green color and cyan color, respectively **(A)**. Ramachandran plots of the unrefined **(B)** and refined models **(C)**. Z-score of the unrefined **(D)** and refined models **(E)**.

3.4 Designed multi-epitope vaccine showed significant binding affinities toward TLR4

The binding affinity of the multi-epitope vaccine construct toward the Toll-like receptor-4 (TLR4) was evaluated using the ClusPro 2.0 web server, which generated nearly 29 clusters of potential docked conformations, and the cluster 17 for its best conformations of the docked complex. The binding energy of the multi-epitope vaccine construct and TLR4 docked complex was observed to be −1,117.5 kcal/mol, indicating the high binding affinity and favorable interaction between TLR4 and the multi-

10.3389/fimmu.2025.1576557

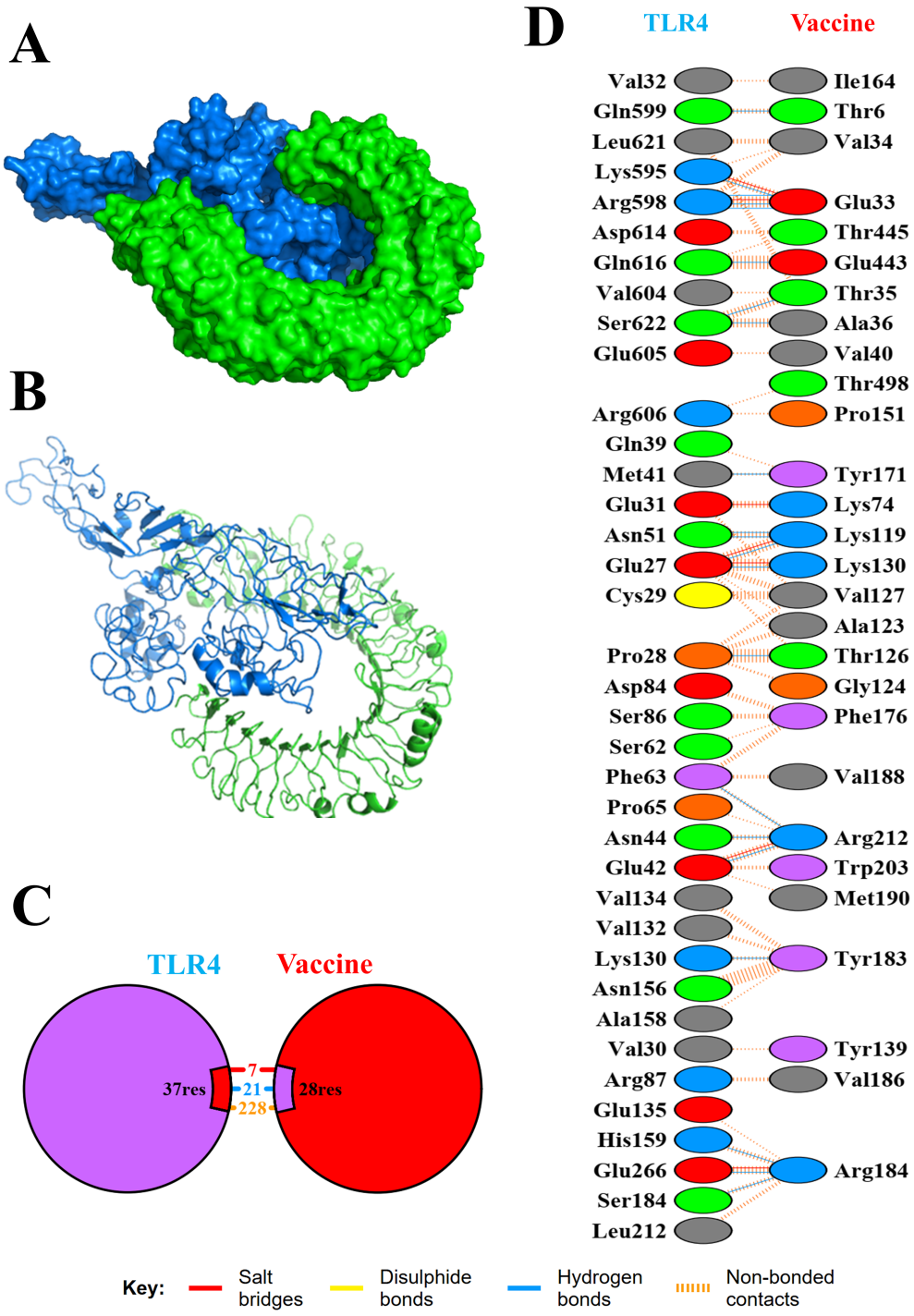


FIGURE 6
Molecular interaction of TLR4 with the designed multi-epitope vaccine construct. The docked complexes are shown in cartoon model **(A)** and surface model **(B)**. Also, total numbers of interactions **(C)** and interacting residues **(D)** of the TLR-4 vaccine complex are shown.

epitope vaccine construct. Upon binding, the vaccine showed a 1,762-Å^2 interface area with 28 interacting residues and the TLR4 showed a 1,671-Å^2 interface area with 37 interacting residues. Also, it revealed that it formed 7 salt bridges, 21 H-bonds, and 228 non-bonded contacts, as shown in Figure 6. Then, the molecular dynamics simulation trajectories were analyzed to study the conformational behavior of the TLR, vaccine, and TLR–vaccine complex over 100 ns. RMSD values were used to assess the local flexibility of the proteins, reflecting their atomic mobility. Higher RMSD values indicate increased mobility, whereas lower values suggest greater structural stability. During the simulation, the average RMSD values for the TLR, vaccine, and TLR–vaccine complex were 0.16, 0.22, and 0.28 nm, respectively. Then, RMSF plots revealed that TLR4 had fluctuations at the 120–170 AA and

310–320 AA regions, vaccine had fluctuations at the 320–325 AA region, and the TLR4–vaccine docked complexes exhibited the same fluctuations; however, these regions are denoted as loop regions. The average Rg values for the TLR, vaccine, and TLR–vaccine complex were 2.15, 2.16, and 2.18 nm, respectively, exhibiting the compactness of the structures. The average SASA values for the TLR, vaccine, and TLR–vaccine complex were 173, 176, and 186 nm^2, respectively, as shown in Figure 7.

3.5 NMA of the multi-epitope vaccine construct

The protein deformation analysis of the multi-epitope vaccine construct and TLR4 docked complex was predicted as normal mode analysis (NMA). The flexibility and stability of the docked complexes were evaluated from various plots such as B-factor/mobility, eigenvalue, variance, and co-variance map of the elastic network of the TLR4–vaccine complex, as illustrated in Figure 8. The B-factor/mobility indicates less deformation of the TLR4–vaccine complex at all amino acid residues and hinges, indicating that it maintains structural integrity. Notably, a lower eigenvalue of 2.77e−07 indicates less deformability of the docked complex, than the TLR4 alone, which showed an eigenvalue of 3.31e−05. In addition, the individual and cumulative variances indicate the contribution of each normal mode to the overall motion. The co-variance map revealed the presence of correlated, uncorrelated, and anti-correlated residue pairs, providing insights into the cooperative movements within the complex. Furthermore, flexibility was also observed from the elastic network.

3.6 Designed multi-epitope vaccine has the potential to induce immune response

The immune response simulation of the designed multi-epitope vaccine construct was predicted using the C-ImmSim web server at three dosage days. The immunological parameters such as the antibody titers, cytokine production, B-cell populations, B-cell populations per state, TH-cell populations, and TH-cell populations per state were predicted as shown in Figure 9. In the antibody titers plot, we have observed that IgG and IgM are significantly increased post-vaccine injection, indicating a robust humoral immune response. Also, the cytokine levels of IFN-γ were elevated notably, which suggested a strong activation of cellular immunity. Furthermore, the B-cell population (cells/m^3) was elevated, reflecting the activation and proliferation of B cells in response to the vaccine. The total TH-cell population (cells/m^3) also showed an increase, indicating enhanced helper T-cell responses, with a significant proportion of TH cells in active states, further

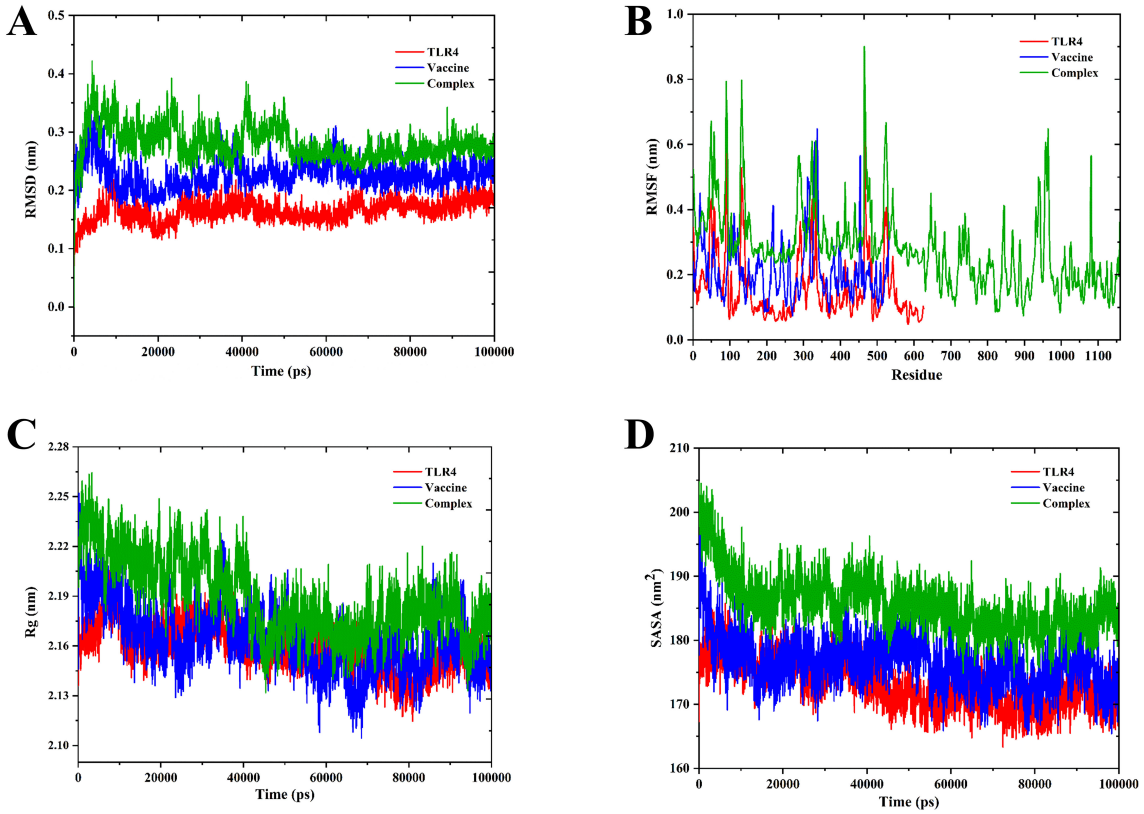


FIGURE 7
Molecular dynamics simulation of TLR4, vaccine, and docked complexes. The RMSD **(A)**, RMSF **(B)**, Rg **(C)**, and SASA **(D)** plots of TLR4, vaccine, and docked complexes are shown in red, blue, and green colors, respectively.

10.3389/fimmu.2025.1576557

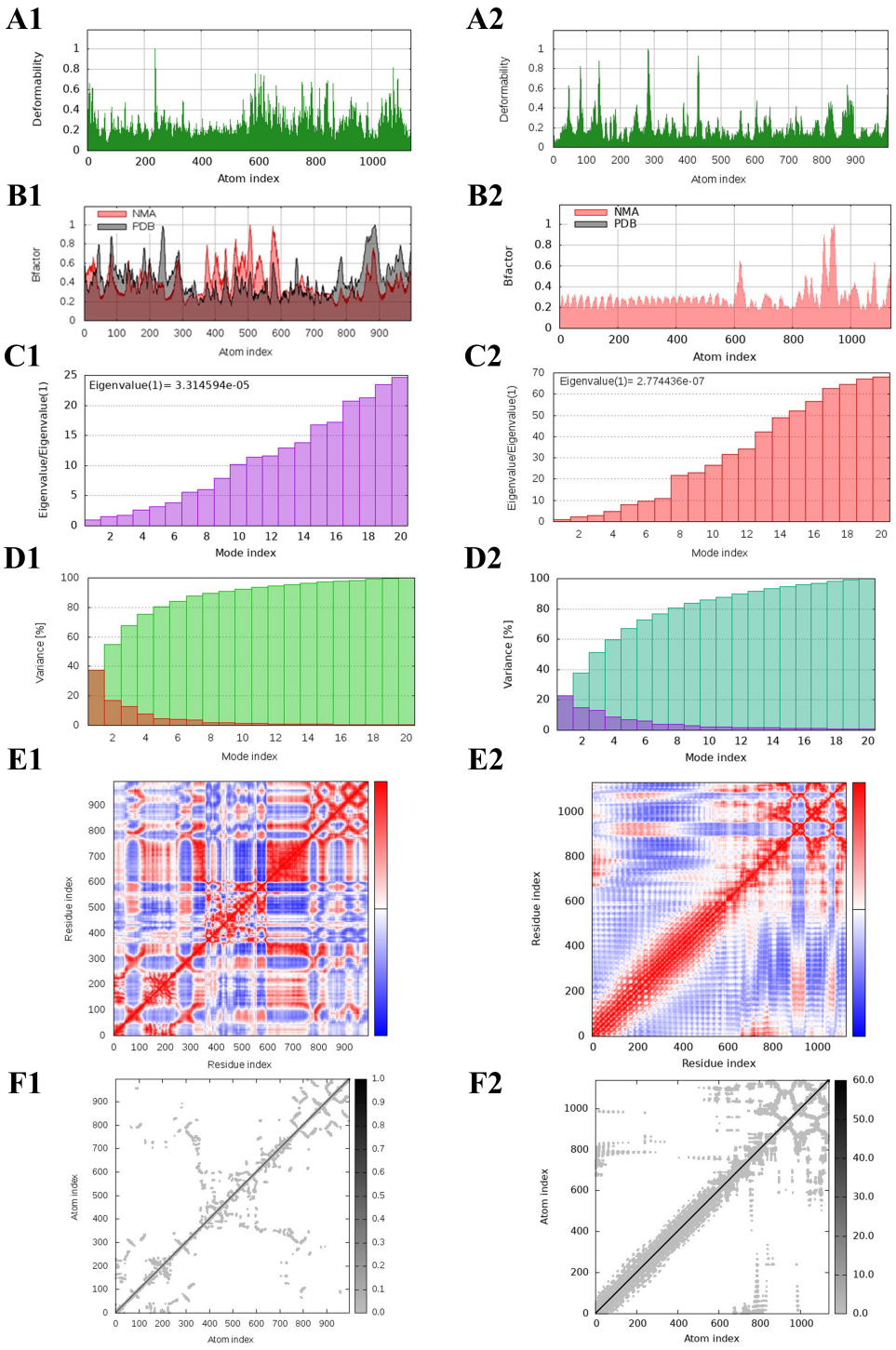


FIGURE 8
NMA of TLR4 and the designed multi-epitope vaccine construct–TLR4 docked complex. The deformability, B-factor, eigenvalues, variance, co-variance, and elastic network of the TLR4 **(A1-F1)** and TLR4–vaccine docked complex **(A2-F2)** are shown.

corroborating the vaccine's efficacy. Notably, all the predicted parameters such as the antibody titers (IgG and IgM), cytokine production (IFN-γ, ILs), B-cell populations (Total), B-cell populations per state (active state), TH-cell populations (Total), and TH-cell populations per state (active state) showed elevated peaks at the vaccine dosage days, indicating that the designed vaccine construct is highly efficient in inducing the immune responses in a time-dependent manner. These findings highlight the potential effectiveness of the multi-epitope vaccine construct in eliciting a comprehensive immune response, demonstrating its ability to induce both humoral and cellular immunity, which is crucial for long-term protection and memory formation.

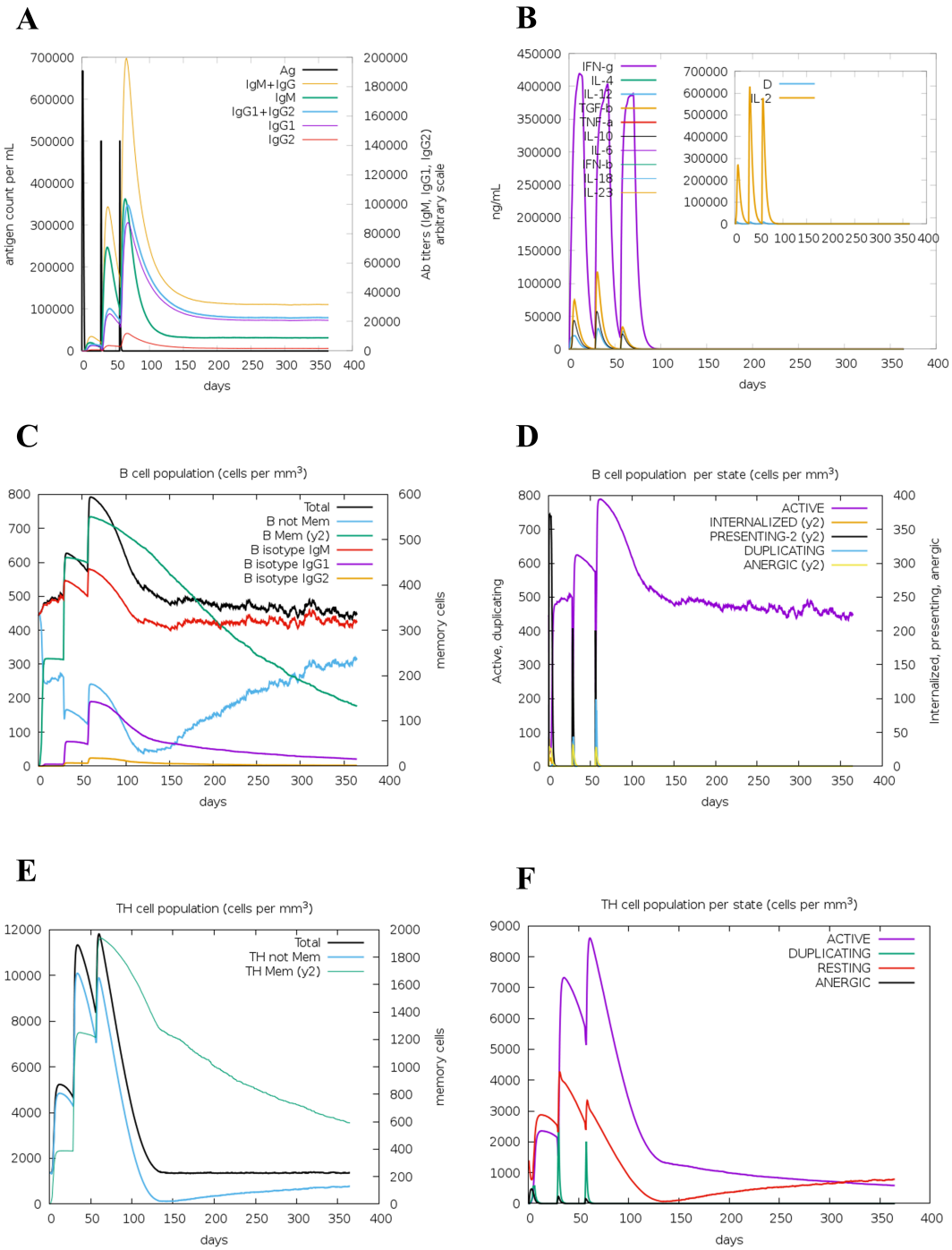


FIGURE 9
Immune simulation of the multi-epitope vaccine construct. Antibody titer **(A)**, cytokine production **(B)**, B-cell population **(C)**, B-cell population per state **(D)**, TH-cell population **(E)**, and TH-cell population per state **(F)** are shown.

3.7 Codon optimization and *insilico* cloning of the multi-epitope vaccine construct

The designed multi-epitope vaccine construct was further reverse-translated and optimized to be cloned by employing *Escherichia coli* K12 as an expression system. Then, the optimized sequence containing 552 nucleotides was obtained with the CAI-value as 1, and GC% as 53.8%. Also, the GC% of *E. coli* strain K12 was observed as 50.73%. Then, the restriction sites of SalI (GAGCTC) and EcoRI (GAATTC) were added at the N-terminal and C-terminal of the optimized DNA sequence. Following this, the optimized vaccine construct sequence (564 nucleotides) was cloned into the pET-28a (+) vector-6xHis-TEV-ORF9c (5554 bp) at restriction sites of SacI (GAGCTC) and EcoRI (GAATTC) using

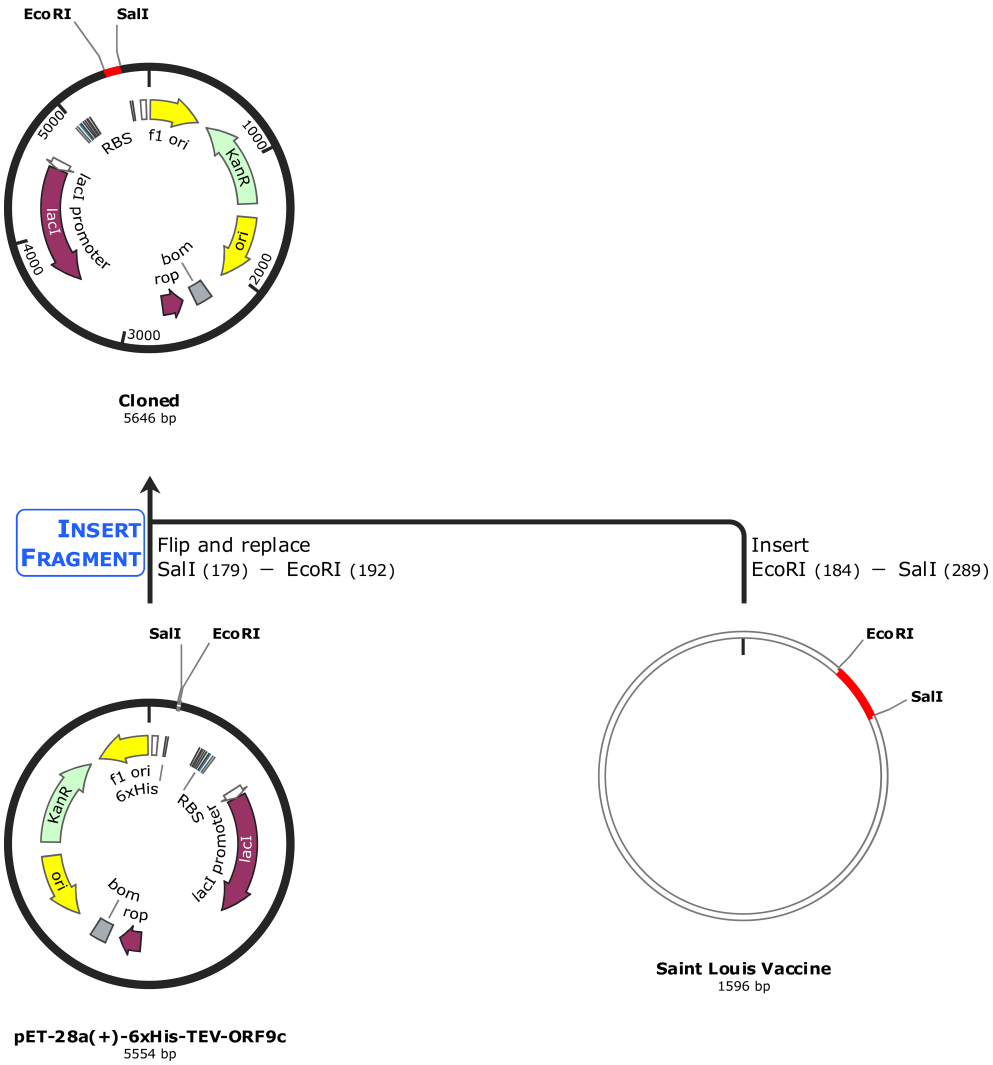


FIGURE 10
In silico cloning of the optimized vaccine construct. The codon-optimized multi-epitope vaccine sequence shown in red was cloned in the pET-28a (+) expression vector (5,554 bp) shown in black between restriction sites SalI and EcoRI, and the final cloned vaccine construct is shown (5,646 bp).

the SnapGene tool, and the final cloned product (5646 bp) is shown in Figure 10.

4 Discussion

Infectious diseases have posed significant challenges to human health throughout history, manifesting in acute, chronic, and often lethal forms caused by various pathogenic microorganisms with widespread morbidity and mortality worldwide (1, 3). The emergence of antimicrobial resistance (AMR) and antigenic shifts and drifts challenge our advances in the medical field (62, 63). The recurring outbreaks of SARS, MERS, and COVID-19 underscore the adaptive potential of RNA viruses, which can mutate to exploit new niches (11, 64). SLEV infection is strongly associated with potential central nervous system impairment that highly targets adults and still lacks potential treatment strategies (17). Alongside, the peptide vaccines constructed with multiple epitopes have recently gained attention due to their ability to amplify immune responses against pathogens (65). Both B cells and T cells can be used for vaccine development, but mostly T cell-based vaccines are preferred to some reasons such as high specificity adaptive immunity. CD8+ T cells uniquely recognize and eliminate infected cells via MHC, long-term immune memory, and broader immunological coverage. Notably, for viral infections that have antigenic variation, the B cell-mediated antibody responds less effectively (66). Although the traditional vaccines have the potency to induce strong humoral, cellular responses, and need fewer boosters than peptide vaccines, they are limited by their stability, risk of reversion to virulence, allergic reactions, and live-attenuated rapid mutation rates that lead to low efficacy in immune-compromised patients (67). On the other hand, the peptide vaccines are made of epitopes that specifically induce the stimulation of CTLs, HTLs, or B cells and have minimum off-target effects

indicating low adverse reactions. Also, peptide vaccines are easy to design, produce, and store; cost-effective; and, mostly importantly, safer for immunocompromised individuals (68, 69). Hossain et al. have also employed the immunoinformatics approach to design a vaccine against SLEV and showed that it has potential against SLEV. However, they have predicted the multi-epitopes only for the envelope protein E (outer membrane protein) (29). In addition, to the best of our knowledge, this is the only report we found for employing reverse vaccinology and immunoinformatics to design multi-epitope vaccine construct against SLEV. Thus, we have designed the multi-epitope vaccine construct toward various key proteins of SLEV such as the membrane glycoprotein M, envelope protein E, and the anchored capsid protein anchC to increase its therapeutic potential in SLEV treatment. So, in the present study, we have designed and developed a multi-epitope vaccine construct against the SLEV by employing reverse vaccinology and immunoinformatics approaches.

The cytotoxic T lymphocytes (CTLs) (CD8+ T cell epitopes) are involved in the recognition, direct killing, and clearance of the virally infected cells, whereas the helper T lymphocytes (HTLs) (CD4+ T cell epitopes) are involved in the immune activation, antibody production, and cytokine secretion respectively, and thus they play a vital role in the vaccine design (70, 71). Also, to elucidate a proper immune response, the epitopes should be antigenic, non-allergenic, and non-toxic and have the potential to induce IFN-γ (HTL epitope) production (72, 73). In our study, we have predicted the possible CTL and HTL epitopes against the various key proteins of SLEV such as the membrane glycoprotein M, envelope protein E, and the anchored capsid protein anchC, and we have selected 5 CTL epitopes and 17 HTL epitopes based on the abovementioned criteria, as shown in Supplementary Table S9. Additionally, the sequence conservation analysis was performed toward Dengue virus 1 (taxid:11053), Zika virus (taxid:64320), Yellow fever virus (taxid:11089), West Nile virus (taxid:11082), and Japanese encephalitis virus (taxid:11072), which are closely related to Flavivirus family. Notably, the West Nile virus and Japanese encephalitis virus shared similarity percentages of most predicted epitopes, and Dengue virus 1 and Zika virus shared similarity percentages with 1 CTL epitope and 1 HTL epitope. This similarity-conserved epitopes have the potential to induce cross-reactive T-cell responses and broaden protection toward other species such as West Nile virus and Japanese encephalitis virus, indicating that the developed vaccine construct was broad-spectrum (74). HLA-A*02:01 (MHC-I) and DRB1*01:01 (MHC-II) are the most frequently expressed alleles that could bind with CTL and HTL epitopes, respectively (75, 76). For instance, HLA-A*02:01 belongs to the A2 supertype possesses supertypic representation, which covers multiple related alleles, expanding their population coverage, whereas HLA-DRB1*01:01 is immunodominant, binds a broad spectrum of peptides, and significantly elicits CD4+ T-cell responses (76, 77). We observed that the selected CTL and HTL epitopes exhibited significant binding affinities toward their respective allele and their binding energy was predicted as shown in Table 1. Unlike the mRNA vaccines, the peptide vaccines have the advantage of adding adjuvants along with the peptides, which could induce more antigenic-mediated immune responses (78, 79). We have utilized the C-terminal region of the large ribosomal subunit protein bL12 of *Mycobacterium tuberculosis* as the adjuvant, which highly prevents the autoimmune reactions.

The linkers play a vital role in the designing of the vaccine construct to elucidate proper structural and functional properties. The EAAAK linker elevates the antigenic nature of the vaccine, the AAY linker promotes the presentation of antigens, and the GPGPG linker promotes solubility and movement (78, 80, 81). Likewise, the adjuvant was connected with EAAAK linkers, CTL epitopes were connected with AAY linkers, and the HTL epitopes were connected with GPGPG linkers, and the designed multi-epitope vaccine construct comprising 532 amino acids has 5 CTL epitopes, 17 HTL epitopes, 1 adjuvant, 4 CTL linkers, 16 HTL linkers, and 1 adjuvant linker, as shown in Figure 2. The designed vaccine should be stable to have a longer half-life period and immunogenicity retention and avoid degradation, and should be soluble to have an enhanced bioavailability nature, to prevent aggregation and for efficient delivery (82, 83). Based on these criteria, several vaccines have been designed and developed against various diseases and infections (31, 84–86). Similarly, we have observed that our designed multi-epitope vaccine construct was soluble with a probability of 0.902 and stable with an instability index of 23.84 (less than 40) indicating the stability of the vaccine. Also, we observed that the addition of adjuvant increased the vaccine's antigenic nature from 0.787234 to 0.898972, as shown in in Table 2.

Structural properties of the vaccine alter its functional properties such as the antigen presentation and stimulation of T lymphocytes and B lymphocytes (73). We have predicted the 3D structure of the designed vaccine construct and further refined it. Furthermore, we validated by the Ramachandran plot that showed highly favored regions in the refined model, and by the Z-score that showed high confidence in the refined 3D model. These analyses underscore the critical improvements in structural prediction and refinement processes, ensuring the vaccine construct's robustness and potential efficacy. The refined model's superior quality and stability are indicative of its potential to elicit a strong and effective immune response, thereby validating its design and functional applicability (87). Unlike the other Toll-like receptors (TLRs), Toll-like receptor-4 is observed to be overexpressed and also involved in various functions such as promoting the production of pro-inflammatory cytokine and chemokine and regulation of homeostasis, and thus plays a vital role in various diseases including SLEV infection (87, 88). Thus, we have docked our vaccine construct with TLR4, which showed significant binding affinities with a binding energy of −1,117.5 kcal/mol. The low binding energy profile suggests that the multi-epitope vaccine construct is likely to form a stable and effective complex with TLR4, potentially enhancing its immunogenic efficacy and contributing to a robust immune response. The molecular dynamics simulation (MDS) revealed that the TLR4-vaccine docked complex was stable throughout the simulation period compared with TLR4 and vaccine alone, indicating the structural compatibility of the docked complex as shown in Figure 7. Also, it indicated that

there is no flip on the residues of the TLR4–vaccine complex confirmed through MDS. Additionally, the TLR4–vaccine docked complex was also observed to be stable through the protein deformation analysis evaluated from various plots such as B-factor/mobility, eigenvalue, variance, and co-variance map of the elastic network of the TLR4–vaccine complex, as illustrated in Figure 8.

Generally, the vaccine-induced immune response is crucial, and multifaceted, encompassing both innate and adaptive immunity (89, 90). From our study, we observed that the designed multi-epitope vaccine construct elevates the levels of antibody titers, cytokine production, B-cell populations, B-cell populations per state, TH-cell populations, and TH-cell populations per state, as shown in Figure 9. These findings highlight the potential effectiveness of the multi-epitope vaccine construct in eliciting a comprehensive immune response, demonstrating its ability to induce both humoral and cellular immunity, which is crucial for long-term protection and memory formation. Also, for the experimental validation, the designed vaccine construct has to be produced in higher quantities, and thus usually it will be cloned in a suitable vector (91). In our study, the designed multi-epitope vaccine construct was reversed translated, codon-optimized, and cloned in a suitable vector pET-28a (+) vector-6xHis-TEV-ORF9c (5554 bp) at the restriction sites of SacI (GAGCTC) and EcoRI (GAATTC), as shown in Figure 10.

On the other hand, this study mostly used bioinformatics tools and databases for the study, and these computational validations may be less reliable when compared with the experimental validations (92, 93). For instance, the NetCTL 1.2 and NetMHCII 2.3 web servers mainly focus on the limited set of common HLA alleles, potentially overlooking epitopes relevant to underrepresented populations, whereas the VaxiJen v2.0, AllerTOP v.2, ToxinPred, IFNepitope, Expasy ProtParam, and ANTIGENpro web server are commonly used in immunoinformatics and vaccine design approaches; however, these predictions are based on a broad training dataset and do not yield high efficacy as the experimental validations (93, 94). Thus, we strongly recommend to validate the designed vaccine construct in *in vitro* and *in vivo* experimental settings to evaluate their completely therapeutic potential against SLEV. Overall, by employing reverse vaccinology and immunoinformatics approaches, we have designed a multi-epitope cancer vaccine against various key proteins of SLEV such as the membrane glycoprotein M, envelope protein E, and the anchored capsid protein anchC, and we further recommend evaluating its therapeutic potential by *in vitro* and *in vivo* studies in the near future. Furthermore, the deployment of these types of vaccines in regions where diseases are endemic offers significant opportunities to enhance public health and mitigate the disease burden (95). Achieving these outcomes, however, necessitates addressing complex logistical, sociocultural, and economic challenges through well-designed strategies and sustained international cooperation (96). These hurdles could be overcome by strengthening the infrastructure, community engagement, financial support, innovative delivery models, policy and governance, and integrated health programs. Effectively overcoming these barriers is critical to ensuring equitable vaccine access and advancing global objectives in health security and disease control.

5 Conclusion

SLEV infection poses a significant public health threat, particularly in regions prone to mosquito-borne diseases. Despite the availability of supportive treatments, there is a critical need for effective therapeutics/vaccines to prevent SLEV infections. In our study, we have designed, constructed, and validated a multi-epitope vaccine targeting key proteins of SLEV such as the membrane glycoprotein M, envelope protein E, and the anchored capsid protein anchC by employing reverse vaccinology and immunoinformatics approaches. Our results indicated that the vaccine construct is structurally stable, antigenic, non–allergic, and non–toxic and has soluble properties. Also, the vaccine exhibited strong binding affinity and structural compactness with the TLR4 upon binding confirmed by docking and molecular dynamics simulations respectively. Furthermore, it also indicated that it has the potential to induce an immune response. Also, it has been cloned in the pET–28a (+) expression vector for the experimental validation by *in vitro* and *in vivo* studies to evaluate the vaccine's therapeutic efficacy in the near future. Further research and experimental studies are warranted to validate the efficacy, safety, and immunogenicity of the proposed vaccine construct in preclinical and clinical settings.

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

PR: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. MA: Formal Analysis, Investigation, Software, Writing – review & editing, Visualization. MH: Data curation, Formal Analysis, Visualization, Writing – review & editing. SE: Data curation, Investigation, Methodology, Writing – review & editing. GC: Methodology, Resources, Software, Writing – review & editing. PB: Resources, Software, Validation, Writing – review & editing. JM: Resources, Validation, Writing – review & editing. KY: Supervision, Validation, Writing – review & editing. SA: Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. PR would like to thank the Council for Scientific and Industrial Research (CSIR) for providing him the Senior Research Fellowship (File No.: 09/0844(18240)/2024-

EMR-I). This work was supported by the Gachon University research fund of 2024(GCU-202403950001,202405290001) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03038996).

Acknowledgments

The authors would like to thank their respective institutes/organizations for providing the necessary facilities to carry out this work.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2025.1576557/full#supplementary-material

References

1. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. (2022) 20:193–205. doi: 10.1038/s41579-021-00639-z

2. IHME Pathogen Core Group. Global burden associated with 85 pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Infect Dis*. (2024) 24:868–95. doi: 10.1016/S1473-3099(24)00158-0

3. Zhang C, Fu X, Liu Y, Zhao H, Wang G. Burden of infectious diseases and bacterial antimicrobial resistance in China: a systematic analysis for the global burden of disease study 2019. *Lancet Reg Heal West Pacific*. (2024) 43:100972. doi: 10.1016/j.lanwpc.2023.100972

4. Diaz A, Coffey LL, Burkett-Cadena N, Day JF. Reemergence of st. Louis encephalitis virus in the americas. *Emerg Infect Dis*. (2018) 24:2150–7. doi: 10.3201/eid2412.180372

5. Kumata R, Sasaki A. Antigenic escape is accelerated by the presence of immunocompromised hosts. *Proc Biol Sci*. (2022) 289:20221437. doi: 10.1098/rspb.2022.1437

6. Sasaki A, Lion S, Boots M. Antigenic escape selects for the evolution of higher pathogen transmission and virulence. *Nat Ecol Evol*. (2022) 6:51–62. doi: 10.1038/s41559-021-01603-z

7. Salam MA, Al-Amin MY, Salam MT, Pawar JS, Akhter N, Rabaan AA, et al. Antimicrobial resistance: A growing serious threat for global public health. *Healthc (Basel Switzerland)*. (2023) 11. doi: 10.3390/healthcare11131946

8. Tang KWK, Millar BC, Moore JE. Antimicrobial resistance (AMR). *Br J BioMed Sci*. (2023) 80:11387. doi: 10.3389/bjbs.2023.11387

9. Ellwanger JH, Veiga ABGD, Kaminski VDL, Valverde-Villegas JM, Freitas AWQD, Chies JAB. Control and prevention of infectious diseases from a One Health perspective. *Genet Mol Biol*. (2021) 44:e20200256. doi: 10.1590/1678-4685-GMB-2020-0256

10. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nat Commun*. (2023) 14:8179. doi: 10.1038/s41467-023-43954-0

11. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-coV-2 and COVID-19. *Nat Rev Microbiol*. (2021) 19:141–54. doi: 10.1038/s41579-020-00459-7

12. Gowd KK, Veerababu D, Reddy VR. COVID-19 and the legislative response in India: The need for a comprehensive health care law. *J Public Aff*. (2021) 21:e2669. doi: 10.1002/pa.2669

13. Khalifa SAM, Swilam MM, El-Wahed AAA, Du M, El-Seedi HHR, Kai G, et al. Beyond the pandemic: COVID-19 pandemic changed the face of life. *Int J Environ Res Public Health*. (2021) 18. doi: 10.3390/ijerph18115645

14. Marou V, Vardavas CI, Aslanoglou K, Nikitara K, Plyta Z, Leonardi-Bee J, et al. The impact of conflict on infectious disease: a systematic literature review. *Confl Health*. (2024) 18:27. doi: 10.1186/s13031-023-00568-z

15. Zhang X-X, Jin Y-Z, Lu Y-H, Huang L-L, Wu C-X, Lv S, et al. Infectious disease control: from health security strengthening to health systems improvement at global level. *Glob Heal Res Policy*. (2023) 8:38. doi: 10.1186/s41256-023-00319-w

16. Zaman K, Khan SA, Chaturongakul S, Sahoo B, Novais Â. Editorial: One Health approaches to combat infectious diseases. *Front Microbiol*. (2023) 14:1321134. doi: 10.3389/fmicb.2023.1321134

17. Zoladek J, Nisole S. Mosquito-borne flaviviruses and type I interferon: catch me if you can! *Front Microbiol*. (2023) 14:1257024. doi: 10.3389/fmicb.2023.1257024

18. Swetnam DM, Stuart JB, Young K, Maharaj PD, Fang Y, Garcia S, et al. Movement of St. Louis encephalitis virus in the Western United States, 2014- 2018. *PloS Negl Trop Dis*. (2020) 14:e0008343. doi: 10.1371/journal.pntd.0008343

19. Tolsá-García MJ, Wehmeyer ML, Lühken R, Roiz D. Worldwide transmission and infection risk of mosquito vectors of West Nile, St. Louis encephalitis, Usutu and Japanese encephalitis viruses: a systematic review. *Sci Rep*. (2023) 13:308. doi: 10.1038/s41598-022-27236-1

20. Moser SK, Barnard M, Frantz RM, Spencer JA, Rodarte KA, Crooker IK, et al. Scoping review of Culex mosquito life history trait heterogeneity in response to temperature. *Parasit Vectors*. (2023) 16:200. doi: 10.1186/s13071-023-05792-3

21. Day JF. Predicting St. Louis encephalitis virus epidemics: lessons from recent, and not so recent, outbreaks. *Annu Rev Entomol*. (2001) 46:111–38. doi: 10.1146/annurev.ento.46.1.111

22. Lupenza E, Gasarasi DB, Minzi OM. Lymphatic filariasis, infection status in Culex quinquefasciatus and Anopheles species after six rounds of mass drug administration in Masasi District, Tanzania. *Infect Dis poverty*. (2021) 10:20. doi: 10.1186/s40249-021-00808-5

23. Habarugira G, Suen WW, Hobson-Peters J, Hall RA, Bielefeldt-Ohmann H. West nile virus: an update on pathobiology, epidemiology, diagnostics, control and "One health" Implications. *Pathog (Basel Switzerland)*. (2020) 9. doi: 10.3390/pathogens9070589

24. Amor S. Virus infections of the central nervous system. *Manson's Trop Dis*. (2009), 853–83. doi: 10.1016/B978-1-4160-4470-3.50052-5

25. Yang D, Li X-J, Tu D-Z, Li X-L, Wei B. Advances in viral encephalitis: Viral transmission, host immunity, and experimental animal models. *Zool Res*. (2023) 44:525–42. doi: 10.24272/j.issn.2095-8137.2023.025

26. Danforth ME, Snyder RE, Feiszli T, Bullick T, Messenger S, Hanson C, et al. Epidemiologic and environmental characterization of the Re-emergence of St. Louis Encephalitis Virus in California, 2015-2020. *PloS Negl Trop Dis*. (2022) 16:e0010664. doi: 10.1371/journal.pntd.0010664

27. Kumar R. Understanding and managing acute encephalitis. *F1000Research*. (2020) 9. doi: 10.12688/f1000research.20634.1

28. Curren EJ, Lindsey NP, Fischer M, Hills SL. St. Louis encephalitis virus disease in the United States, 2003-2017. *Am J Trop Med Hyg*. (2018) 99:1074–9. doi: 10.4269/ajtmh.18-0420

29. Hossain MS, Hossan MI, Mizan S, Moin AT, Yasmin F, Akash AS, et al. Immunoinformatics approach to designing a multi-epitope vaccine against Saint Louis Encephalitis Virus. *Inf Med Unlocked*. (2021) 22:100500. doi: 10.1016/J.IMU.2020.100500

30. Blaney JEJ, Speicher J, Hanson CT, Sathe NS, Whitehead SS, Murphy BR, et al. Evaluation of St. Louis encephalitis virus/dengue virus type 4 antigenic chimeric viruses in mice and rhesus monkeys. *Vaccine*. (2008) 26:4150–9. doi: 10.1016/j.vaccine.2008.05.075

31. Alsowayeh N, Albutti A, Al-Shouli ST. Reverse vaccinology and immunoinformatic assisted designing of a multi-epitopes based vaccine against nosocomial burkholderia cepacia. *Front Microbiol*. (2022) 13:929400. doi: 10.3389/fmicb.2022.929400

32. Ramalingam PS, Arumugam S. Reverse vaccinology and immunoinformatics approaches to design multi-epitope based vaccine against oncogenic KRAS. *Med Oncol*. (2023) 40:283. doi: 10.1007/s12032-023-02160-0

33. Rice CM, Aebersold R, Teplow DB, Pata J, Bell JR, Vorndam AV, et al. Strauss JH. Partial N-terminal amino acid sequences of three nonstructural proteins of two flaviviruses. *Virology*. (1986) 151:1–9. doi: 10.1016/0042-6822(86)90098-x

34. Khan AR, Baker BM, Ghosh P, Biddison WE, Wiley DC. The structure and stability of an HLA-A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. *J Immunol*. (2000) 164:6398–405. doi: 10.4049/jimmunol.164.12.6398

35. Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure*. (1997) 5:1385–96. doi: 10.1016/s0969-2126(97)00288-8

36. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf*. (2007) 8:424. doi: 10.1186/1471-2105-8-424

37. 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

38. Vita R, Blazeska N, Marrama D, Duesing S, Bennett J, Greenbaum J, et al. The immune epitope database (IEDB): 2024 update. *Nucleic Acids Res*. (2025) 53:D436–43. doi: 10.1093/nar/gkae1092

39. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf*. (2007) 8:4. doi: 10.1186/1471-2105-8-4

40. 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

41. Sharma N, Naorem LD, Jain S, Raghava GPS. ToxinPred2: an improved method for predicting toxicity of proteins. *Brief Bioinform*. (2022) 23. doi: 10.1093/bib/bbac174

42. Dhanda SK, Vir P, Raghava GPS. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct*. (2013) 8:30. doi: 10.1186/1745-6150-8-30

43. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res*. (2016) 44:W449–54. doi: 10.1093/nar/gkw329

44. Zhou P, Jin B, Li H, Huang S-Y. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res*. (2018) 46:W443–50. doi: 10.1093/nar/gky357

45. Ayyagari VS, Srirama K. Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *J Biomol Struct Dyn*. (2022) 40:2963–77. doi: 10.1080/07391102.2020.1850357. T C V, K AP.

46. Zaib S, Akram F, Liaqat ST, Altaf MZ, Khan I, Dera AA, et al. Bioinformatics approach for the construction of multiple epitope vaccine against omicron variant of SARS-CoV-2. *Sci Rep*. (2022) 12:19087. doi: 10.1038/s41598-022-23550-w

47. 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*. (1999) 112:531–52. doi: 10.1385/1-59259-584-7:531

48. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics*. (2009) 25:2200–7. doi: 10.1093/bioinformatics/btp386

49. Laskowski RA, Thornton JM. PDBsum extras: SARS-CoV-2 and AlphaFold models. *Protein Sci*. (2022) 31:283–9. doi: 10.1002/pro.4238

50. Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, et al. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat Protoc*. (2022) 17:2326–53. doi: 10.1038/s41596-022-00728-0

51. 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

52. 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

53. 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

54. Miryala SK, Anbarasu A, Ramaiah S. Systems biology studies in Pseudomonas aeruginosa PA01 to understand their role in biofilm formation and multidrug efflux pumps. *Microb Pathog*. (2019) 136:103668. doi: 10.1016/j.micpath.2019.103668

55. Miryala SK, Anbarasu A, Ramaiah S. Role of SHV-11, a class A β-lactamase, gene in multidrug resistance among klebsiella pneumoniae strains and understanding its mechanism by gene network analysis. *Microb Drug Resist*. (2020) 26:900–8. doi: 10.1089/mdr.2019.0430

56. Miryala SK, Anbarasu A, Ramaiah S. Gene interaction network to unravel the role of gut bacterial species in cardiovascular diseases: E. coli O157:H7 host-bacterial interaction study. *Comput Biol Med*. (2021) 133:104417. doi: 10.1016/j.compbiomed.2021.104417

57. Miryala SK, Anbarasu A, Ramaiah S. Organ-specific host differential gene expression analysis in systemic candidiasis: A systems biology approach. *Microb Pathog*. (2022) 169:105677. doi: 10.1016/j.micpath.2022.105677

58. 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

59. Chakraborty A, Bayry J, Mukherjee S. Immunoinformatics approaches in designing vaccines against COVID-19. *Methods Mol Biol*. (2023) 2673:431–52. doi: 10.1007/978-1-0716-3239-0_29

60. Rapin N, Lund O, Castiglione F. Immune system simulation online. *Bioinformatics*. (2011) 27:2013–4. doi: 10.1093/bioinformatics/btr335

61. 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

62. Sharma S, Chauhan A, Ranjan A, Mathkor DM, Haque S, Ramniwas S, et al. Emerging challenges in antimicrobial resistance: implications for pathogenic microorganisms, novel antibiotics, and their impact on sustainability. *Front Microbiol*. (2024) 15:1403168. doi: 10.3389/fmicb.2024.1403168

63. Ho CS, Wong CTH, Aung TT, Lakshminarayanan R, Mehta JS, Rauz S, et al. Antimicrobial resistance: a concise update. *Lancet Microbe*. (2024), 6:100947. doi: 10.1016/j.lanmic.2024.07.010

64. Luu B, McCoy-Hass V, Kadiu T, Ngo V, Kadiu S, Lien J. Severe acute respiratory syndrome associated infections. *Physician Assist Clin*. (2023) 8:495–530. doi: 10.1016/j.cpha.2023.03.002

65. 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

66. Song X, Li Y, Wu H, Qiu H, Sun Y. T-cell epitope-based vaccines: A promising strategy for prevention of infectious diseases. *Vaccines*. (2024) 12:1181. doi: 10.3390/VACCINES12101181

67. Gupta S, Pellett S. Recent developments in vaccine design: from live vaccines to recombinant toxin vaccines. *Toxins (Basel)*. (2023) 15. doi: 10.3390/toxins15090563

68. Black M, Trent A, Tirrell M, Olive C. Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists. *Expert Rev Vaccines*. (2010) 9:157–73. doi: 10.1586/erv.09.160

69. Kalita P, Tripathi T. Methodological advances in the design of peptide-based vaccines. *Drug Discov Today*. (2022) 27:1367–80. doi: 10.1016/j.drudis.2022.03.004

70. Banerjee S, Majumder K, Gutierrez GJ, Gupta D, Mittal B. Immuno-informatics approach for multi-epitope vaccine designing against SARS-CoV-2. *bioRxiv Prepr Serv Biol*. (2020). doi: 10.1101/2020.07.23.218529

71. Khan MS, Khan IM, Ahmad SU, Rahman I, Khan MZ, Khan MSZ, et al. Immunoinformatics design of B and T-cell epitope-based SARS-CoV-2 peptide vaccination. *Front Immunol*. (2022) 13:1001430. doi: 10.3389/fimmu.2022.1001430

72. Lari A, Lari N, Biabangard A. Immunoinformatics approach to design a novel subunit vaccine against visceral leishmaniasis. *Int J Pept Res Ther*. (2022) 28:34. doi: 10.1007/s10989-021-10344-3

73. Sanami S, Rafieian-Kopaei M, Dehkordi KA, Pazoki-Toroudi H, Azadegan-Dehkordi F, Mobini G-R, et al. In silico design of a multi-epitope vaccine against HPV16/18. *BMC Bioinf*. (2022) 23:311. doi: 10.1186/s12859-022-04784-x

74. Hou B, Chen H, Gao N, An J. Cross-reactive immunity among five medically important mosquito-borne flaviviruses related to human diseases. *Viruses*. (2022) 14. doi: 10.3390/v14061213

75. Singh A, Thakur M, Sharma LK, Chandra K. Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci Rep*. (2020) 10:16219. doi: 10.1038/s41598-020-73371-y

76. Pedersen SR, Christensen JP, Buus S, Rasmussen M, Korsholm KS, Nielsen M, et al. Immunogenicity of HLA class I and II double restricted influenza A-derived peptides. *PloS One*. (2016) 11:e0145629. doi: 10.1371/journal.pone.0145629

77. Kohlgruber AC, Dezfulian MH, Sie BM, Wang CI, Kula T, Laserson U, et al. High-throughput discovery of MHC class I- and II-restricted T cell epitopes using synthetic cellular circuits. *Nat Biotechnol*. (2024) 43(4):623–34. doi: 10.1038/s41587-024-02248-6

78. Ebrahimi SM, Tebianian M, Toghyani H, Memarnejadian A, Attaran HR. Cloning, expression and purification of the influenza A (H9N2) virus M2e antigen and truncated Mycobacterium tuberculosis HSP70 as a fusion protein in Pichia pastoris. *Protein Expr Purif*. (2010) 70:7–12. doi: 10.1016/j.pep.2009.11.001

79. Javid B, MacAry PA, Oehlmann W, Singh M, Lehner PJ. Peptides complexed with the protein HSP70 generate efficient human cytolytic T-lymphocyte responses. *Biochem Soc Trans*. (2004) 32:622–5. doi: 10.1042/BST0320622

80. 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

81. Kavoosi M, Creagh AL, Kilburn DG, Haynes CA. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in Escherichia coli. *Biotechnol Bioeng*. (2007) 98:599–610. doi: 10.1002/bit.21396

82. Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, et al. Immunoinformatics and vaccine development: an overview. *ImmunoTargets Ther*. (2020) 9:13–30. doi: 10.2147/ITT.S241064

83. Adam KM. Immunoinformatics approach for multi-epitope vaccine design against structural proteins and ORF1a polyprotein of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). *Trop Dis Travel Med vaccines*. (2021) 7:22. doi: 10.1186/s40794-021-00147-1

84. Beikzadeh B. Immunoinformatics design of multi-epitope vaccine using OmpA, OmpD and enterotoxin against non-typhoidal salmonellosis. *BMC Bioinf*. (2023) 24:63. doi: 10.1186/s12859-023-05183-6

85. Prawiningrum AF, Paramita RI, Panigoro SS. Immunoinformatics approach for epitope-based vaccine design: key steps for breast cancer vaccine. *Diagnostics (Basel Switzerland)*. (2022) 12. doi: 10.3390/diagnostics12122981

86. 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) 11:15431. doi: 10.1038/s41598-021-92176-1

87. Wei J, Zhang Y, Li H, Wang F, Yao S. Toll-like receptor 4: A potential therapeutic target for multiple human diseases. *BioMed Pharmacother*. (2023) 166:115338. doi: 10.1016/j.biopha.2023.115338

88. Kim H-J, Kim H, Lee J-H, Hwangbo C. Toll-like receptor 4 (TLR4): new insight immune and aging. *Immun Ageing*. (2023) 20:67. doi: 10.1186/s12979-023-00383-3

89. Arunachalam AB. Vaccines induce homeostatic immunity, generating several secondary benefits. *Vaccines*. (2024) 12. doi: 10.3390/vaccines12040396

90. Larsen SE, Williams BD, Rais M, Coler RN, Baldwin SL. It takes a village: the multifaceted immune response to mycobacterium tuberculosis infection and vaccine-induced immunity. *Front Immunol*. (2022) 13:840225. doi: 10.3389/fimmu.2022.840225

91. Gauci C, Merli M, Muller V, Chow C, Yagi K, Mackenstedt U, et al. Molecular cloning of a vaccine antigen against infection with the larval stage of Echinococcus multilocularis. *Infect Immun*. (2002) 70:3969–72. doi: 10.1128/IAI.70.7.3969-3972.2002

92. Premkumar T, Sajitha Lulu S. Molecular crosstalk between COVID-19 and Alzheimer's disease using microarray and RNA-seq datasets: A system biology approach. *Front Med*. (2023) 10:1151046. doi: 10.3389/fmed.2023.1151046

93. Sasikumar DSN, Thiruselvam P, Sundararajan V, Ravindran R, Gunasekaran S, Madathil D, et al. Insights into dietary phytochemicals targeting Parkinson's disease key genes and pathways: A network pharmacology approach. *Comput Biol Med*. (2024) 172:108195. doi: 10.1016/j.compbiomed.2024.108195

94. Ramalingam PS, Raj DBTG, Subramanian M, Arumugam S. ML-Based Screening of miRNA Inhibitors and Intervention of lncRNA/miRNA/mRNA Axis in Oncogenic KRAS-Associated Colorectal Cancer. *Adv Public Heal*. (2024) 2024:9436238. doi: 10.1155/2024/9436238

95. Gupta SS, Nair GB, Arora NK, Ganguly NK. Vaccine development and deployment: opportunities and challenges in India. *Vaccine*. (2013) 31 Suppl 2:B43–53. doi: 10.1016/j.vaccine.2012.11.079

96. Kumraj G, Pathak S, Shah S, Majumder P, Jain J, Bhati D, et al. Capacity building for vaccine manufacturing across developing countries: the way forward. *Hum Vaccin Immunother*. (2022) 18:2020529. doi: 10.1080/21645515.2021.2020529

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 20 May 2025
DOI 10.3389/fimmu.2025.1540253

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College & Hospital, India

REVIEWED BY
Patrícia C. C. Neves,
Oswaldo Cruz Foundation, Brazil
Vinothkannan Ravichandran,
Amity University, Mumbai, India
Satyavani Kaliamurthi,
Saveetha Institute of Medical and Technical Sciences, India

*CORRESPONDENCE
John J. Georrge
johnjgeorrge@gmail.com

RECEIVED 05 December 2024
ACCEPTED 26 March 2025
PUBLISHED 20 May 2025

CITATION
Kumar Mishra S, Kumar N, Or Rashid MH, Sultana S, Dawoud TM, Bourhia M and Georrge JJ (2025) An integrated mutation-based immunoinformatic approach incorporating variability in epitopes: a study based on HIV subtype C. *Front. Immunol.* 16:1540253. doi: 10.3389/fimmu.2025.1540253

An integrated mutation-based immunoinformatic approach incorporating variability in epitopes: a study based on HIV subtype C

Saurav Kumar Mishra[1], Neeraj Kumar[2], Md. Harun Or Rashid[3,4], Sharifa Sultana[4], Turki M. Dawoud[5], Mohammed Bourhia[6] and John J. Georrge[1]*

[1]Department of Bioinformatics, University of North Bengal, Darjeeling, West Bengal, India, [2]Department of Pharmaceutical Chemistry, Bhupal Nobles' College of Pharmacy, Udaipur, Rajasthan, India, [3]School of Engineering, Macquarie University, Sydney, NSW, Australia, [4]Computational Biology Research Laboratory, Department of Pharmacy, Faculty of Health and Life Sciences, Daffodil International University, Dhaka, Bangladesh, [5]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia, [6]Department of Chemistry and Biochemistry, Faculty of Medicine and Pharmacy, Ibn Zohr University, Laayoune, Morocco

Currently, HIV (human immunodeficiency virus) infection is one of the leading complications in public health and causes acquired immunodeficiency syndrome (AIDS), especially in the African region. No specific vaccine is available to combat this, with multi-strain variability being one of the hurdles. In this investigation, we employed variability in the epitope of the HIV subtype C targets to introduce mutations and construct an epitope-based vaccine. Four targets were examined to predict the B and T cells (major histocompatibility complex class I and II). Among the predicted epitopes, immunodominant epitopes were selected and were mapped with the identified variable amino acid to incorporate mutation. These selected and mutated epitopes were used for the non-mutated and mutated vaccine construction, considering linker for fusion and adjuvant to improve the activity. The vaccine's structure was modeled and examined to validate its structural quality, and a high population coverage was also found. The docking investigation of the non-mutated and mutated vaccine with Toll-like receptor 3 shows remarkable activity followed by strong binding affinity, and the simulation of over 100 ns revealed the constancy of the complex system. The immune response revealed its strong effectiveness by generating multiple immunoglobulins followed by the time step of infection, and further, *in silico* cloning demonstrated a high expression in *Escherichia coli* based on their favorable Codon Adaptation Index and GC value. The integrated approach in this investigation will help to plan a potent immunodominant vaccine that can work for multiple strains of HIV infection.

KEYWORDS

HIV, epitope, mutation, vaccine, TLR3, simulation

Introduction

Acquired immunodeficiency syndrome (AIDS) is an ongoing public health concern caused by HIV (1, 2). According to a recent World Health Organization (WHO) report, nearly 39.9 million are living with HIV; at the end of 2023, approximately 1.3 million people acquired the infection, while 63,000 died due to HIV-related complications (https://www.who.int/news-room/fact-sheets/detail/hiv-aids) (3). Moreover, the WHO African region remains to have the highest HIV burden (https://www.who.int/data/gho/data/themes/hiv-aids). The AIDS pandemic is led by the two types of HIV, i.e., HIV-1 and HIV-2, with the former being more prevalent than the latter (4, 5). Moreover, several antiretroviral therapies (ARTs) were designed and used, but none of them will lead to combatting this infection completely (6, 7). A few vaccines were developed to combat this, but they did not prove efficient due to a lack of appropriate immune response and effectiveness. Apart from that, one of the hurdles behind the efficiency is the variability and the mutation within the strains. Despite the various hurdles, the most effective vaccine, RV144, was developed, providing only 31.2% protection against this infection (7–9). This emphasizes the need to create a potent vaccine to address the challenges of strain variability due to mutations in controlling HIV infection. HIV-1 is classified into four (M, N, O, and P) groups; among them, only group M causes 95% to be classified into various subtypes (A, B, C, D, F, G, H, J, and K) (10–12). However, subtype B is prevalent in Australia, America, and Western Europe, whereas subtype C is prevalent in Africa and India (10, 11). Moreover, subtype C is the most prevalent strain worldwide (46.6%) and dominates in Asia and Africa, followed by subtypes A and B (13–15). Furthermore, a recent systematic review reported that subtype C accounted for 50.4% of worldwide HIV based on data (from 2016 to 2021) and found a significant increase in the cases compared to the previous dataset (from 2010 to 2015) (16). HIV employs various strategies to evade immune surveillance, including antigenic variation, MHC downregulation, and immune cell dysfunction (17). Subsequently, several key mutations are mainly responsible for escaping immune mechanisms, such as N332 glycan shift (escape broadly neutralizing antibodies by altering glycan shielding) and T242N (reduces recognition by CTL), among others (18, 19).

The HIV genome comprises several effective structural, regulator, and accessory genes. However, structural genes, i.e., the envelope glycoprotein, protease, reverse transcriptase, and integrase, are crucial for host–pathogen interaction and its replication (5, 11, 20, 21). Their role in viral mechanisms makes them an ideal candidate for therapeutic development. At present, using immunological data, immunoinformatics-assisted vaccine design has been identified as a suitable strategy, along with reverse vaccinology and advanced computational approaches (11, 22–25) targeting several other pathogens, because time efficiency, cost-effectiveness, and high accuracy are essential for a successful vaccine design. Immunoinformatics-assisted studies on HIV have successfully targeted various components, including gp120 (21), the whole HIV genome (5), and Gag polyprotein (7), among others (11, 20), highlighting the reliability of this approach without incorporating the variability of epitopes. In addition, no such study was reported on subtype C that contributed to higher HIV infection than the other strains in group M. The main hurdle in combating HIV is the strain variability caused by mutations, which has yet to be fully addressed (26–28) and also remains a major obstacle behind the successful vaccine formulation and the effective potent activity of the available vaccine. Moreover, a few studies were designed to incorporate mutation within epitopes against COVID-19 successfully (22, 29). Compared to conventional vaccine design methods, the advantage of employing immunoinformatics enables the screening of potential epitopes that are effective for multiple strains, the assessment of their immune activity, and other key factors to enhance vaccine development, which is vital for the effective vaccine design (5, 7).

Therefore, this study examined subtype C to formulate a potent vaccine considering variability. The B cells and MHC class I and class II (MHC I and MHC II) epitopes were identified and selected based on their high antigenicity score in this investigation. These epitopes were further mapped considering the variability of amino acids identified via multiple strains. The non-mutated and mutated vaccines were formulated, and their molecular activity and stability toward the TRL were analyzed via docking and dynamics. The immune activity based on the vaccination steps and the expression of the formulated vaccine were performed and analyzed.

Materials and methods

The employed steps corresponding to the methodology are illustrated in Figure 1.

Collection of the target sequence and their immune assessment

The vital target (essential for host–pathogen interaction, replication, and pathogenesis) sequences within the HIV mechanism were retrieved from UniProt (https://www.uniprot.org/). The vaccine protein must have strong immunological properties and be non-allergenic to confirm a potent immune response (29). These retrieved sequences were further subjected to the antigen and allergen assessment via VaxiJen v2.0 (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (30) considering virus as a target and a threshold value of 0.4 and the AllerTOP v.2.0 (https://www.ddg-pharmfac.net/allertop_test/) (31) server. The VaxiJen server is mainly based on alignment-based prediction methods, while the AllerTOP server is alignment-free and grounded on the target's physicochemical properties.

Identification of B-cell epitope

Two subsequent servers—ABCpred (http://crdd.osdd.net/raghava/abcpred/) (32), which utilized the artificial neural network, and BepiPred 2.0 (http://tools.iedb.org/bcell/) grounded on the sequence features of the antigen (33), available at IEDB—

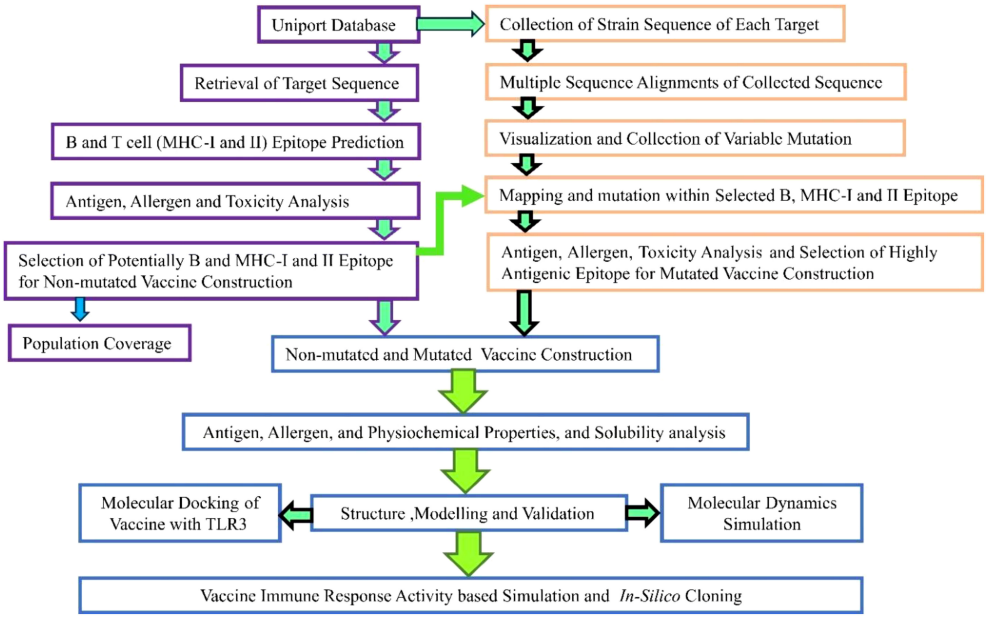


FIGURE 1
Overview of employed steps in the designed study.

have different algorithms to detect more potential linear B-lymphocyte (LBL) epitopes considering collected sequences as input with default parameters. For peptide vaccines, recognizing B cells is crucial, as their receptors recognize peptides to trigger an effective immune response (29). However, the epitopes were further considered based on their presence in both servers and examined via VaxiJen v2.0 (30), AllerTOP v.2.0 (31), and ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/) along with default parameter (34) servers.

Identification of T-cell (MHC I and MHC II) epitopes and their immune assessment

$CD8^+$ T lymphocytes recognize MHC I epitopes. When a cell is infected or has aberrant proteins (such as in viral infections or cancer), MHC I molecules present these peptides on the cell surface, prompting $CD8^+$ T cells to kill the infected or abnormal cells (35–37). On the other hand, $CD4^+$ helper T cells recognize MHC II epitopes. Antigen-presenting cells (APCs) internalize and process foreign antigens, presenting peptides on MHC II molecules to activate $CD4^+$ T cells, which then help coordinate the broader immune response (38, 39). The MHC I and MHC II within the targets were identified using Tepitool (http://tools.iedb.org/tepitool/), which computes the epitopes based on seven prediction methods (IEDB recommended, consensus, NetMHCIIpan, NN-align, SMM-align, Sturniolo, and the combinatorial library method) (40). For MHC I, 27 and MHC II, 7, the most frequent alleles with the restricted 9- and 15-mer length were selected, and all other IEDB-recommended parameters were selected (24, 29, 40). Furthermore, the immune assessment was done similarly to the abovementioned one to screen out the potential epitopes.

Epitope mapping of B and T cells with the variable amino acid

To formulate a mutation-proof vaccine, the designed vaccine should be highly effective in both mutated and non-mutated forms (29, 41). The available sequence concerning each target was collected from UniProt (https://www.uniprot.org/). These sequences were subjected for multiple sequence alignments via Clustal Omega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo), which is based on the seeded guide trees and the HMM technique (42), and the variable amino acid was visualized and collected using the JalView (43) software. These variable amino acids were further mapped with the final selected B- and T-cell epitopes to incorporate mutation.

Vaccine formulation and immune and physiological assessments

The highly antigenic score followed by non-allergenic and non-toxic-based LBL, MHC I, and MHC II epitopes were selected from each target for the vaccine formulation, leading to a robust immune response against the infection. These epitopes were joined via different subsequent linkers (EAAAK, AAY, KK, and GPGPG) (21, 41). Furthermore, to enhance, activate, and purify, the adjuvant, PADRE, and His-tag were also attached at the N and C terminals of the vaccine construct. In contrast, His-tag was attached using the RVRR linkers (5, 7, 11, 21). Moreover, considering combination, six different vaccines were constructed to identify additional potential combinations with high antigenic properties (score). However, the adjuvant (beta-defensin), PADRE at the N, and His-tag at the C terminal were kept in different distinct vaccine constructs (44).

Moreover, the EAAAK offers an extended, uncharged spacer that can reduce steric hindrance in the region, AAY enhances the immunogenicity and improves pathogen-specific immunity while reducing junctional immunogenicity, KK linkers enhance solubility and are crucial proteases required for antigen processing, GPGPG linkers will aid to avoid aggregation and sustain flexibility, and His-tag is vital for the recognition and separation and facilitates efficient purification (7, 29, 44–46). The antigen and allergen predictions were used similarly to those mentioned above to identify vaccine candidate combinations with optimal immunological and antigenic properties. The combination with the highest antigen score was analyzed for its physicochemical activity via the ProtParam server (https://web.expasy.org/protparam/) (47), considering default parameters. The selected vaccine combination also underwent solubility analysis via Protein-sol (https://protein-sol.manchester.ac.uk/) (48), which is based on weighted scores considering default parameters.

Population coverage analysis of the selected MHC I and MHC II epitopes

The selection of potential must be validated based on its population coverage, which can be crucial for vaccine development and helpful for most of the world's population (29). The final MHC I and MHC II epitopes with their restricted alleles were utilized for the analysis via population coverage (http://tools.iedb.org/population/) (49), which estimates the fraction of responders to epitopes with known MHC restrictions.

Mutated vaccine formulation and immune and physiological assessments

The variable positions identified through multiple sequence alignment were mapped onto the selected epitope to introduce variability and design a mutated epitope to formulate a mutated vaccine that can be helpful in combating multi-strain. The mutated vaccine was constructed, and its immune and physiological assessments were performed similarly to those of the non-mutated vaccine.

Structure modeling and quality assessment

The SOPMA (based on the homology modeling) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) (50) and PSIPRED [based on machine learning (ML)] (http://bioinf.cs.ucl.ac.uk/psipred/) (51) were employed to examine the secondary structure of non-mutated and mutated vaccine construct following the default parameters. However, structure was modeled via the Robetta (https://robetta.bakerlab.org/) (52) server based on deep learning methods using RoseTTAFold. These models were enhanced via the GalaxyRefine (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) (53) server, and the most promising enhanced models were further examined for their structure quality validation via PROCHCK (https://saves.mbi.ucla.edu/) (54) and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php), which is grounded on the statistical analysis following the available structure (55).

Identification of discontinuous epitopes

Discontinuous epitopes are crucial for encoding the immune system's specificity and complexity in responding to infectious agents, leading to more robust and protective immune responses (36). Therefore, the presence of these epitopes within the non-mutated and mutated vaccine was examined via Ellipro (http://tools.iedb.org/ellipro/) (56), which is grounded on geometrical properties of structure, considering the vaccine model structure.

Docking analysis of vaccine with TLR

Potent vaccines must be able to bind with the receptor to activate an immune activity. Therefore, the formulated vaccine (non-mutated and mutated) was docked with the TLR via the ClusPro (https://cluspro.org/login.php) (57) webserver, which utilized the PIPER docking algorithm following the default parameters, whereas the TLR3 structure was collected via the Protein Data Bank (PDB) (ID: 1ZIW) (https://www.rcsb.org/) database. The obtained docked complexes were examined, and the most potent complexes were selected based on their lowest negative energy, demonstrating strong binding. The binding affinities of complex chosen were computed via the PRODIGY (https://rascar.science.uu.nl/prodigy/) (58) sever, and their interaction was visualized through the PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) (59) and PyMOL.

Molecular dynamics simulation

To examine the docked complex's stability (vaccine with TLR), the Desmond software on an Acer workstation with Ubuntu 20.04 was used (60). The OPLS-2005 Force field was employed to generate the coordinates and topology file of the vaccine and TLR complex to define bonded and non-bonded interactions. The system was prepared, solvated (in the TIP3P model), and further neutralized to mimic the physiological condition via Na+ and Cl– counter ions with 0.15 M salt concentration. Furthermore, the simulations were carried out at 300 K temperature and 1.0325 bar pressure for 100 ns, and the system was minimized and relaxed using the default protocol considering all other criteria that were earlier described (23, 60–63). Furthermore, the trajectory file was examined by root mean square deviation (RMSD) and root mean square fluctuation (RMSF) to evaluate the system's stability.

Vaccine-assisted immune activity via immune simulation

The immune activity produced via vaccine (non-mutated and mutated) was analyzed via C-ImmSim (https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php) (64), which employs an ML algorithm. This server assesses the host's immune activity and the ensuing vaccine administration. Default parameters were used following the adjustment based on previously reported data corresponding to the vaccine construct sequence. Additionally, time steps were modified to reflect the administration of three doses at 1, 84, and 168, with 1,050 set as the simulation step, while all other parameters remained the same (5, 21, 65).

Optimization and cloning of vaccine

The formulated vaccine (non-mutated and mutated) must have a high expression level for a robust response. Therefore, the constructed sequence was optimized via the VectorBuilder (https://en.vectorbuilder.com/tool/codon-optimization.html) server, considering *E. coli K12* with default parameters. The Codon Adaptation Index (CAI) and GC% should be 0.8–1.0 and 30%–70% for the maximum expression, respectively (25, 66). Furthermore, the optimized sequence was incorporated and cloned in pET-28a (+) via SnapGene (https://www.snapgene.com/) software, considering a specific restriction site as previously reported (5, 7, 11).

Results

Collection of the target sequence and their immune assessment

The selected proteins, envelope glycoprotein (Q75008), protease (Q75002), reverse transcriptase (Q75002), and integrase (Q75002), were retrieved from the UniProt database, which is a part of the human immunodeficiency virus type 1 group M subtype C (isolate ETH2220), and are crucial in the infection mechanism (11, 67). The immune assessment of the target sequence demonstrated (Table 1) that the required properties can be utilized for vaccine formulation.

Identification of B-cell epitope and their immune assessment

The crucial B-cell epitope within targets was identified via ABCpred (32) and BepiPred 2.0 (33). Via the ABCpred server, 87 envelope glycoprotein (Supplementary Table 1), 9 protease (Supplementary Table 2), 56 reverse transcriptase (Supplementary Table 3), and 28 integrase (Supplementary Table 4) epitopes, and simultaneously via BepiPred, 28 envelope glycoprotein (Supplementary Table 5), 4 protease (Supplementary Table 6), 20 reverse transcriptase (Supplementary Table 7), and 9 integrase (Supplementary Table 8) epitopes were predicted. Moreover, 25 envelope glycoprotein, 4 protease, 24 reverse transcriptase, and 13 integrase epitopes were selected to screen out the more precise assessments, which overlapped in both (Supplementary Table 9). The immune evaluation of these final epitopes revealed that several epitopes have potential, having antigen, non-allergen, and non-toxic features, and the epitopes with high antigen scores from each target (Supplementary Table 9, highlighted in blue) were selected for vaccine formulation as in Table 2.

Identification of T-cell (MHC I and MHC II) epitopes and their immune assessment

The MHC I and MHC II epitopes were identified within the targets via Tepitool (40), considering the most frequent alleles (29). The MHC I assessment revealed 238 envelope glycoprotein (Supplementary Table 10), 24 protease (Supplementary Table 11), 170 reverse transcriptase (Supplementary Table 12), and 76 integrase (Supplementary Table 13) epitopes. Simultaneously, the MHC II assessment revealed 80 envelope glycoprotein (Supplementary Table 14), 12 protease (Supplementary Table 15), 61 reverse transcriptase (Supplementary Table 16), and 32 integrase (Supplementary Table 17) epitopes. The immune assessments of the epitope in MHC I and MHC II revealed several leading immunodominant properties, as shown in Table 3. Furthermore, one epitope with many covering alleles and a high antigenic score (Table 3) from each respective target was selected for vaccine formulation, as in Table 4.

Epitope mapping of B and T cells (MHC I and MHC II) with the variable amino acid

To compute the variability of amino acids across different variants, the total reviewed sequences concerning each target were retrieved from UniProt, and their MSA was accomplished via Clustal Omega (42). The MSA was visualized via the JalView (43) software, which revealed several variable positions across the variant (Supplementary Figures 1–4). In the case of the B-cell epitope, a total of 38 amino acids from envelope glycoprotein, 29 from protease, 16 from reverse transcriptase, and 9 from integrase were found and mapped (Supplementary Table 18) with the selected final epitope (Table 2), whereas 73 amino acids from envelope glycoprotein, 8 from protease, 21 from reverse transcriptase, and 12 from integrase for the combined MHC I and II were found and successfully mapped (Supplementary Tables 19–22) with the selected epitope (Table 4, non-mutated vaccine formulation). These mapped amino acids were further incorporated (highlighted in red), and the variability was

TABLE 1 List of selected targets with their immune attributes.

Properties	Envelope	Protease	Reverse T	Integrase
Antigen	0.5425 (Yes)	0.4639 (Yes)	0.5039 (Yes)	0.4628 (Yes)
Allergen	No	No	No	No

10.3389/fimmu.2025.1540253

TABLE 2 List of final selected promising LBL epitopes with their immune properties.

Position	Peptide	Antigen	Allergen	Toxic
Envelope glycoprotein				
78–93	PSPQELGLENVTENFN	1.0049 (Yes)	No	No
Protease				
54–69	IKVRQYDQIIIEICGK	0.5430 (Yes)	No	No
Reverse transcriptase				
349–364	LKTGKFAKRGTAHTND	1.1808 (Yes)	No	No
Integrase				
188–203	RGGIGGYSAGERIIDI	0.8048 (Yes)	No	No

introduced in the selected non-mutated B- and T-cell epitope (Tables 2, 4). Furthermore, the mutated epitope (Supplementary Tables 19–22) concerning to non-mutated epitopes were examined for antigen, allergen, and toxicity assessment, similar to those mentioned for non-mutated epitopes, and several potential epitopes were found to have antigenic, non-allergenic, and non-toxic properties (Supplementary Tables 19–22). Among the potential epitopes, the epitopes with high antigenic scores (Supplementary Tables 19–22, highlighted in blue) were further selected for mutated vaccine formulation.

Vaccine formulation and immune and physiological assessments

Among the predicted epitopes, four LBL (Table 3), four MHC I, and four MHC II (Table 4) were selected based on their high immunodominant activity for the non-mutated vaccine formulation. In contrast, four LBL, four MHC I, and four MHC II mutated epitopes concerning the non-mutated vaccine, based on the introduced variability having high antigenic scores, were used for mutated vaccine formulation, as in Table 5. These selected epitopes were joined via EAAAK, AAY, KK, and GPGPG linkers to attain the most immunodominant combination; six distinct non-mutated vaccines were constructed considering the selected epitope and different linkers, adjuvants, and other essential attributes. Moreover, the adjuvant, PADRE, and His-tag were kept as in the N and C terminal end, and the LBL, MHC I, and MHC II were framed in different positions (11, 44) for the vaccine construction, as shown below, and the final constructed sequence was of 276 amino acids.

1. Adjuvant-PADRE-LBL-MHC I-MHC II-His-tag (V1)
2. Adjuvant-PADRE-LBL-MHC II-MHC I-His-tag (V2)
3. Adjuvant-PADRE-MHC(I)-MHC (II)-LBL-His-tag (V3)
4. Adjuvant-PADRE-MHC(II)-MHC (I)-LBL-His-tag (V4)
5. Adjuvant-PADRE-MHC(II)-LBL-MHC (I)-His-tag (V5)
6. Adjuvant-PADRE-MHC(I)-LBL-MHC (II)-His-tag (V6)

Furthermore, antigenicity and allergenicity revealed that the V2 combination was found to have the highest antigenic score among the different combinations, as shown in Supplementary Table 23. Moreover, all the constructed vaccines in different forms have an antigenic nature and a non-allergenic feature, which ensures that the selected epitope is highly promising in various forms. These V2 combinations (Figure 2) were similarly applied for the mutated

TABLE 3 Immune assessment of MHC I and MHC II epitopes of the targets.

Targets	Total epitopes	Antigen	Non-antigen	Allergen	Non-allergen	Toxic	Non-toxic
MHC I							
Envelope glycoprotein	238	128	110	103	135	1	237
Protease	24	16	8	14	10	0	24
Reverse transcriptase	170	96	74	90	80	0	176
Integrase	76	41	35	44	32	1	75
MHC II							
Envelope glycoprotein	80	48	32	37	43	0	80
Protease	12	6	6	10	2	0	12
Reverse transcriptase	61	38	23	25	36	0	61
Integrase	32	22	10	14	18	0	32

10.3389/fimmu.2025.1540253

TABLE 4 Selected highly antigenic MHC I and MHC II epitopes within all targets and their immune properties.

Position	Peptide	Alleles	Antigen	Allergen	Toxic
MHC I					
Envelope glycoprotein					
206–214	SLDPIPIHY	HLA-A*30:02 HLA-A*01:01 HLA-B*15:01 HLA-A*32:01 HLA-B*35:01 HLA-A*26:01 HLA-A*11:01 HLA-A*02:06 HLA-B*53:01 HLA-A*03:01 HLA-A*02:01 HLA-B*58:01 HLA-B*44:02 HLA-B*44:03 HLA-A*23:01 HLA-B*57:01	2.0650 (Yes)	No	No
Protease					
91–99	TQLGRTLNF	HLA-B*15:01 HLA-A*32:01 HLA-A*23:01 HLA-A*24:02 HLA-A*30:02 HLA-A*02:06 HLA-B*08:01 HLA-A*26:01	1.3043 (Yes)	No	No
Reverse transcriptase					
381–389	VIWGKTPKF	HLA-A*32:01 HLA-A*23:01 HLA-A*24:02 HLA-B*15:01 HLA-A*26:01 HLA-A*30:02 HLA-B*58:01 HLA-B*57:01 HLA-B*08:01 HLA-B*53:01 HLA-A*02:06	0.4408 (Yes)	No	No
Integrase					
75–83	VAVHVASGY	HLA-A*30:02 HLA-B*35:01 HLA-A*26:01 HLA-B*15:01 HLA-A*01:01 HLA-B*53:01 HLA-B*58:01	0.5921 (Yes)	No	No
MHC II					
Envelope glycoprotein					
351–365	NKTIEFKPSSGGDLE	HLA-DRB1*07:01 HLA-DRB1*15:01 HLA-DRB3*01:01 HLA-DRB3*02:02 HLA-DRB4*01:01 HLA-DRB5*01:01	1.3159 (Yes)	No	No
Protease					
42–56	WKPKMIGGIGGFIKV	HLA-DRB5*01:01	0.6796 (Yes)	No	No

(Continued)

TABLE 4 Continued

Position	Peptide	Alleles	Antigen	Allergen	Toxic
Reverse transcriptase					
343–357	QEPFKNLKTGKFAKR	HLA-DRB1*07:01 HLA-DRB1*15:01 HLA-DRB3*01:01 HLA-DRB3*02:02 HLA-DRB4*01:01 HLA-DRB5*01:01	0.7494 (Yes)	No	No
Integrase					
253–267	DNSDIKVVPRRKAKI	HLA-DRB1*03:01 HLA-DRB1*15:01 HLA-DRB3*02:02 HLA-DRB4*01:01 HLA-DRB5*01:01	1.2710 (Yes)	No	No

vaccine formulation of 276 amino acids, and their antigenicity and allergenicity were analyzed (Table 6). Furthermore, the physiochemical properties and solubility analysis revealed suitable properties of non-mutated (Supplementary Table 24) and mutated vaccines, as in Table 6.

Population coverage analysis of the selected MHC I and MHC II epitopes

For effectiveness, a potent vaccine must have a wide range of coverage (29). These eight epitopes (four MHC I and four MHC II) were examined together, and according to the restricted alleles, there was 97.41% coverage, which shows the broader coverage (Figure 3) of the employed epitope in the vaccine formulation.

Structure modeling and quality assessment

The secondary assessment revealed that the non-mutated vaccine has a helix, 23.91%; strand, 23.91%; and coil, 52.17% (Supplementary Figure 5), whereas the mutated has a helix, 20.65%; strand, 25.72%; and coil, 53.62% (Figure 4).

The model structure via Robetta (52) servers revealed a confidence score of 0.42 for the non-mutated and 0.41 for the mutated vaccine, which lies within the better-quality range. These models were further refined, and based on their various parameters, model 3 for the non-mutated (Supplementary Figure 6A) (Supplementary Table 25, highlighted in blue) and model 1 for the mutated vaccine (Figure 5A) (Table 7, highlighted in blue) were found suitable.

The structure quality validation via PROCHECK (54) demonstrated that the non-mutated vaccine has 87.3% residue in the most favored region, 8.6% residue in the additional allowed region, 1.8% residue in the generously allowed region, and 2.3% residue in the disallowed region (Supplementary Figure 6B), followed by 88.3% residue in the most favored region, 9.5% residue in the additional allowed region, 0.9% residue in the generously allowed region, and 1.4% residue in the disallowed region as in Figure 5B for the mutated vaccine. The Ramachandran plot shows that both non-mutated (Supplementary Figure 6B) and mutated (Figure 5B) vaccine models have only five and three residues in the disallowed regions and are scattered, suggesting less likely to cause significant structural instability. Moreover, most of the residue lies in the favored region, suggesting the overall reliable backbone geometry of the model (22, 68). Furthermore, the *Z*-score assessment done via ProSA-web (55) revealed that the non-mutated vaccine has a −6 score (Supplementary Figure 6C) and the mutated vaccine has a −5.54 score (Figure 5C); the negative score represents the superior structure model. Based on structural validation, the assessment demonstrated the good quality of the non-mutated and mutated vaccines (22).

Identification of discontinuous epitopes

The non-mutated and mutated vaccine structure was subjected to the Ellipro (56) server to compute the discontinuous epitope within the vaccine. The subjected non-mutated vaccine revealed that seven epitopes covered 139 amino acids; their range score varied from 0.618 to 0.815 (Supplementary Table 26). In contrast, six epitopes were found for the mutated vaccine, covering 147 residues, followed by the score range from 0.588 to 0.967 (Table 8). The discontinuous epitopes with both vaccines show that the construct vaccine will lead to a remarkable immune response (69).

Docking analysis of the non-mutated and mutated vaccine with TLR

The molecular activity of formulated non-mutated and mutated vaccines with the TLR3 was accomplished via ClusPro (7). The TLR3 can recognize double-stranded RNA (dsRNA) and single-stranded RNA (ssRNA) and is also vital in antiviral immune responses. Moreover, its activation stimulates dendritic cell activation mediated by HIV-1, which makes it an ideal target (7, 70). Among the generated multiple docked complexes of subjected TLR3 and vaccine, model 6 for the non-mutated (Supplementary Table 27) and model 7 for the mutated vaccine (Supplementary Table 28) were

10.3389/fimmu.2025.1540253

TABLE 5 Selected mutated B- and T-cell (MHC I and MHC II) epitopes were mapped with non-mutated epitopes, whereas the mutation was highlighted in blue.

Position	Epitope	A.Pos	R.Pos	V.Amino acid	M.Epitope	Antigen	Allergen	Toxic
B cell								
Envelope glycoprotein								
78–93	PSPQELGLENVTENFN	E86	96	G	PSPQELGL**G**NVTENFN	1.4187 (Yes)	No	No
Protease								
54–69	IKVRQYDQIIIEICGK	I63	63	C	IKVRQYDQ**C**IEICGK	1.1378 (Yes)	No	No
Reverse transcriptase								
349–364	LKTGKFAKRGTAHTND	F354	354	Y	LKTGK**Y**AKRGTAHTND	1.1961 (Yes)	No	No
Integrase								
188–203	RGGIGGYSAGERIIDI	R197	197	R	RGGIGGYSA**R**ERIIDI	1.5032 (Yes)	No	No
MHC I								
Envelope glycoprotein								
206–214	SLDPIPIHY	S206	237	N	**N**LDPIPIHY	2.4487 (Yes)	No	No
Protease								
91–99	TQLGRTLNF	L93	93	I	TQ**I**GRTLNF	1.3254(Yes)	No	No
Reverse transcriptase								
381–389	VIWGKTPKF	T386	387	S	VIWGK**S**PKF	0.5451 (Yes)	No	No
Integrase								
75–83	VAVHVASGY	Y83	83	F	VAVHVASG**F**	0.5744 (Yes)	No	No
MHC II								
Envelope glycoprotein								
351–365	NKTIEFKPSSGGDLE	S359	401	K	NKTIEFKPK**S**GGDLE	1.6779 (Yes)	No	No
Protease								
42–56	WKPKMIGGIGGFIKV	M46	46	I	WKPK**I**IGGIGGFIKV	0.5336(Yes)	No	No
Reverse transcriptase								
343–357	QEPFKNLKTGKFAKR	P345	345	E	QE**E**FKNLKTGKFAKR	0.9871 (Yes)	No	No
Integrase								
253–267	DNSDIKVVPRRKAKI	S255	255	N	DN**N**DIKVVPRRKAKI	1.2852 (Yes)	No	No

A.Pos, Absolute position; R.Pos, Relative position; V.Amino acid, Variable amino acid; M.Epitope, Mutation incorporated based on mapped variability data.

found most suitable, having high negative energies of −1,120.2 and −1,275.4 kcal/mol, respectively. The binding affinity of complexes was computed via PRODIGY (58), and the score was obtained at −12.8 kcal/mol (TLR3-Non-mutated) and −24.0 kcal/mol (TLR3-Mutated). These complexes were visualized for their various types of interaction followed by the H bond via PDBSum (59). The TLR3-Non-mutated complex shows 16 H bonds followed by 4 salt bridges and 196 non-bonded contacts as in Supplementary Figure 7. In contrast, the TLR3-Mutated vaccine revealed 40 H bonds followed by 8 salt bridges and 364 non-bonded contacts, as in Figure 6. Moreover, the interface residue is demonstrated in Supplementary Figure 7, Figure 6. The docking analysis revealed that the vaccine is strongly bound via molecular connection with TLR3, and the incorporated variability in the epitopes does not affect the interaction; rather, it improves, followed by a high number of hydrogen bonds.

Molecular dynamics simulation

The docked TLR3 with the non-mutated and mutated vaccines was analyzed via the Desmond software, followed by considering steps of the parameter (23, 61, 62) to examine their stability. The examination shows that the non-mutated and mutated vaccines remained bound with the TLR3 over the simulation period

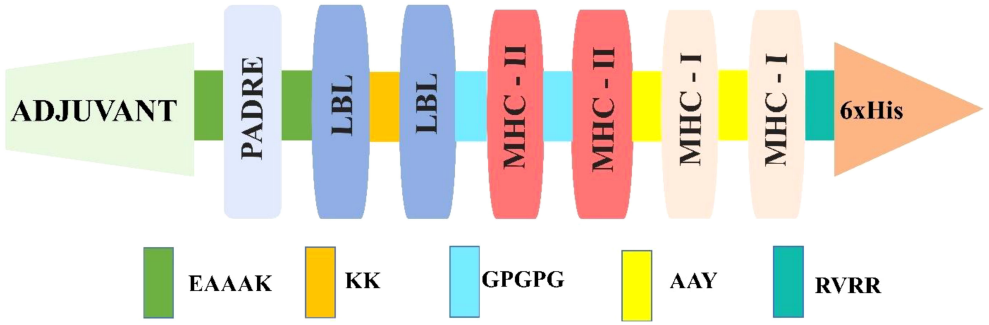


FIGURE 2
Illustration of vaccine construction followed by different attributes.

(Figure 7, Supplementary Figure 8). The RMSD investigation shows that the Cα of the mutated vaccine–TLR3 complex stabilized after 20 ns, followed by approximately 3.0–3.5 Å deviation, and the side chains were comparably slightly higher at approximately 4.5–5.0 Å, which shows the local conformational adjustments (Figure 7A), whereas the non-mutated vaccine–TLR3 complex was gradually stabilized after 20 ns and the Cα atoms rise between 6.0 and 6.5 Å, and the side changes merely followed a similar trend but are slightly higher and stabilized (6.5–7.0 Å) (Supplementary Figure 8A). The higher range of RMSD revealed great flexibility, and the complex maintained its structural stability (71, 72). Moreover, the RMSF investigation shows that the alpha of the mutated vaccine–TLR3 complex was less than 2 Å, and the side chain surpassed 4 Å at specific residues, which shows higher fluctuation (Figure 7B). In contrast, the alpha of the non-mutated vaccine–TLR3 complex remains below 3 Å, and their side chain was comparably higher with a minor exceeding 6–8 Å at certain regions (Supplementary Figure 8B). The minor high peaks in the RMSF of both docked complexes recommend confined rigidity, which is essential for interaction (60, 71, 72).

TABLE 6 Computed antigen, allergen, physicochemical, and solubility properties of the mutated vaccine.

Sl. no.	Properties	Mutated vaccine
1.	Antigen	0.8889 (Probable antigen)
2.	Allergen	Non-allergen
3.	Residue count	276
4.	Molecular weight	30,121.02
5.	Theoretical pI	10.16
6.	Formula	$C_{1348}H_{2166}N_{408}O_{361}S_8$
7.	Estimated half-life	30 h (mammalian reticulocytes, *in vitro*) >20 h (yeast, *in vivo*) >10 h (*Escherichia coli*, *in vivo*)
8.	Instability index	31.76
9.	Aliphatic index	70.43
10.	Grand average of hydropathicity (GRAVY)	−0.593
11.	Solubility	0.674 (Higher than scaled solubility)

Vaccine-assisted immune response activity

The ML accomplished vaccine immune activity and assisted the C-IMMsim server in considering the time steps of the injection interval, as in Figure 8 (Mutated) and Supplementary Figure 9 (Non-mutated). In the case of the non-mutated vaccine, the primary administration shows a high peak of antigen level (700,000 mL) and high generation of immunoglobin, followed by secondary and tertiary administration having an antigen count level of 500,000 each, which further instantly completely reduced, and further, the generated immunoglobins (IgM +IgG, IgM, IgG1+IgG2, IgG1, and IgG2) spiked (650,000) and continued to increase, as shown in Supplementary Figure 9A. In contrast, the mutated vaccine shows antigen counts of approximately 700,000, 300,000, and 50,000 per mL at the primary, secondary, and tertiary response levels, respectively. In contrast, the generated immunoglobin level shows a more promising spike (IgM+IgG, IgM, IgG1+IgG2, IgG1, and IgG2) followed by nearly 800,000, which is higher than the non-mutated immunoglobin level as in Figure 8A. Moreover, the generated cytokine and interleukins show the highest peaks (IFN-γ, IL-2, IL-4, and TNF-α) at nearly 450,000 ng/mL for non-mutated (Supplementary Figure 9B), nearly similar to the mutated vaccine (Figure 8B). The repeated exposure of the immunoglobin and cytokine level followed by steps of injection shows that the vaccine is capable of remarkable immune activity in both forms (Mutated, Figure 8; and Non-mutated, Supplementary Figure 9), and the incorporated variability does not reduce the vaccine's effectiveness.

Optimization and cloning of vaccine

The queried non-mutated and mutated vaccine optimized sequence was 831 for each. The CAI value was 0.95 and GC% was 54.27 for the non-mutated vaccine. In contrast, for the mutated vaccine, the CAI was 0.95, and the GC% was 53.43, demonstrating the significant expression in the bacterial system of both vaccines as the obtained value lies in favor of the expression level. Furthermore,

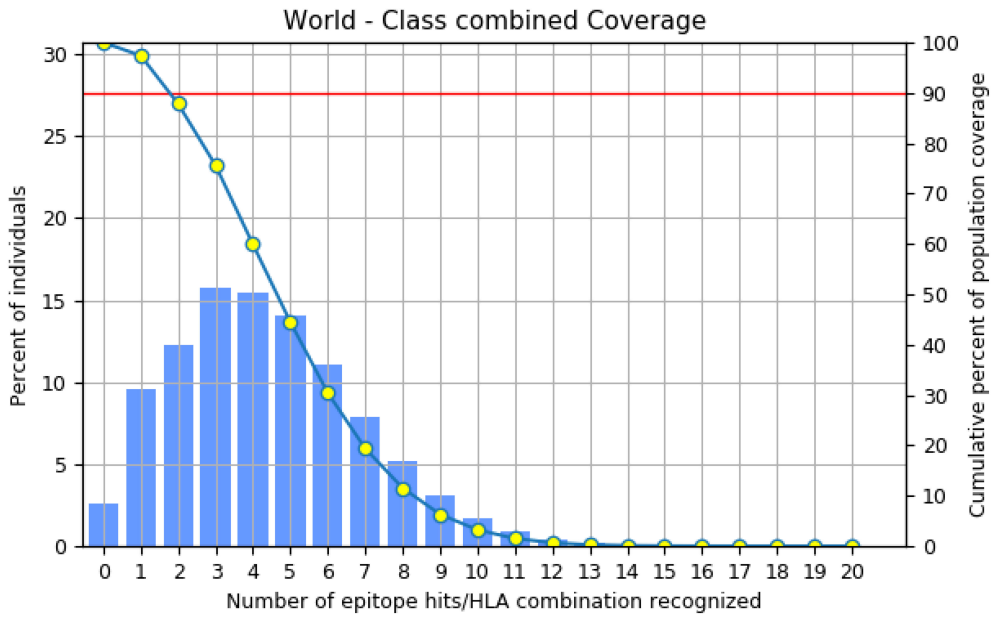


FIGURE 3
Illustration of selected (MHC I and MHC II) epitope-based population coverage.

the optimized mutated and non-mutated vaccines (red) were cloned in the pET28a (+) vector in Figure 9; Supplementary Figure 10.

Discussion

Vaccine formulation for emerging and re-emerging infections presents a promising strategy for effective disease control, offering broad coverage and cost-efficiency. In this context, researchers have leveraged bioinformatics, immunoinformatics, and reverse vaccinology approaches to develop successful multi-epitope vaccines (5, 22, 69, 71). HIV is one of the ongoing endemic concerns due to high infection. No specific vaccine are available to completely eradicate the infection due to its strain variability (26, 28). Previously, researchers applied various approaches for the successful vaccine development towards this infection (5, 11, 21, 70) considering the viral targets mostly from subtype B without incorporating variability in epitopes. In HIV infection, subtype C accounts for the majority of infections, compared to other subtypes, which have not been fully explored yet (13, 14). Therefore, this investigation formulated a potent multi-epitope vaccine by examining subtype C's four potential targets and incorporating variability (mutation) in epitopes to fight against multiple strains of infection. Based on the antigen, allergen, toxicity, and incorporated variability in the epitopes, four LBL, four MHC I, and four MHC II were found as highly immunodominant epitopes and were selected for the non-mutated and mutated (based on the introduced variability) vaccine formulation. The vaccine's immune activity was enhanced by including the adjuvant, PADRE, and 6×His-Tag in the construction (5, 21). The antigenicity and allergenicity assessment confirmed that both the mutated and non-mutated vaccines are antigenic, with scores of 0.8889 and 0.7657, and these values are consistent with previous findings (11, 21, 70) and indicate that both vaccines are non-allergenic and the incorporated mutation in the non-mutated vaccine does not compromise its antigenic effectiveness. Furthermore, mutated and non-mutated vaccines' physiochemical attributes and solubility levels were found suitable and improved (5, 21). The MHC I and MHC II epitopes involved in the vaccine formulation revealed high population coverage, i.e., 97.41%, based on the combined investigation, which is nearly similar to and has improved from the earlier reported study (5, 21, 70). The secondary structural assessment of mutated and non-mutated vaccines showed 20.65% and 23.91% as helix, which is nearly similar to the previous data (5, 11), revealing structural stability. Moreover, the tertiary structure modeling of both mutated and non-mutated vaccines and their validation confirmed that the modeled structures are of favorable quality and closely resemble previously reported data (21, 70). The presence of discontinuous epitopes in vaccines demonstrated their ability to induce protective immunity, as they can produce the antibodies that identify the infection (73, 74). Previously, studies found that the activation of TLR3 can potentially lead to combat HIV infection. Moreover, it can also recognize the dsRNA and ssRNA and initiate the stimulation of dendritic cells facilitated by HIV infection (5, 70). Subsequently, the activation of TLR3 in the viral infection was found to be most suitable, as reported previously by researchers (5, 21, 70, 75). The docking analysis of both non-mutated and mutated vaccine models with TLR3 demonstrated accurate binding, with the incorporated mutation maintaining and enhancing the molecular interaction. This enhancement was reflected in the increased number of interacting residues, with the non-mutated vaccine forming 16 hydrogen bonds with TLR3, while the mutated vaccine formed 40 hydrogen bonds. Furthermore, the binding affinity of both vaccine–TLR3 complexes indicates the favored stability of the

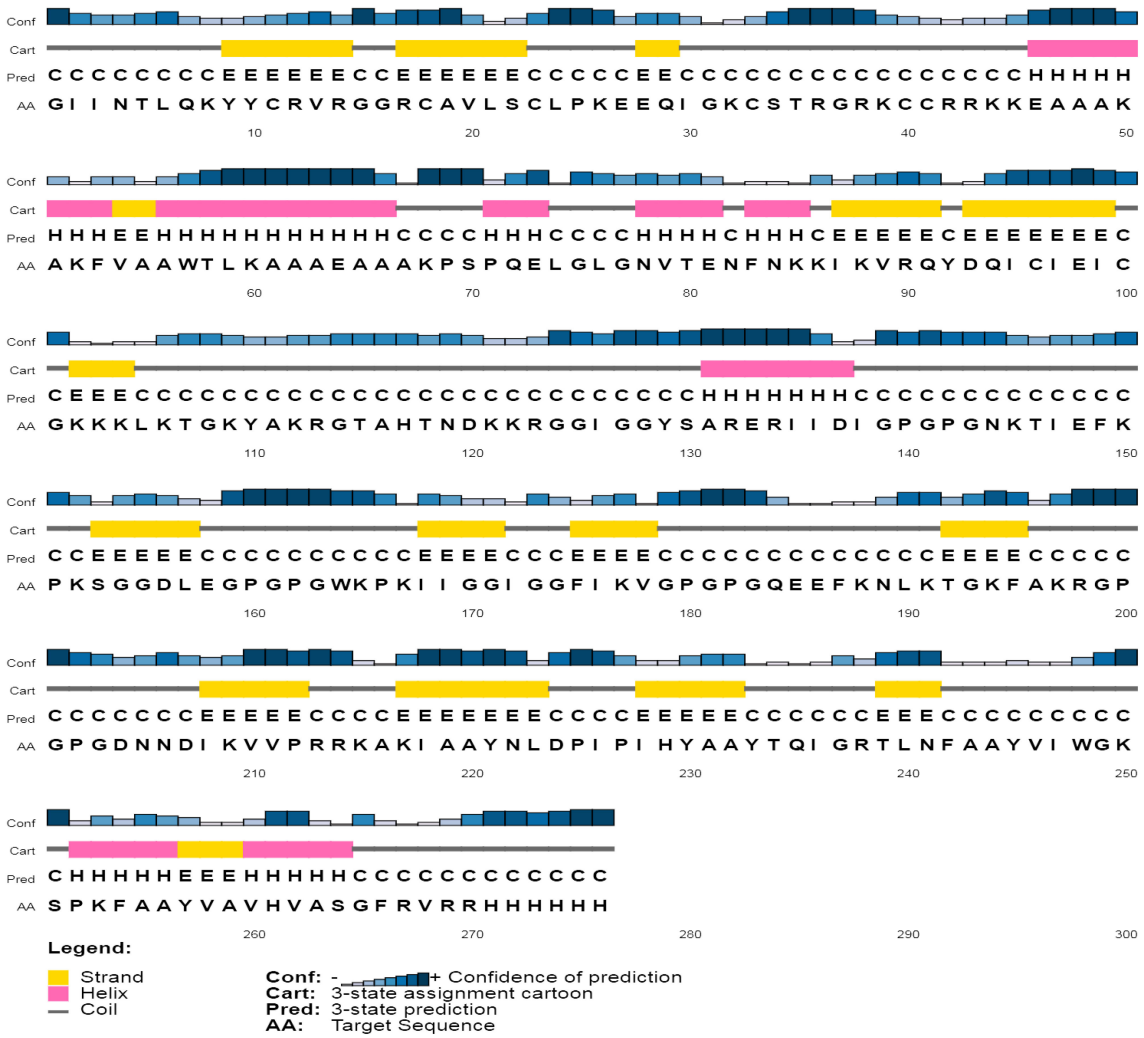


FIGURE 4
Illustration of secondary composition based on their attributes of the mutated vaccine.

system (75). The obtained binding affinity was −12.8 kcal/mol (TLR3-Non-mutated) and −24.0 kcal/mol (TLR3-Mutated). Moreover, nearly similar binding affinities calculated via PRODIGY, i.e., −10.8 kcal/mol (76) and −20.0 kcal/mol (77), were previously reported. Subsequently, a study based on the variability in epitopes reported −20.7 kcal/mol (non-mutated) and −19.5 kcal/mol (mutated) (29). Moreover, Habib et al. found that among the various TLRs (TLR-2, TLR-3, TLR-4, TLR-5, TLR-8, and TLR-9), the designed vaccine exhibited a greater number of interactions towards the TLR-2 followed by 12 H bonds (21). Moreover, exhibited strong interactions specifically with TLR3 and TLR5 among the various TLRs (77). The vaccine-assisted immune simulation activity demonstrated that repeated exposure to formulated vaccines revealed high immunoglobulins and decreased antigen levels. The presence of the IgM in the vaccine will help in the early stage of immune regulation (78), whereas the presence and generation of IgG1 and IgG2 towards the antigens suggest the robust immune response followed by antibody production and neutralization of the viral part (21). Furthermore, the different cytokines and interleukins generated in response to antigens, i.e., IFN-γ (activation of macrophages), IL-2 (stimulates the IFN-γ), IL-4 (B-cell activation), and TNF-α (activation of dendritic cells and T cells), demonstrated the protective immune activity (21, 79, 80) of the formulated non-mutated and mutated vaccine and successfully suppressed and nearly similar to previously reported studies (5, 7, 21), and the introduced mutation does affect and reduce the effectiveness of the production of immune activity. The *in silico* cloning of the non-mutated and mutated vaccine into the pET28a(+) vector within the *E. coli* K12 system demonstrated the maximum expression level, with a CAI value of 0.95 for each and GC% values of 54.27 and 53.43. These values fall within the favored range for optimal expression, aligning closely with previously predicted CAI and GC% values (5, 7, 11, 70). Moreover, the cloning of the designed vaccine into the pET28a(+) vector was deemed suitable for viral infection-based studies owing to its

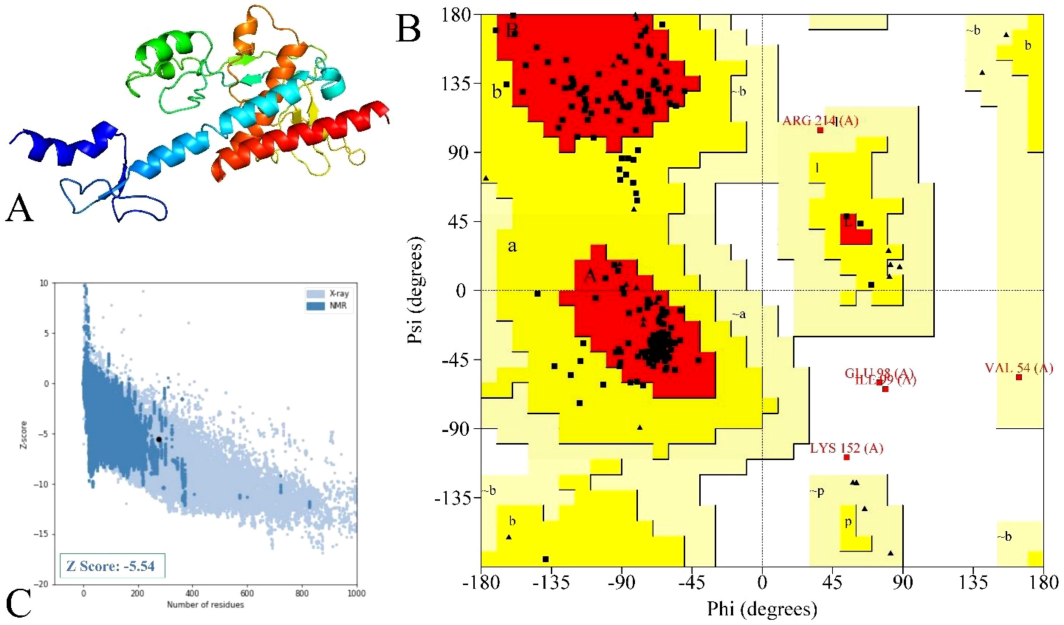


FIGURE 5
Illustration and modeled mutated vaccine and their quality assessments. **(A)** Designed vaccine model, **(B)** residue representation in various regions, and **(C)** quality evaluation via the *Z*-score value.

TABLE 7 List of enhanced mutated vaccine models with their attributes.

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	1.466	2.1	0.0	92.0
Model 1	0.9819	0.322	1.805	8.3	0.5	94.9
Model 2	0.9764	0.329	1.884	9.7	0.5	94.5
Model 3	0.9755	0.335	1.897	10.6	0.5	94.9
Model 4	0.9792	0.329	1.912	9.9	0.5	94.2
Model 5	0.9728	0.349	1.918	10.6	0.9	94.5

TABLE 8 List of obtained discontinuous epitopes within the mutated vaccine.

Sl. No.	Residue	No. of residues	Score
1.	A:F187, A:K188, A:N189, A:L190, A:K191, A:T192, A:G193, A:K194, A:F195	9	0.967
2.	A:G1, A:I2, A:I3, A:N4, A:T5, A:L6, A:Q7, A:K8, A:Y9, A:Y10, A:C11, A:R12, A:V13, A:R14, A:G15, A:G16, A:R17, A:C18, A:A19, A:V20, A:L21, A:S22, A:C23, A:L24, A:P25, A:K26, A:E27, A:E28, A:Q29, A:I30, A:G31, A:K32, A:C33, A:S34, A:T35, A:R36, A:G37, A:R38, A:K39, A:C40, A:C41, A:R42, A:R43	43	0.804
3.	A:P151, A:G154, A:D156, A:L157, A:E158, A:G159, A:P160, A:G161, A:P162, A:G163, A:W164, A:K165, A:P166, A:G179, A:P180, A:G181, A:P182, A:G183, A:Q184, A:E185, A:E186, A:A196, A:K197, A:R198, A:G199, A:P200, A:G201, A:P202, A:G203, A:D204	30	0.689
4.	A:K217, A:A220, A:Y221, A:N222, A:L223, A:D224, A:P225, A:I226, A:P227, A:H229, A:Y230	11	0.648
5.	A:K68, A:P69, A:S70, A:P71, A:Q72, A:E73, A:L74, A:G77, A:N78, A:V79, A:T80, A:E81, A:N82, A:F83, A:K85, A:D137, A:I138, A:G139, A:P140, A:G141, A:P142, A:G143, A:N144, A:K145, A:S264, A:G265, A:R267, A:V268, A:R269, A:R270, A:H271, A:H272, A:H274, A:H275, A:H276	35	0.599
6.	A:C100, A:G101, A:K102, A:K104, A:L105, A:T107, A:G108, A:K109, A:A111, A:K112, A:G114, A:T115, A:A116, A:H117, A:T118, A:N119, A:D120, A:K121, A:K122	19	0.588

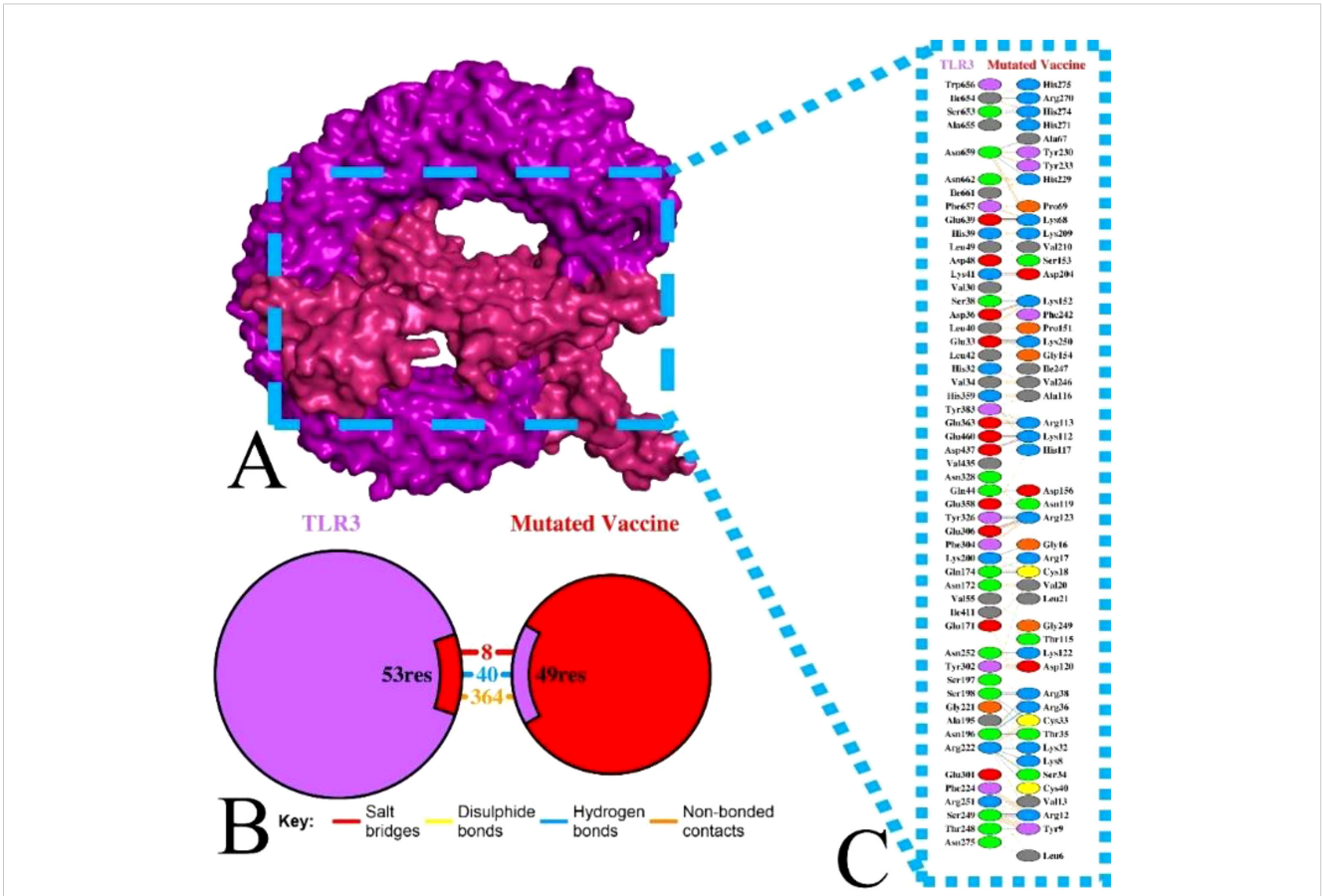


FIGURE 6
Illustration of TLR3 with mutated vaccine. **(A)** Surface interaction, **(B)** interface residue connection, and **(C)** residual interaction.

capability to efficiently express viral proteins and the presence of multiple cloning sites, which streamline the cloning process (5, 29, 81).

One of the major hurdles to combating HIV is the low immune response and strain variability. Compared to conventional methods, immunoinformatics-based approaches offer a more precise, rapid, and cost-effective method for vaccine formulation. This study's major findings demonstrate that the designed vaccine elicits a significant immune response, effectively triggering cellular and humoral activity to combat the infection. Furthermore, based on strain variability, the incorporated mutation does not affect its effectiveness, highlighting its potential to address multi-strain variability. Overall, this study confirms that the formulated vaccines possess immunodominant activity and are capable of effectively fighting HIV infection.

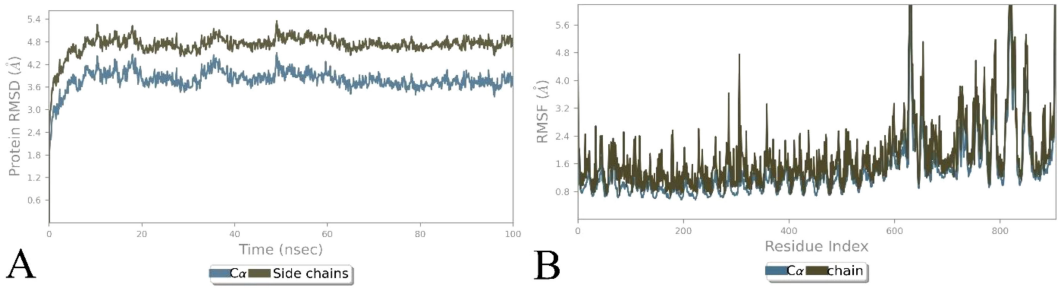


FIGURE 7
Illustration of simulation-based investigation of the docked complex (mutated vaccine with TLR3). **(A)** The RMSD-based trajectories analysis of the complex, and **(B)** the RMSF-based trajectories analysis of the complex.

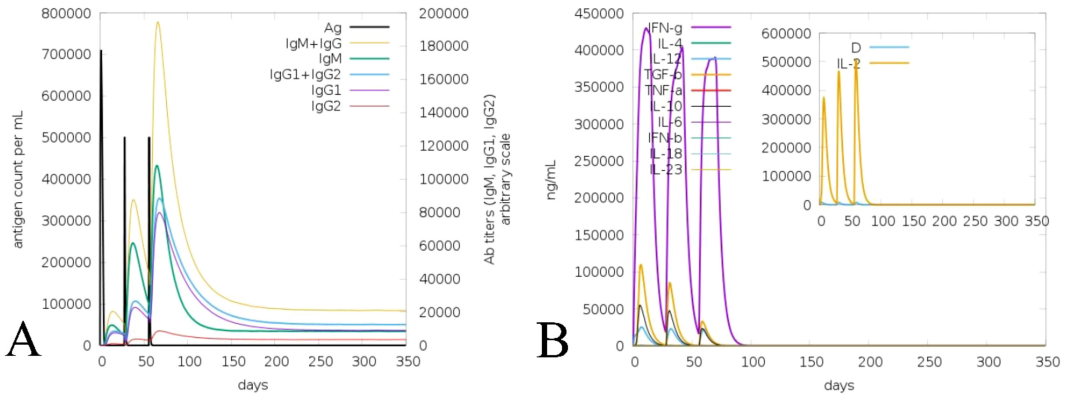


FIGURE 8
Illustration of immune activity response of the mutated vaccine considering injection steps. **(A)** Vaccine-assisted antigen and antibody level. **(B)** Generated cytokine and interleukin level.

Limitations and future scope

Strain variability remains a significant challenge in HIV vaccine development. In this study, we successfully designed both non-mutated and mutated vaccine constructs, incorporating epitope variability to address this issue. The vaccines demonstrated remarkable immune activity, highlighting their potential effectiveness. Several steps of investigation and examination were employed via integrating the computational and immunoinformatic approach, which is associated with accuracy and promise. While the formulated vaccine revealed strong immune activity, future steps, including experimental validation, multi-strain efficacy, immune response evaluation, and clinical trials, are essential to ensure its protection and immune activity.

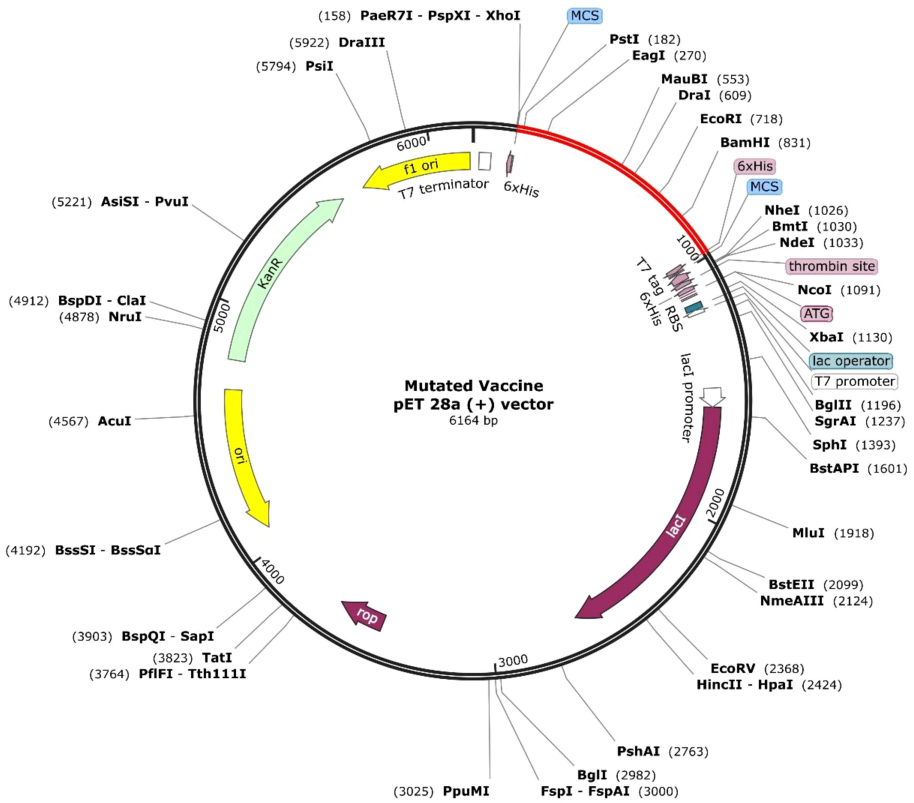


FIGURE 9
Illustration of the incorporated mutated vaccine in pET28a(+).

Conclusion

In this combined mutation-based immunoinformatic investigation, a potent peptide vaccine against HIV infection was successfully formulated by incorporating variability (mutations) in the epitopes utilized in the vaccine. The formulated vaccine effectively evokes a robust immune response based on the fusion of immunodominant epitopes. The docking and dynamics investigation of non-mutated and mutated vaccines with the TLR3 demonstrated strong and stable binding, which ensures the ability of the vaccine activity towards the signaling receptor to trigger the immune response. The vaccine-generated immune response, followed by the injection time step, effectively stimulates immune cells. Additionally, the *in silico*-assisted cloning revealed the high expression levels of non-mutated and mutated vaccines. The strategy employed in this investigation suggests a potent framework for formulating a vaccine capable of addressing strain variability.

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

SK: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. NK: Software, Validation, Visualization, Writing – review & editing. MO: Software, Validation, Visualization, Writing – review & editing. SS: Software, Validation, Visualization, Writing – review & editing. TD: Formal analysis, Software, Validation, Writing – review & editing. MB: Formal analysis, Software, Validation, Writing – review & editing. JJG: Project administration, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is financially supported by the Researchers Supporting Project number (RSP2025R197), King Saud University, Riyadh, Saudi Arabia.

Acknowledgments

Saurav Kumar Mishra acknowledges the Department of Bioinformatics, University of North Bengal, for the Hardware and Software support. The authors extend their appreciation to the Researchers Supporting Project number (RSP2025R197), 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2025.1540253/full#supplementary-material

References

1. Matsuda K, Maeda K. HIV reservoirs and treatment strategies toward curing HIV infection. *Int J Mol Sci.* (2024) 25:2621. doi: 10.3390/ijms25052621

2. Moss AR, Bacchetti P. Natural history of HIV infection. *Aids.* (1989) 3:55–62. doi: 10.1097/00002030-198902000-00001

3. Wu X, Zhou X, Chen Y, Lin Y-F, Li Y, Fu L, et al. Global, regional, and national burdens of HIV/AIDS acquired through sexual transmission 1990–2019: an observational study. *Sexual Health.* (2024) 21. doi: 10.1071/SH24056

4. Martinez-Steele E, Awasana AA, Corrah T, Sabally S, van der Sande M, Jaye A, et al. Is HIV-2-induced AIDS different from HIV-1-associated AIDS? Data from a West African clinic. *Aids.* (2007) 21:317–24. doi: 10.1097/QAD.0b013e328011d7ab

5. Hashempour A, Khodadad N, Akbarinia S, Ghasabi F, Ghasemi Y, Nazar MMKA, et al. Reverse vaccinology approaches to design a potent multiepitope vaccine against the HIV whole genome: immunoinformatic, bioinformatics, and molecular dynamics approaches. *BMC Infect Dis.* (2024) 24:873. doi: 10.1186/s12879-024-09775-2

6. Ghassabi F, Hashempour T, Moghadami M, Davarpanah M, Kalani M, Chatrabnous N, et al. Bacterial etiology and antibiotic resistance pattern of septicemia in HIV and non-HIV patients admitted to tertiary care hospitals, Shiraz, South of Iran. *Cell Mol Biol.* (2017) 63:115–21. doi: 10.14715/cmb/2017.63.9.20

7. Hashempour A, Khodadad N, Bemani P, Ghasemi Y, Akbarinia S, Bordbari R, et al. Design of multivalent-epitope vaccine models directed toward the world's

population against HIV-Gag polyprotein: Reverse vaccinology and immunoinformatics. *PloS One*. (2024) 19:e0306559. doi: 10.1371/journal.pone.0306559

8. Pavlakis GN, Felber BK. A new step towards an HIV/AIDS vaccine. *Lancet*. (2018) 392:192–4. doi: 10.1016/S0140-6736(18)31548-4

9. Johnson MM, Jones CE, Clark DN. The Effect of treatment-associated mutations on HIV replication and transmission cycles. *Viruses*. (2022) 15:107. doi: 10.3390/v15010107

10. Peng S. HIV-1 M group subtype classification using deep learning approach. *Comput Biol Med*. (2024) 183:109218. doi: 10.1016/j.compbiomed.2024.109218

11. Pandey RK, Ojha R, Aathmanathan VS, Krishnan M, Prajapati VK. Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine*. (2018) 36:2262–72. doi: 10.1016/j.vaccine.2018.03.042

12. Williams A, Menon S, Crowe M, Agarwal N, Biccler J, Bbosa N, et al. Geographic and population distributions of human immunodeficiency virus (HIV)–1 and HIV-2 circulating subtypes: A systematic literature review and meta-analysis (2010–2021). *J Infect Dis*. (2023) 228:1583–91. doi: 10.1093/infdis/jiad327

13. Siddappa NB, Dash PK, Mahadevan A, Jayasuryan N, Hu F, Dice B, et al. Identification of subtype C human immunodeficiency virus type 1 by subtype-specific PCR and its use in the characterization of viruses circulating in the southern parts of India. *J Clin Microbiol*. (2004) 42:2742–51. doi: 10.1128/JCM.42.6.2742-2751.2004

14. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis*. (2019) 19:143–55. doi: 10.1016/S1473-3099(18)30647-9

15. Alexiev I, Mavian C, Paisie T, Ciccozzi M, Dimitrova R, Gancheva A, et al. Analysis of the origin and dissemination of HIV-1 subtype C in Bulgaria. *Viruses*. (2022) 14:263. doi: 10.3390/v14020263

16. Nair M, Gettins L, Fuller M, Kirtley S, Hemelaar J. Global and regional genetic diversity of HIV-1 in 2010–21: systematic review and analysis of prevalence. *Lancet Microbe*. (2024) 5:100912. doi: 10.1016/S2666-5247(24)00151-4

17. Masenga SK, Mweene BC, Luwaya E, Muchaili L, Chona M, Kirabo A. HIV–host cell interactions. *Cells*. (2023) 12:1351. doi: 10.3390/cells12101351

18. Balasubramaniam M, Davids B-O, Bryer A, Xu C, Thapa S, Shi J, et al. HIV-1 mutants that escape the cytotoxic T-lymphocytes are defective in viral DNA integration. *PNAS nexus*. (2022) 1:pgac064. doi: 10.1093/pnasnexus/pgac064

19. Jeffy J, Parthasarathy D, Ahmed S, Cervera-Benet H, Xiong U, Harris M, et al. Herschhorn: Alternative substitutions of N332 in HIV-1AD8 gp120 differentially affect envelope glycoprotein function and viral sensitivity to broadly neutralizing antibodies targeting the V3-glycan. *Mbio*. (2024) 15:e02686–23. doi: 10.1128/mbio.02686-23

20. Khairkhah N, Namvar A, Kardani K, Bolhassani A. Prediction of cross-clade HIV-1 T-cell epitopes using immunoinformatics analysis. *Proteins: Structure Function Bioinf*. (2018) 86:1284–93. doi: 10.1002/prot.v86.12

21. Habib A, Liang Y, Xu X, Zhu N, Xie J. Immunoinformatic Identification of Multiple Epitopes of gp120 Protein of HIV-1 to Enhance the Immune Response against HIV-1 Infection. *Int J Mol Sci*. (2024) 25:2432. doi: 10.3390/ijms25042432

22. Bhattacharya M, Alshammari A, Alharbi M, Dhama K, Lee S-S, Chakraborty C. A novel mutation-proof, next-generation vaccine to fight against upcoming SARS-CoV-2 variants and subvariants, designed through AI enabled approaches and tools, along with the machine learning based immune simulation: A vaccine breakthrough. *Int J Biol macromolecules*. (2023) 242:124893. doi: 10.1016/j.ijbiomac.2023.124893

23. Kumar P, Kumar P, Shrivastava A, Dar MA, Lokhande KB, Singh N, et al. Immunoinformatics-based multi-epitope containing fused polypeptide vaccine design against visceral leishmaniasis with high immunogenicity and TLR binding. *Int J Biol Macromolecules*. (2023) 253:127567. doi: 10.1016/j.ijbiomac.2023.127567

24. Moustafa RI, Faraag AH, El-Shenawy R, Agwa MM, Elsayed H. Harnessing immunoinformatics for developing a multiple-epitope peptide-based vaccination approach against SARS-CoV-2 spike protein. *Saudi J Biol Sci*. (2023) 30:103661. doi: 10.1016/j.sjbs.2023.103661

25. Rani NA, Robin TB, Prome AA, Ahmed N, Moin AT, Patil RB, et al. Development of multi epitope subunit vaccines against emerging carp viruses Cyprinid herpesvirus 1 and 3 using immunoinformatics approach. *Sci Rep*. (2024) 14:11783. doi: 10.1038/s41598-024-61074-7

26. Govindan R, Stephenson KE. HIV vaccine development at a crossroads: new B and T cell approaches. *Vaccines*. (2024) 12:1043. doi: 10.3390/vaccines12091043

27. Zubair A, Bibi B, Habib F, Sujan A, Ali M. Clinical trials and recent progress in HIV vaccine development. *Funct Integr Genomics*. (2024) 24:143. doi: 10.1007/s10142-024-01425-9

28. Scott GY, Worku D. HIV vaccination: Navigating the path to a transformative breakthrough—A review of current evidence. *Health Sci Rep*. (2024) 7:e70089. doi: 10.1002/hsr2.70089

29. Mishra SK, Priya P, Rai GP, Haque R, Shanker A. Coevolution based immunoinformatics approach considering variability of epitopes to combat different strains: A case study using spike protein of SARS-CoV-2. *Comput Biol Med*. (2023) 163:107233. doi: 10.1016/j.compbiomed.2023.107233

30. 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

31. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. *J Mol modeling*. (2014) 20:1–6. doi: 10.1007/s00894-014-2278-5

32. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure Function Bioinf*. (2006) 65:40–8. doi: 10.1002/prot.21078

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. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Consortium OSDD, et al. In silico approach for predicting toxicity of peptides and proteins. *PloS One*. (2013) 8:e73957. doi: 10.1371/journal.pone.0073957

35. Peters B, Nielsen M, Sette A. T cell epitope predictions. *Annu Rev Immunol*. (2020) 38:123–45. doi: 10.1146/annurev-immunol-082119-124838

36. 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

37. Voogd L, Ruibal P, Ottenhoff TH, Joosten SA. Antigen presentation by MHC-E: a putative target for vaccination? *Trends Immunol*. (2022) 43:355–65. doi: 10.1016/j.it.2022.03.002

38. Chatzileontiadou DS, Sloane H, Nguyen AT, Gras S, Grant EJ. The many faces of CD4+ T cells: Immunological and structural characteristics. *Int J Mol Sci*. (2020) 22:73. doi: 10.3390/ijms22010073

39. Couture A, Garnier A, Docagne F, Boyer O, Vivien D, Le-Mauff B, et al. HLA-class II artificial antigen presenting cells in CD4+ T cell-based immunotherapy. *Front Immunol*. (2019) 10:1081. doi: 10.3389/fimmu.2019.01081

40. Paul S, Sidney J, Sette A, Peters B. TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol*. (2016) 114:18.19.1–18.19.24. doi: 10.1002/0471142735.2016.114.issue-1

41. Tan C, Zhou J, Wu A, Li C. In silico development of a novel anti-mutation, multi-epitope mRNA vaccine against MPXV variants of emerging lineage and sub-lineages by using immunoinformatics approaches. *J Biomolecular Structure Dynamics*. (2024), 1–18. doi: 10.1080/07391102.2024.2325109

42. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. (2018) 27:135–45. doi: 10.1002/pro.v27.1

43. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. (2009) 25:1189–91. doi: 10.1093/bioinformatics/btp033

44. Singh S, Rao A, Kumar K, Mishra A, Prajapati VK. Translational vaccinomics and structural filtration algorithm to device multiepitope vaccine for catastrophic monkeypox virus. *Comput Biol Med*. (2023) 153:106497. doi: 10.1016/j.compbiomed.2022.106497

45. Khamjan NA, Lohani M, Khan MF, Khan S, Algaissi A. Immunoinformatics strategy to develop a novel universal multiple epitope-based COVID-19 vaccine. *Vaccines*. (2023) 11:1090. doi: 10.3390/vaccines11061090

46. Shetty S, Dash S, Kumar A, Vishwanath S, Kini SG, Brand A. Immunoinformatics design of a multi-epitope vaccine for Chlamydia trachomatis major outer membrane proteins. *Sci Rep*. (2024) 14:1–18. doi: 10.1038/s41598-024-81736-w

47. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*. (2005) 112:571–607. doi: 10.1385/1-59259-584-7:531

48. 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

49. Bui 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

50. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. (1995) 11(6):681–4. doi: 10.1093/bioinformatics/11.6.681

51. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. (2000) 16(4):404–5. doi: 10.1093/bioinformatics/16.4.404

52. 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

53. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res*. (2012) 40:W294–7. doi: 10.1093/nar/gks493

54. 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

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. 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

57. 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

58. 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) 32:3676–8. doi: 10.1093/bioinformatics/btw514

59. 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.v27.1

60. Akhtar N, Joshi A, Singh J, Kaushik V. Design of a novel and potent multivalent epitope based human cytomegalovirus peptide vaccine: an immunoinformatics approach. *J Mol Liquids*. (2021) 335:116586. doi: 10.1016/j.molliq.2021.116586

61. Mishra SS, Kumar N, Karkara BB, Sharma C, Kalra S. Identification of potential inhibitors of Zika virus targeting NS3 helicase using molecular dynamics simulations and DFT studies. *Mol Diversity*. (2023) 27:1689–701. doi: 10.1007/s11030-022-10522-5

62. Hussain M, Kanwal N, Jahangir A, Ali N, Hanif N, Ullah O. Computational modeling of cyclotides as antimicrobial agents against neisseria gonorrhoeae porB porin protein: integration of docking, immune, and molecular dynamics simulations. *Front Chem*. (2024) 12:1493165. doi: 10.3389/fchem.2024.1493165

63. Samad A, Meghla NS, Nain Z, Karpiński TM, Rahman MS. Immune epitopes identification and designing of a multi-epitope vaccine against bovine leukemia virus: a molecular dynamics and immune simulation approaches. *Cancer Immunology Immunotherapy*. (2022) 71:2535–48. doi: 10.1007/s00262-022-03181-w

64. 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

65. Yun J-S, Kim AR, Kim SM, Shin E, Ha S-J, Kim D, et al. In silico analysis for the development of multi-epitope vaccines against Mycobacterium tuberculosis. *Front Immunol*. (2024) 15:1474346. doi: 10.3389/fimmu.2024.1474346

66. Ahmed S, Rahman MN, Hasan M, Hasan MA, Mia MM. Immunogenic multi-epitope-based vaccine development to combat cyclosporiasis of immunocompromised patients applying computational biology method. *Exp Parasitology*. (2023) 248:108497. doi: 10.1016/j.exppara.2023.108497

67. Gartner MJ, Roche M, Churchill MJ, Gorry PR, Flynn JK. Understanding the mechanisms driving the spread of subtype C HIV-1. *EBioMedicine*. (2020) 53:102682. doi: 10.1016/j.ebiom.2020.102682

68. Kleywegt GJ, Jones TA. Phi/psi-chology: ramachandran revisited. *Structure*. (1996) 4:1395–400. doi: 10.1016/S0969-2126(96)00147-5

69. Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee S-S. Immunoinformatics approach for the identification and characterization of T cell and B cell epitopes towards the peptide-based vaccine against SARS-CoV-2. *Arch Med Res*. (2021) 52:362–70. doi: 10.1016/j.arcmed.2021.01.004

70. Abdulla F, Adhikari UK, Uddin MK. Exploring T & B-cell epitopes and designing multi-epitope subunit vaccine targeting integration step of HIV-1 lifecycle using immunoinformatics approach. *Microbial pathogenesis*. (2019) 137:103791. doi: 10.1016/j.micpath.2019.103791

71. Bhattacharya K, Shamkh IM, Khan MS, Lotfy MM, Nzeyimana JB, Abutayeh RF, et al. Multi-epitope vaccine design against monkeypox virus via reverse vaccinology method exploiting immunoinformatic and bioinformatic approaches. *Vaccines*. (2022) 10:2010. doi: 10.3390/vaccines10122010

72. Islam SI, Mou MJ, Sanjida S. Application of reverse vaccinology to design a multi-epitope subunit vaccine against a new strain of Aeromonas veronii. *J Genet Eng Biotechnol*. (2022) 20:118. doi: 10.1186/s43141-022-00391-8

73. Ferdous S, Kelm S, Baker TS, Shi J, Martin AC. B-cell epitopes: Discontinuity and conformational analysis. *Mol Immunol*. (2019) 114:643–50. doi: 10.1016/j.molimm.2019.09.014

74. Ahmad TA, Eweida AE, Sheweita SA. B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials Vaccinology*. (2016) 5:71–83. doi: 10.1016/j.trivac.2016.04.003

75. Saha R, Ghosh P, Burra VP. Designing a next generation multi-epitope based peptide vaccine candidate against SARS-CoV-2 using computational approaches. *3 Biotech*. (2021) 11:47. doi: 10.1007/s13205-020-02574-x

76. Khairunisa SQ, Rachman BE, Fahmi M, Dinana IA, Ito M. Designing a multi-epitope vaccine targeting the HIV-1 subtype CRF01_AE in Indonesia. *Comput Biol Med*. (2025) 187:109758. doi: 10.1016/j.compbiomed.2025.109758

77. Zubair A, Al-Emam A, Ali M, Hussain SM, Elmagzoub RM. Targeting HIV-1 conserved regions: An immunoinformatic pathway to vaccine innovation for the Asia. *PloS One*. (2025) 20(3):e0317382

78. Ruggiero A, Piubelli C, Calciano L, Accordini S, Valenti MT, Dalle Carbonare L, et al. SARS-CoV-2 vaccination elicits unconventional IgM specific responses in naïve and previously COVID-19-infected individuals. *EBioMedicine*. (2022) 77:103888. doi: 10.1016/j.ebiom.2022.103888

79. Gupta A, Righi E, Konnova A, Sciammarella C, Spiteri G, Van Averbeke V, et al. Interleukin-2-mediated CD4 T-cell activation correlates highly with effective serological and T-cell responses to SARS-CoV-2 vaccination in people living with HIV. *J Med Virol*. (2024) 96:e29820. doi: 10.1002/jmv.29820

80. Vaidya SA, Korner C, Sirignano MN, Amero M, Bazner S, Rychert J, et al. Tumor necrosis factor α is associated with viral control and early disease progression in patients with HIV type 1 infection. *J Infect Dis*. (2014) 210:1042–6. doi: 10.1093/infdis/jiu206

81. Sethi G, Kim YK, Han S-C, Hwang JH. Designing a broad-spectrum multi-epitope subunit vaccine against leptospirosis using immunoinformatics and structural approaches. *Front Immunol*. (2025) 15:1503853. doi: 10.3389/fimmu.2024.1503853

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 05 June 2025
DOI 10.3389/fimmu.2025.1589083

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College & Hospital, India

REVIEWED BY
Deeksha Tripathi,
University of Rajasthan, India
Md Sadique Hussain,
Uttaranchal University, India

*CORRESPONDENCE
Leah Kashiri
lpadya@medsch.uz.ac.zw
Wonderful T. Choga
wchoga@bhp.org.bw

RECEIVED 06 March 2025
ACCEPTED 14 May 2025
PUBLISHED 05 June 2025

CITATION
Kashiri L, Choga WT, Musasa T, Nziramasanga P, Gutsire RB, Zijenah LS, Mukarati NL, Gaseitsiwe S, Moyo S and Chin'ombe N (2025) *In silico* multi-epitope-based vaccine design for *Mycobacterium avium* complex species.
Front. Immunol. 16:1589083.
doi: 10.3389/fimmu.2025.1589083

In silico multi-epitope-based vaccine design for *Mycobacterium avium* complex species

Leah Kashiri[1]*, Wonderful T. Choga[2,3,4]*, Tinashe Musasa[1], Pasipanodya Nziramasanga[1], Rutendo B. Gutsire[5], Lynn S. Zijenah[5], Norman L. Mukarati[6], Simani Gaseitsiwe[2,7], Sikhulile Moyo[2,7,8,9] and Nyasha Chin'ombe[1]

[1]Medical Microbiology Unit, Department of Laboratory Diagnostic and Investigative Sciences, Faculty of Medicine and Health Sciences, University of Zimbabwe, Harare, Zimbabwe, [2]Department of Medical Sciences, Botswana Harvard Health Partnership, Gaborone, Botswana, [3]School of Allied Health Sciences, Faculty of Health Sciences, Gaborone, Botswana, [4]Department of Applied Biology and Biochemistry, National University of Science and Technology, Bulawayo, Zimbabwe, [5]Immunology Unit, Department of Laboratory Diagnostic and Investigative Sciences, Faculty of Medicine and Health Sciences, University of Zimbabwe, Harare, Zimbabwe, [6]Department of Clinical Veterinary Sciences, Faculty of Veterinary Sciences, University of Zimbabwe, Harare, Zimbabwe, [7]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, United States, [8]School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa, [9]Division of Medical Virology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa

Introduction: The *Mycobacterium avium* complex (MAC)—comprising *M. colombiense*, *M. avium*, and*M. intracellulare*—is an emerging group of opportunistic pathogens responsible for significant morbidity and mortality, particularly in immunocompromised individuals. Despite this growing burden, no vaccines currently provide cross-species protection. In silico vaccine design offers a rapid, cost-effective strategy to identify immunogenic epitopes and assemble multi-epitope constructs with optimized safety and efficacy. Accordingly, we aimed to develop a candidate multi-epitope vaccine (MEV) targeting conserved antigens across multiple MAC species.

Methods: From a genomic survey of nontuberculous mycobacteria (NTM) in Zimbabwe, we assembled complete genomes for *M. colombiense* (MCOL), *M. avium* (MAV), and *M. intracellulare* (MINT). Using both local and global reference datasets, we screened the conserved immunodominant proteins 85A, 85B, and 85C for high-affinity T-helper lymphocyte (THL) epitopes. Promising epitopes were further evaluated for antigenicity, immunogenicity, physicochemical stability, and population coverage.

Results: Epitope mapping across the nine target proteins yielded 82 THL epitopes predicted to bind 13 MHC class II (DRB*) alleles, ensuring broad coverage within Zimbabwean and pan-African populations. Clustering analyses consolidated 26 unique epitopes into 11 consensus peptides, 65.4% of which derived from the 85B proteins. *In silico* immune simulations predicted robust humoral and cellular responses, including elevated IgG titers, T-helper and T-cytotoxic cell proliferation and increased secretion of IFN-γ and IL-2 following MEV administration.

Conclusion: These findings indicate that our construct possesses strong immunogenic potential and cross-species applicability. We present here a rationally designed MEV candidate that merits further experimental validation as a broad-spectrum vaccine against multiple MAC species.

KEYWORDS

Epitopes, Mycobacterium avium complex, Vaccine, Antigen85, mycolyltransferase, Th1 helper T-cell, immunodominance, promiscuous epitopes

1 Introduction

Mycobacterium avium complex (MAC) encompasses a group of twelve species of mycobacteria that are opportunistic pathogens responsible for significant morbidity and mortality in both humans and animals (1, 2). To improve on clarity, may we request change of this statement to 'MAC species cause pulmonary disease in humans and animals. They are clinically significant both in immunocompromised patients—such as people living with HIV/AIDS or those with chronic lung disease—and, less commonly, in otherwise healthy individuals. (3). In animals, MAC infections pose significant threats to livestock and wildlife resulting in economic losses and serving as potential reservoirs for zoonotic transmission (4, 5). Infections due to MAC species are increasing globally, particularly among immunocompromised individuals and patients with underlying lung disease, with notable prevalence reported in Australia, America, Europe, and Asia, thereby underscoring the urgent need for effective vaccine strategies worldwide (1, 3, 6–9). The growing global burden of MAC infections and multidrug resistance among MAC calls for an urgent and constant '*One Health Approach*' in development of effective prevention and control strategies in both humans and animals (2). This vaccine development effort aligns with the '*One Health approach*' addressing human, animal and environmental health by targeting pathogens at the human-animal-environment interface. Despite the widespread use of Bacillus Calmette–Guérin (BCG) as a vaccine for tuberculosis and also providing partial immunity to NTM infections, further studies to either increase the efficacy of the BCG/recombinant BCG vaccine or to create new vaccines or booster vaccines that induce an optimal immune response against NTM is required (10). To address this waning efficacy, booster vaccines are essential as they help achieve long-term immunity (11). However, the recent advances in vaccine technology, *in silico* predictions provide a more efficient, cost-effective alternative for screening candidate epitopes that can elicit strong immune responses, identifying and optimizing vaccine candidates used in the development of therapeutics and vaccines.

Given that effective vaccines must enhance immune mechanisms responsible for pathogen elimination, understanding the nature of pathogen clearance becomes essential. Pathogen clearance often relies on multi-specific, polyclonal, and robust T cell-mediated responses. Major histocompatibility complexes (MHCs), known as human leukocyte antigens (HLAs) in humans, are crucial for the host immune system, presenting antigenic peptides (epitopes) to CD8+ cytotoxic T cells (CTLs) and CD4+ T helper (Th) cells [Helper T-lymphocytes (HTL)] (12–14). HLA class I molecules present endogenous peptides to CTLs, while HLA class II molecules present exogenous peptides to HTLs (13). The HTLs have a key role in adaptive immunity. These activate the B-cells along with the CTLs for production of antibodies and eventually killing infected/damaged cells (15). The HTL epitopes for the selected protein can be calculated using the prediction tool for MHC-II epitope (http://tools.iedb.org/main/tcell/).

Epitope-based vaccines represent a novel approach for generating a specific immune response and avoiding responses against other unfavourable epitopes (like epitopes that may drive immunopathogenic or immune modulating responses) in the complete antigen. Potential advantages of epitope-based vaccines also include increased safety, the opportunity to rationally engineer the epitopes for increased potency and breadth, and the ability to focus immune responses on conserved epitopes (16, 17). The repertoire of peptides presented by HLAs is influenced by the structural features of the HLA binding groove and the peptide's amino acid composition (18, 19). *In silico* tools can predict MHC-presented epitopes and profile immune escape mutations, though such analyses remain complex and underexplored for bacterial genomes. Additionally, pathogens frequently mutate within immunogenic epitopes to evade recognition by T cells and agents, posing significant challenges for developing potent vaccines and therapeutics for diseases like tuberculosis (TB) and others (20).

The antigen 85 (Ag85) complex, comprising a cascade of 85A, 85B, and 85C proteins is the main secretory antigen playing an important role in the pathogenicity of mycobacteria (21). Ag85 complex molecules are widely being explored as tools in diagnostic methods and in vaccine research including recombinant attenuated vaccines, DNA vaccines and subunit vaccines because of their ability to allow bacteria to evade host immune responses through preventing formation of phagolysosomes (21). These highly conserved fibronectin-binding proteins also promote immune responses in host by inducing the production of IFN-g and have

been shown to confer protection against TB (22). Research into Ag85 proteins continues to hold great promise for improving TB vaccines, particularly in high-burden settings.

By leveraging computational approaches, it is possible to predict T-cell from MAC antigens, assess their immunogenicity, and design multi-epitope vaccine constructs. In this study, we utilized immunoinformatics to analyze the 85A, 85B, and 85C proteins of MAV, MINT and MCOL, the commonly found strains in Zimbabwe, to identify candidate epitopes for vaccine development. This work aims to contribute to the development of safe and effective vaccines to combat MAC infections in both humans and animals, addressing a critical need in the global fight against mycobacterial diseases.

2 Materials and methods

2.1 Screening for Ag85A-C genes of the three MAC species

From a genomic survey project using NTM samples in Zimbabwe, we generated complete MAV, MINT and MCOL genomes. These genomes provided a foundational dataset for vaccine modelling, enabling a genotyping approach through whole genome sequencing (WGS) to identify immunologically relevant targets. To enhance the multi-epitope vaccine design, we screened for Ag85A-C genes across the three MAC species. The amino acid sequences of the target proteins Ag85A, Ag85B, and Ag85C of MAV, MINT and MCOL were retrieved from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov) in FASTA format. Subsequently, the nine protein sequences were grouped into three multiple sequence alignments (MSAs), corresponding to Ag85A, Ag85B, and Ag85C for MAV, MINT, and MCOL. Each MSA was visualized using AliView (https://github.com/AliView/AliView), and conserved regions were identified from the alignments. The Epitope Conservancy Analysis tool, available through the Immune Epitope Database (IEDB; http://tools.iedb.org/conservancy/), was used to evaluate the variability of epitopes based on the sequence alignment of the three MAC species.

2.2 Prediction of T-cell and designing of the multi-epitope subunit vaccine

We used NetMHCIIpan 4.3 to predict MHC class II-binding peptides (NetMHCIIpan 4.3 - DTU Health Tech - Bioinformatic Services) for all the 3 MSA containing proteins (Ag85A, Ag85B, and Ag85C) for multiple MAC species. Epitope selection thresholds were based on established immunoinformatics criteria. Only epitopes that fulfilled multiple criteria, high antigenicity, strong MHC binding, and IFN-γ induction were shortlisted for vaccine construct design. These selection thresholds have been widely adopted in previous epitope-based vaccine design studies to ensure that the predicted peptides are likely to be immunogenic and broadly recognized across different HLA types (16, 17, 23). As a selection criteria, the strong binding promiscuous epitopes were considered for downstream analyses towards final vaccine construct. To enhance vaccine efficacy, the GPGPG linker was used to connect amino acid sequences, ensuring optimal individual functionality (23, 24). GPGPG linkers reduced junctional immunogenicity. Since immune adjuvants are a key requirement in vaccine formulation and play a critical role in enhancing the efficacy of vaccines, the *Mycobacterium tuberculosis* 50S ribosomal protein L7/L12 (RL7_MYCTU), P9WHE3 was retrieved from the UniProt database (https://www.uniprot.org/) and used as an adjuvant for the immune interaction based on its ability to act as an agonist for TLR437 (25). P9WHE3 was then integrated at the N-terminal of the construct and connected to the antigenic epitopes using the EAAAK linker to ensure improved expression, bioactivity stability and structural integrity (26, 27). The final vaccine construct was meticulously designed by assembling the adjuvant, epitopes, and linkers into a unified, functional structure to maintain the structural stability and immunological independence of the epitopes and adjuvant. To analyze epitope similarity with human surface proteins and minimize the risk of autoimmune reactions, BLASTp was also used. Epitopes with a similarity below 70% to human proteins are considered acceptable (28). The analysis was conducted using the BLASTp tool.

2.3 Determination of physicochemical characteristics, immunogenicity and allergenicity prediction

We evaluated the immunogenicity of the multiepitope subunits using the VaxiJen (VaxiJen v3.0) and the ANTIGENpro module of the SCRATCH protein predictor (Scratch Protein Predictor). Allergenicity was assessed with the AllerTOP v. 2.0 (29) and AlgPred servers (http://crdd.osdd.net/raghava/algpred/) to identify potential allergic reactions, ensuring the safety and efficacy of the predicted vaccine candidates. We used the ProtParam tool of the EXPASY database server (http://web.expasy.org/protparam/) to determine the physicochemical parameters (molecular weight, half-life, atomic composition, stability index and mean hydrophilicity) of the vaccine candidates' antigens.

2.4 3D modelling of immunogenic polypeptides and protein subunits

For *ab initio* modelling, we utilized the Swiss-model (SWISS-MODEL), submitting the designed full length chimeric peptide sequence with default settings. One model (Swiss-Model ID: Q63Q02.1) encompassed the entire multi-epitope construct, while the second returned only a truncated peptide fragment and was therefore excluded from further consideration. Model Q63Q02.1 was subsequently validated using MolProbity metrics—MolProbity score, clashscore, and Ramachandran analysis (Table 1)—to ensure stereochemical quality before downstream analyses. Functional

TABLE 1 Structural validation metrics for the chimeric MEV construct (Model Q63Q02.1).

Model ID	MolProbity Score	Clash score	Ramachandran Favored (%)	Ramachandran Outliers (%)
Q63Q02.1	1.73	3.64	88.35	1.94

insights into the targets were derived by rethreading the models through the BioLiP protein function database (BioLiP). The resulting 3D models were visually inspected using PyMOL (PyMOL | pymol.org) for structural validation and analysis. To further evaluate the protein's flexibility and dynamics, Normal Mode Analysis (NMA) was performed using the iMODS server (https://imods.iqf.csic.es/), which provided insights into residue coupling through the covariance matrix and defined the elastic network model to identify regions of rigidity and flexibility based on the stiffness of atomic interactions.

2.5 Population coverage by HTL epitopes

Human leukocyte antigen (HLA) patterns differ across ethnic groups and geographical regions, making it essential to evaluate population coverage when designing effective vaccines. The IEDB population coverage tool (http://tools.iedb.org/population/) was used to calculate global human population coverage for the predicted HTL epitopes, ensuring their broad applicability. The 15-mer peptides overlapping by 14 amino acids were tested for binding to a set of 13 HLA class II alleles—HLA-DRB1*0101, DRB1*0301, DRB1*0302, DRB1*0401, DRB1*0701, DRB1*0802, DRB1*1101, DRB1*1102, DRB1*1301, DRB1*1302, DRB1*1501, and DRB5*0101—that have high population coverage in Zimbabwe and other African populations (http://www.allelefrequencies.net/hla6006a.asp?hla_population=2057).

3 Results

The amino acid sequences of nine proteins, 85A, 85B and 85C proteins of MAV, MINT and MCOL, were retrieved from the GenBank database to design a multi-epitope vaccine targeting MAC. The inclusion of Ag85 complex was guided by its high degree of conservation among mycobacterial species and its established immunogenicity as supported by its wide use in vaccine candidate development for TB (26, 30–32). Although nine proteins were selected, clustering and consensus analysis resulted in 11 distinct peptide sequences, reflecting inter-strain variability and epitope overlap. The eleven protein sequences were MSFIEKVRKLRGAAATMPR, MSFFEKLRGAAATMPRR, PRRLAIAAVGASLLSGVAVAAGGS, PRRLAIAAMGASLLSGL, RLAIAAVGASLLSGL, GLPVEYLEVPSPSMGRNI, SEKVRAWGRRLLVGAAAAVTLPGLIGIAGGAATAN, SEKVRAWGRRLLVGTAAAATLPG, AWGRRLVVG AAAAATLPGLIGLAGGAATAN, PGLPVEYLQVPSAGMGRNI and PVEYLQVPSAGMGRDIKVQFQS. SignalP 4.5 was used to assess functionality, revealing no signal peptides for proteins other than Ag85A/B/C. Functional protein sequences were then subjected to T-cell epitope prediction, identifying 17 high-affinity HTL epitopes for inclusion in the final vaccine construct. These epitopes overlapped with HTL epitopes (Table 2). A BLAST search against the UniProt database confirmed high conservation among the proteins, ranging from 27.7% to 100%.

3.1 Construction of multi-epitope subunit vaccine.

A total of 7 clusters of overlapping 14-mers high-binding HTL epitopes were predicted (Figure 1A). Consequently, 7 consensus sequences representing each cluster were generated at 100% threshold. These were used in designing the chimera using GPGPG linkers. Additionally, an adjuvant was added to the amino terminus of the vaccine peptide using an EAAAK linker in order to potentiate antigen-specific immune responses. The 50S ribosomal protein L7/L12 (RL7_MYCTU, UniProt ID: P9WHE3) was incorporated as an adjuvant at the N-terminal of the vaccine construct to enhance immunogenicity. The adjuvant RL7_MYCTU was an ideal adjuvant for enhancing cell-mediated immunity in this MEV vaccine construct as it is well-documented to have the ability to act as a potent immunostimulatory molecule inducing cytokine production, T cell activation and IFN-γ secretion (25). The final vaccine peptide generated consisted of 423 amino acid residues. The immunogenic peptide identified through epitope-mapping has been patented for further vaccine development.

3.2 Physiochemical properties and solubility prediction

The molecular weight (MW), theoretical isoelectric point (pI), and half-life of the final protein [as assessed in mammalian reticulocytes (*in vitro*) and in yeast and *E. coli* (*in vivo*)] is summarized in Table 2. The protein demonstrated good solubility upon expression, with a solubility score and an Abs 0.1% (1 g/L) value of 0.505. Furthermore, the instability index (II) was calculated as 36.67, classifying the protein as stable, as proteins with an II >40 are typically considered unstable.

3.3 Secondary-structure analysis and tertiary-structure modeling of the chimeric MEV construct using Swiss-model server

The Swiss-Model server was used to generate the two tertiary structure models for the designed chimeric protein. Among these, the **MSVQ63Q02.1** model was identified as the best as it represented the full-length MEV construct and achieved a

TABLE 2 Characteristics of the primary structure of proposed multi-epitope vaccine candidate for MAC species calculated through ProtParam tool.

Characteristics of Vaccine	Assessment
Number of amino acids	423 aa
Molecular weight (kDa)	41.5 kDa
Theoretical pI	10.27
Negatively charged residues (Asp + Glu)	31
Positively charged residues (Arg + Lys)	50
Extinction coefficient ($M^{-1}cm^{-1}$)	20910
Estimated half life	30 hours (mammalian reticulocytes)
Aliphatic index	88.09
Grand average of hydropathicity (GRAVY)	0.099
Instability index	36.67 (Stable)
Rama favoured score of tertiary structure	88.35%
z-score	4.41
Codon Optimization Index (CAI)	0.80
GC content (E coli as vector)	64.38%
Allergenicity (ALLERCATPRO + AlgPred)	No evidence (non-allergenic)
Antigenicity Score: Threshold (0.4)	0.6120 (Probable ANTIGEN).

favorable MolProbity score (1.73) and low clashscore (3.64) and thus was selected for presentation (Table 1, Figure 2A). Ramachandran plot analysis indicated that 88.35% of the residues were in favoured regions, with only 1.94% in disallowed regions (Table 1, Figure 2Bb). The quality and accuracy of the refined 3D model were evaluated using ProSA-web and ERRAT. The ERRAT analysis reported an overall quality factor of 97.5% (Figure 2C), while ProSA-web yielded a Z-score of -4.41 (Figure 2D), confirming the reliability of the refined vaccine protein model. The internal dynamics of the MEV model were further examined using normal mode analysis (NMA) via the iMODS server. The covariance matrix (Figure 2E) illustrated patterns of correlated and anti-correlated motions between residues, while the elastic network model (Figure 2F) highlighted stiffness variations across residue connections, indicating regions of structural rigidity and flexibility.

3.4 IFN-γ inducing epitope prediction

This prediction was consistent with the simulated level of IFN-γ produced after immunization with the peptide using the C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.php). The MEV model we designed managed to elicit a significant increase in T cell population following immunization. Furthermore, the antibody levels (IgM+IgG, IgG1+IgG2, IgM, and IgG) were found to increase during immunizations, accompanied by a decrease in antigen count (Figure 3A). Additionally, both CTL and HTL populations increased following secondary and tertiary immunization (Figures 3B–D).

3.5 Codon optimization and the *in vitro* expression simulations

The Codon Optimization Tool (ExpOptimizer) tool (https://www.novoprolabs.com/tools/codon-optimization), was used for the multi-epitope vaccine model to enhance efficient translation and optimization of the codons for maximal expression in the prokaryotic host system *E. coli* (strain K12). In codon-optimized sequences of the designed vaccine *(MEV-LpKwTC001)*, the codon adaptation index values were 0.80, and the GC content value 64.14%. Additionally, the adapted codon sequences were optimized with sticky end restriction sites of *HindIII* and *NcoI* at the N-terminus and C-terminus to facilitate restriction and cloning and inserted into the recombinant plasmid vector, *pET-30a (+)*, using the Snapgene tool to design and effective cloning strategy (Figure 4).

4 Discussion

Paediatric administration of BCG vaccine is practiced in Zimbabwe. However, with *M. tuberculosis* being endemic and with the rise of NTM infections, there is a great need for new vaccines and booster vaccines for the BCG vaccines to fight both tuberculosis in adults as well as NTM infections particularly MAC infections. The rise of drug-resistant mycobacteria, limited BCG efficacy, and the need for vaccines targeting both humans and animals highlight significant challenges (33–35). The focus has recently shifted towards the development of subunit vaccines as they are associated with better safety profiles and are logistically more feasible, effective vaccines candidates that can be used to control MAC-related infections (36). Bioinformatics (*in silico*) is a good option to be used in designing and development of vaccines and diagnostics for newly emerged pathogens. The use of this approach reduces the time and cost. In order to construct a potent vaccine and effective diagnosis, understanding of the epitope and antibody interaction is required.

This work therefore focused on the *in-silico* design and development of a multi-epitope vaccine peptide generated using different MAC species (*MCOL, MAV, MINT*) and antigens (85A, 85B, and 85C) and has a potential for cross-protection (prophylactic and therapeutic). The proteins that we selected had exhibited potential to be vaccine candidates for *in vitro* studies (32). Epitope mapping of Ag85 protein complex has identified distinct peptides capable of stimulating human T cells, highlighting specific regions that could potentially trigger protective immune responses (37). More than 50% of vaccine candidates development for TB to date, some in advanced clinical trials, incorporated Ag85 (32).

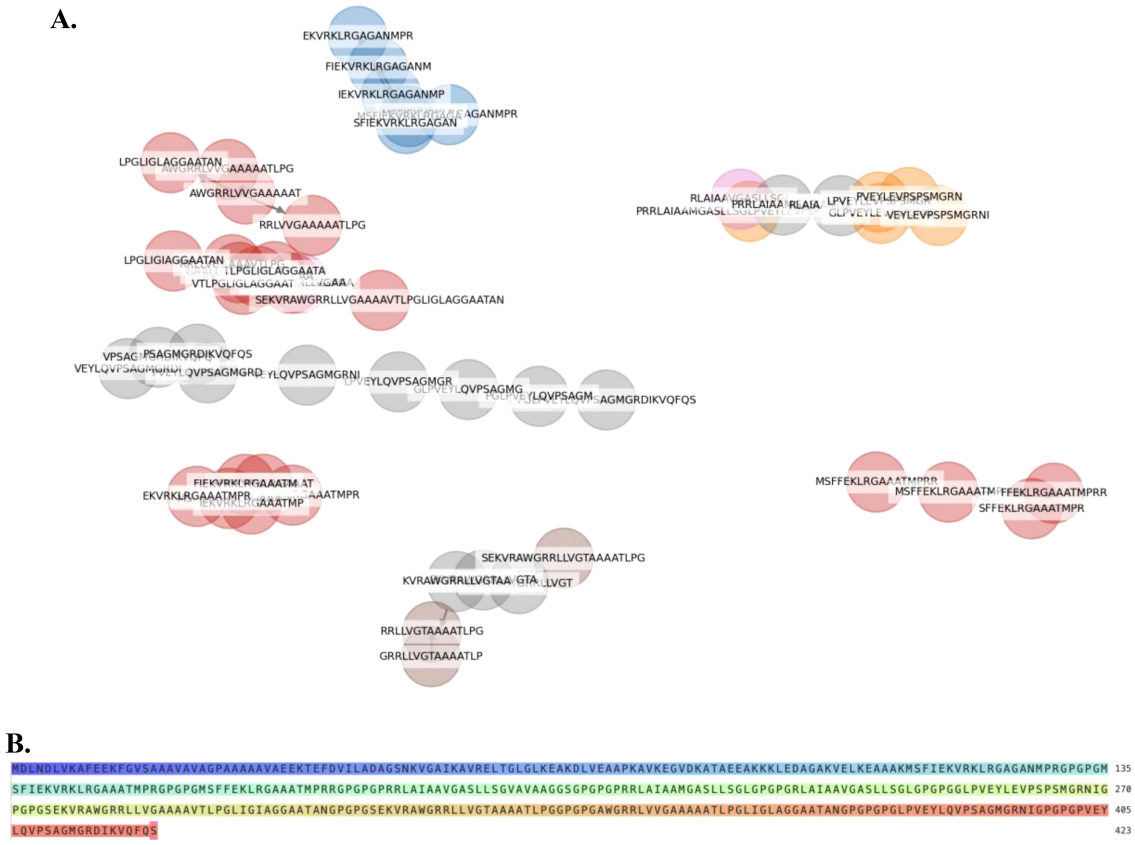


FIGURE 1
Helper T lymphocyte (HTL) Epitope Clustering and Multi-Epitope Vaccine Construct Design **(A)**. Predicted HTL 14-mer epitopes were grouped into clusters based on sequence similarity of their core binding regions. **(B)** A multi-epitope-vaccine construct was designed using consensus sequences derived from each epitope cluster.

Conservation of the Ag85 complex across mycobacterial species also highlights their potential for cross-protection against related pathogens, including those in the MAC (30, 31).

We identified THL epitopes from selected proteins and fused them using linkers to create a multi-epitope peptide. Immunoinformatics showed the vaccine candidate contains many high-affinity MHC Class II epitopes from ag85B. Notably, multi-epitope vaccines are often poorly immunogenic and typically require adjuvants (38, 39); however, the designed protein demonstrated comparable antigenicity scores with or without an added adjuvant sequence. The antigenicity of the final sequence (including the adjuvant sequence) was shown to be probable antigen with a bacteria and also to be non-allergenic. we identified epitopes with accessibility, flexibility, hydrophilicity, and antigenic profiles for the Ag85. The vaccine candidate has a molecular weight of 41.5 kDa and is predicted to be soluble upon expression, aligning with its simulated immunogenicity. Solubility in an *E. coli* host is crucial for biochemical and functional studies.

Based on the predicted GRAVY score, which assesses maintenance ability in hydrophilic or hydrophobic environments, our MEV model displayed negative GRAVY value, suggesting a higher structural stability in a hydrophilic environment. This aspect can be correlated with solubility, critical in determining *in vitro* protein expression. Consequently, *MEV-LpKwTC001*, which demonstrated the highest solubility score, was selected for expression in *E. coli*. Furthermore, the predicting protein solubility is crucial for the selection of highly effective candidate proteins, as it can help avoid protein aggregation, which adversely affects biological activity and can lead to failures in the recombinant protein pipeline. The theoretical pI of 10.27 indicates the protein is alkaline, and the predicted instability index confirms its stability upon expression. The aliphatic index highlights the presence of hydrophobic aliphatic side chains, suggesting thermal stability, which is ideal for use in endemic regions like sub-Saharan Africa. These properties show that this is a potential vaccine design. To the best of our knowledge, no vaccine candidate is in phase III, nor licensed for use in the NTM infection with MAC species. To improve the immunogenicity of the vaccine antigen, we inserted adjuvant and linker sequences between the previously predicted epitopes to increase antigenicity. The fact that it has no allergenic properties further confirms its potential as a vaccine candidate.

Secondary structure analysis shows the protein is primarily composed of coils (67%), with 48% of residues disordered. These structural features, including natively unfolded regions and alpha-helical coiled coils, are known to serve as "*structural antigens*" capable of folding into native structures and being recognized by

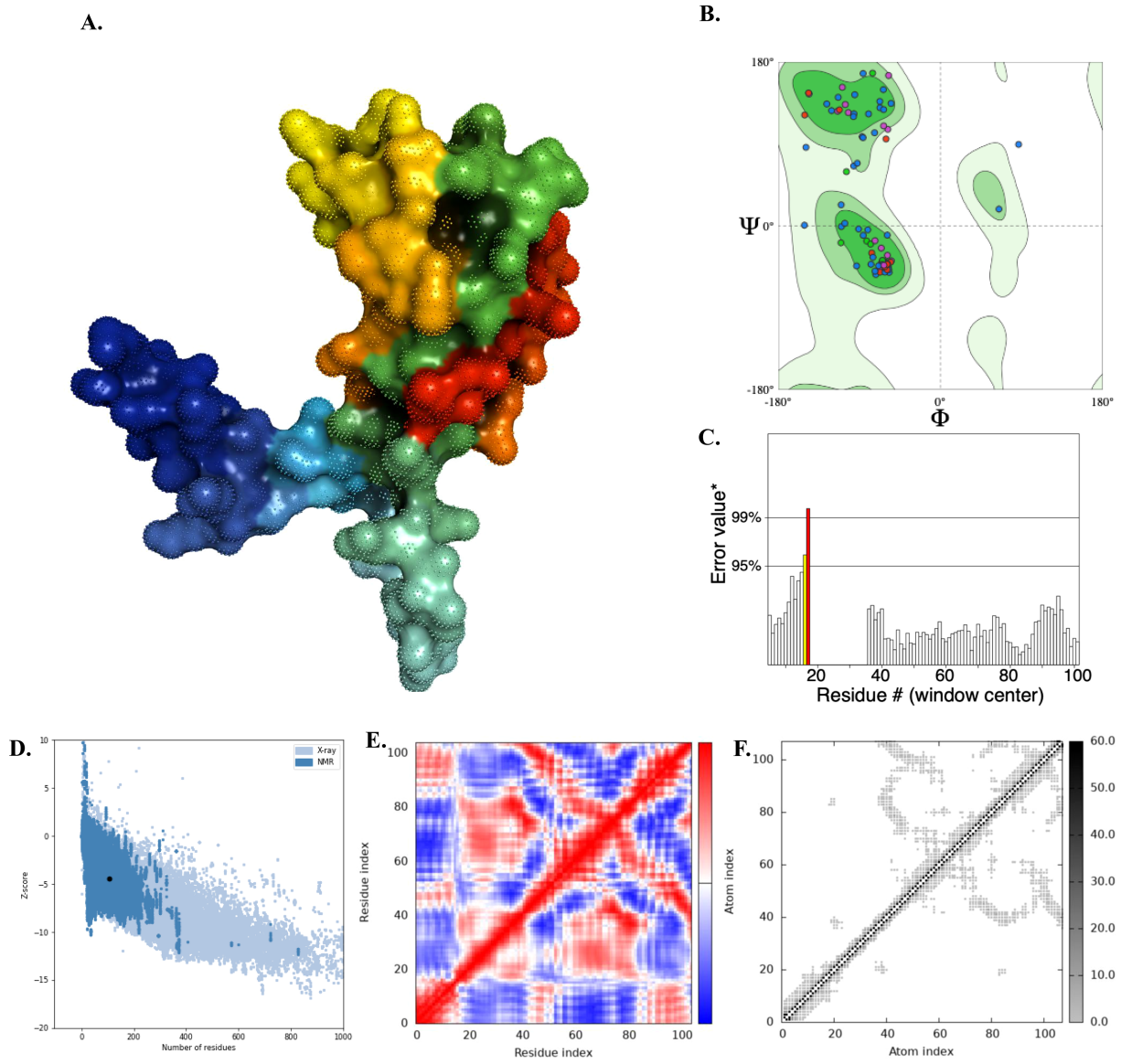


FIGURE 2
Protein 3D modelling, refinement and validation of designed MEV. **(A)** 3D structure of *LpKWTc001* multi-epitope-based vaccine design for MAC species. **(B)** Ramachandran plot **(C)** ERRAT analysis report. **(D)** Xray based on ProSA (Protein Structural Analysis). The Z-score (dark spot) value was **4.41**, within NMR (dark blue) and X-ray (light blue. **(E)** Covariance matrix indicating coupling between pairs of residues, i.e. whether they experience correlated (red), uncorrelated (white) or anti-correlated (blue) motions. **(F)** The elastic network model defining which pairs of atoms are connected by springs. Dots are coloured according to their stiffness, the darker greys indicate stiffer springs and vice versa.

infection-induced antibodies. To evaluate conformational changes of the MEV, protein flexibility was examined using NMA. The MEV showed that a greater part of its peptide chains have high rigid regions which are crucial in the protein's functional dynamics. The 3D structure, refined to improve its quality, exhibited favourable characteristics in the Ramachandran plot, with 85.16% of residues in allowed regions and minimal outliers, confirming the model's reliability and suitability for vaccine design.

Immune simulation showed responses typical of a strong immune reaction, with increased activity after repeated antigen exposure. Following infection with MAC species, IgG1, IgG3, and IgE antibodies are critical for protection, and the vaccine candidate effectively stimulated memory B-cells and T-cells, with B-cell memory lasting several months. The simulations show that THL cells were strongly activated, and levels of IFN-γ and IL-2 spiked after the first injection, staying high with subsequent doses. This suggests strong TH cell activity and efficient antibody production, supporting a robust humoral response. The diversity of the immune response, indicated by the Simpson index, reflects the chimeric peptide's design, which includes multiple B and T-cell epitopes. The dominant IFN-γ-driven TH1-type response, seen in naturally immune individuals, involves higher levels of TH1 cells, cytotoxic

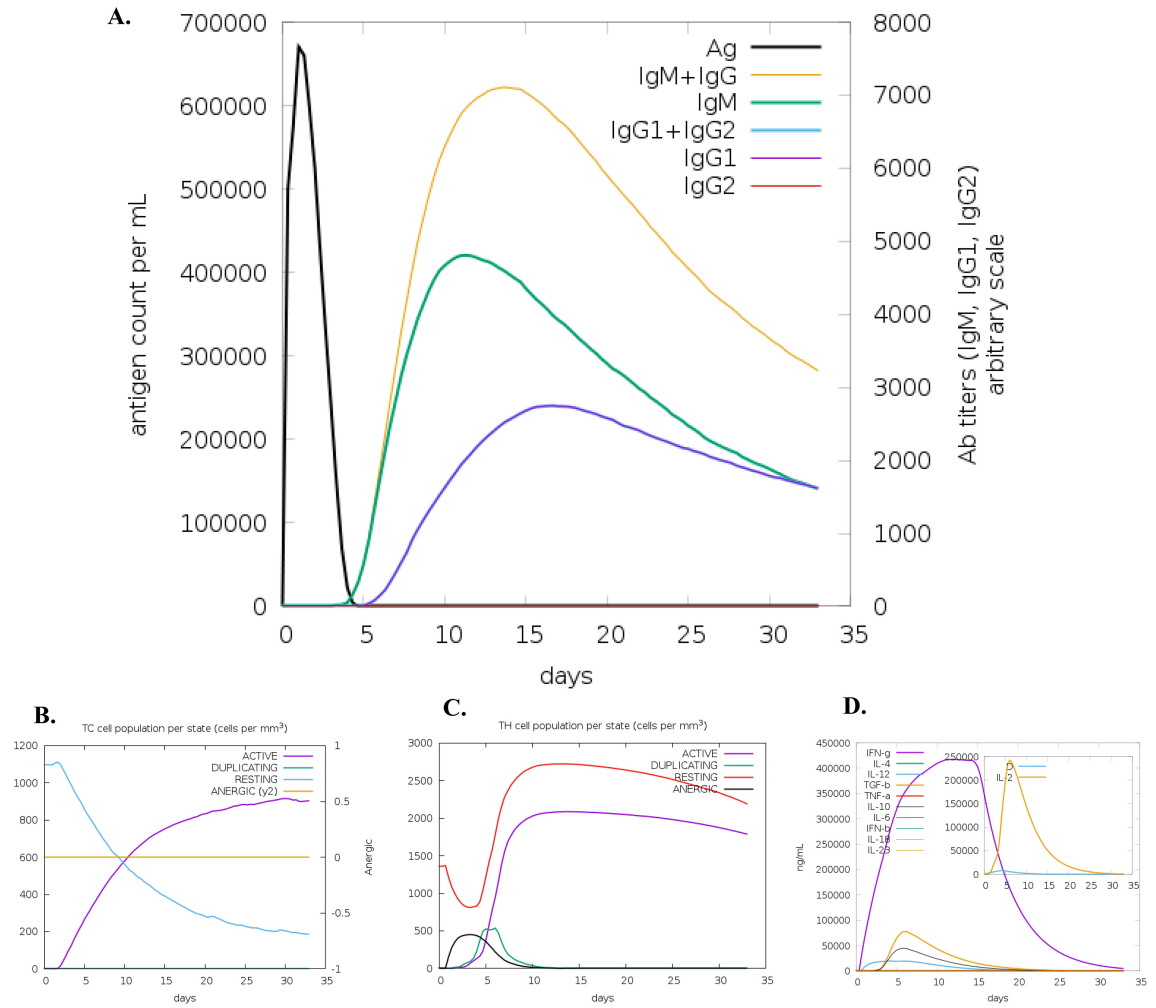


FIGURE 3
C-ImmSim simulation of the cytokine levels induced by the vaccine. **(A)** Antigen and immunoglobulins (antibodies are sub-divided per isotype); **(B)** CD8 T-cytotoxic lymphocytes count per entity-state; **(C)** T helper (TH) cell population, and **(D)** Concentration of cytokines and interleukins.

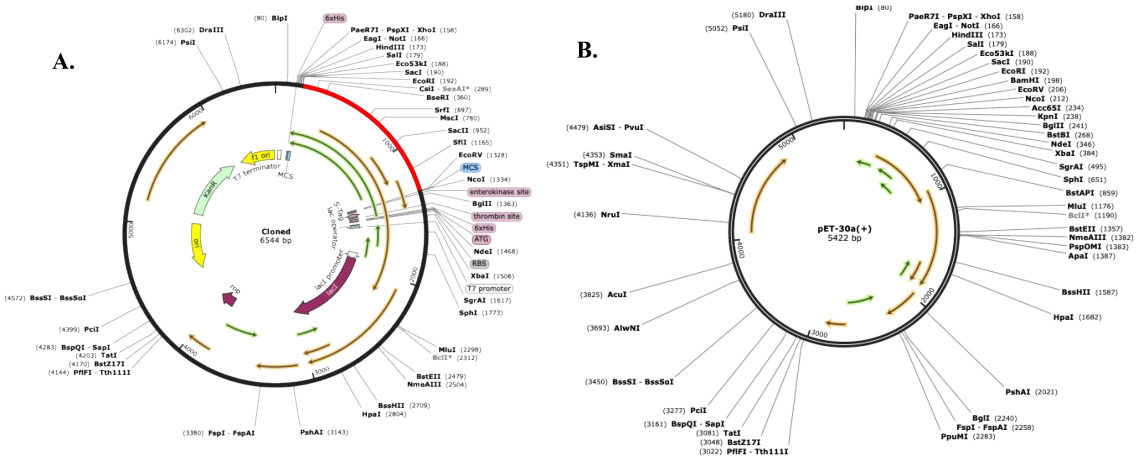


FIGURE 4
Codon optimisation and *in silico* restriction cloning of the multiple-epitope vaccine construct into the expression vector *pET30a(+)*. The codon sequence of each multi-epitope vaccine was inserted in the multiple cloning site (MCS) of the **(A)** *pET30a(+)* expression vector using the Snapgene sequence alignment tool; **(B)** Final clone with fragment *MEV-LpKwTC001*.

CD8+ T cells, neutrophils, and macrophages, further highlighting the vaccine's potential effectiveness.

After obtaining the candidate vaccine, validating a candidate vaccine begins with screening for immunoreactivity using serological analysis, which requires expressing the recombinant protein in a suitable host. *E coli* expression systems, particularly strain K12, are preferred for producing recombinant proteins. To ensure high-level expression of the vaccine protein, codon optimization was performed *in silico*, yielding a favourable codon adaptability index (0.80) and GC content (64.14%).

While our vaccine candidate demonstrated favourable protein characteristics and strong immunogenicity, our study had certain limitations. Since the candidate was designed using human MHC epitopes, its efficacy must be assessed in a humanized mouse model. Additionally, we did not evaluate vaccine efficacy *in vitro*, which remains a limitation of our approach. However, this work represents a crucial foundational step toward experimental vaccine development. Notably, similar multi-epitope vaccines designed through utilization of immunoinformatic tools *in silico* have demonstrated strong immunogenicity in both *in vitro* and *in vivo* models, supporting the reliability of these methods (40–42). Moving forward, to advance the development of a preventive and therapeutic vaccine for MAC, we will validate the proposed vaccine through *in vivo* and *in vitro* studies to corroborate the predicted immunogenic potential.

Additionally, population coverage analysis for our vaccine candidate focused on HLA-DRB alleles. While this approach provides a robust estimation of coverage in African populations, the exclusion of HLA-DQ and HLA-DP loci may result in an underestimation of total MHC class II diversity, particularly in non-African populations. Future studies incorporating these additional loci are warranted to refine global population coverage estimates.

This study highlights a novel vaccine construct capable of eliciting a strong immune response against MAC species, potentially serving as a prototype for vaccines targeting other emerging infectious diseases. Vaccination is an important strategy to induce an immune response against the pathogen by specifically inducing the adaptive immune system. However, current challenges such as the absence of approved vaccines for MAC species, limited epitope-based research, and the lengthy development timelines and high costs associated with traditional vaccine approaches present significant gaps. Addressing these gaps is essential for comprehensive disease control and the reduction of MAC-associated morbidity across human and animal populations. This approach accelerates and lowers the cost of developing diagnostics and vaccines for MAC species, aiding future studies on epitope-based solutions to tackle the NTM challenge.

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 below: https://www.ncbi.nlm.nih.gov/, BioProject ID PRJNA1205738.

Ethics statement

Joint Research Ethics Committee (JREC; under Approval No. JREC/168/22, dated 02.03.2023), Medical Research Council of Zimbabwe (MRCZ; under Approval No. MRCZ/A/29/03, dated 13.06.2023) and Zimbabwe's National Animal Research and Ethics Committee (NAREC). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by- product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

LK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. WC: Data curation, Formal Analysis, Software, Validation, Visualization, Writing – review & editing. TM: Investigation, Writing – review & editing. PN: Conceptualization, Methodology, Supervision, Writing – review & editing. RG: Supervision, Writing – review & editing. LZ: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. NM: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. SG: Formal Analysis, Writing – review & editing. SM: Data curation, Formal Analysis, Funding acquisition, Resources, Software, Validation, Visualization, Writing – review & editing. NC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Research reported in this presentation was supported by the Trials of Excellence in Southern Africa (TESA) Addressing Gender and Diversity Regional Gaps in Clinical Research Capacity (TAGENDI) Project funded by the European and Developing Countries Clinical Trials Partnership (EDCTP) in partnership with the United Kingdom Department of Health and Social Care Award Number PSIA2020AGDG-3319. The content is solely the author's responsibility and does not necessarily represent the official views of the funders.

Acknowledgments

The authors are grateful to EDCTP through TAGENDI Fellowship, the University of Zimbabwe, Medical Microbiology Unit, Department of Laboratory Diagnostic and Investigative Sciences,

Faculty of Medicine and Health Sciences, Biomedical Research and Training Institute (BRTI), National Microbiology Reference Laboratories, Harare and National TB Reference Laboratories, Bulawayo, through the Ministry of Health and Child Care, Zimbabwe and Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Nishiuchi Y, Iwamoto T, Maruyama F. Infection sources of a common non-tuberculous mycobacterial pathogen, mycobacterium avium complex. *Front Med*. (2017) 4:27. doi: 10.3389/fmed.2017.00027

2. Kaczmarkowska A, Didkowska A, Kwiecień E, Stefańska I, Rzewuska M, Anusz K. The Mycobacterium avium complex – an underestimated threat to humans and animals. *Ann Agric Environ Med*. (2022) 29:22–7. doi: 10.26444/aaem/136398

3. Bhanushali J, Jadhav U, Ghewade B, Wagh P. Unveiling the clinical diversity in nontuberculous mycobacteria (NTM) infections: A comprehensive review. *Cureus*. (2023) 15:e48270. doi: 10.7759/cureus.48270

4. Thorel MF, Huchzermeyer H, Weiss R, Fontaine JJ. Mycobacterium avium infections in animals. Literature review. *Vet Res*. (1997) 28:439–47.

5. Inderlied CB, Kemper CA, Bermudez LE. The Mycobacterium avium complex. *Clin Microbiol Rev*. (1993) 6:266–310. doi: 10.1128/CMR.6.3.266

6. Hoefsloot W, Van Ingen J, Andrejak C, Ängeby K, Bauriaud R, Bemer P, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. *Eur Respir J*. (2013) 42:1604–13. doi: 10.1183/09031936.00149212

7. Wetzstein N, Diricks M, Anton TB, Andres S, Kuhns M, Kohl TA, et al. Clinical and genomic features of Mycobacterium avium complex: a multi-national European study. *Genome Med*. (2024) 16:86. doi: 10.1186/s13073-024-01359-8

8. Prevots DR, Marras TK. Epidemiology of human pulmonary infection with nontuberculous mycobacteria. *Clin Chest Med*. (2015) 36:13–34. doi: 10.1016/j.ccm.2014.10.002

9. Adjemian J, Daniel-Wayman S, Ricotta E, Prevots DR. Epidemiology of nontuberculous mycobacteriosis. *Semin Respir Crit Care Med*. (2018) 39:325–35. doi: 10.1055/s-0038-1651491

10. Orujyan D, Narinyan W, Rangarajan S, Rangchaikul P, Prasad C, Saviola B, et al. Protective Efficacy of BCG Vaccine against Mycobacterium leprae and Non-Tuberculous Mycobacterial Infections. *Vaccines*. (2022) 10:390. doi: 10.3390/vaccines10030390

11. Tao ZX, Li RP, Song YY, Xu A. Basic immunization of vaccines is fundamental, and booster immunization is the guarantee: Booster immunization and its public health value. *Zhonghua Yu Fang Yi Xue Za Zhi*. (2022) 56:1401–10. doi: 10.3760/cma.j.cn112150-20220727-00760

12. Adegboro B, Kolawole OM, Lawani O, Folahan F, Seriki AA. A review of the roles of Major Histocompatibility Complex (MHC) molecules in infections. *Afr J Clin Exp Microbiol*. (2022) 23:120–30. doi: 10.4314/ajcem.v23i2.2

13. Krensky AM. The HLA system, antigen processing and presentation. *Kidney Int Suppl*. (1997) 58:S2–7.

14. Thio CL, Thomas DL, Karacki P, Gao X, Marti D, Kaslow RA, et al. Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. *J Virol*. (2003) 77:12083–7. doi: 10.1128/JVI.77.22.12083-12087.2003

15. Gaseitsiwe S, Maeurer MJ. Identification of MHC class II binding peptides: microarray and soluble MHC class II molecules. *Methods Mol Biol Clifton NJ*. (2009) 524:417–26. doi: 10.1007/978-1-59745-450-6_30

16. Pitaloka DAE, Izzati A, Amirah SR, Syakuran LA. Multi epitope-based vaccine design for protection against mycobacterium tuberculosis and SARS-coV-2 coinfection. *Adv Appl Bioinforma Chem*. (2022) 15:43–57. doi: 10.2147/AABC.S366431

17. Zaib S, Rana N, Areeba, Hussain N, Alrbyawi H, Dera AA, et al. Designing multi-epitope monkeypox virus-specific vaccine using immunoinformatics approach. *J Infect Public Health*. (2023) 16:107–16. doi: 10.1016/j.jiph.2022.11.033

18. Sinigaglia F, Hammer J. Defining rules for the peptide-MHC class II interaction. *Curr Opin Immunol*. (1994) 6:52–6. doi: 10.1016/0952-7915(94)90033-7

19. Nielsen M, Andreatta M, Peters B, Buus S. Immunoinformatics: predicting peptide–MHC binding. *Annu Rev BioMed Data Sci*. (2020) 3:191–215. doi: 10.1146/annurev-biodatasci-021920-100259

20. Ernst JD. Antigenic variation and immune escape in the MTBC. *Adv Exp Med Biol*. (2017) 1019:171–90. doi: 10.1007/978-3-319-64371-7_9

21. Karbalaei Zadeh Babaki M, Soleimanpour S, Rezaee SA. Antigen 85 complex as a powerful *Mycobacterium tuberculosis* immunogene: Biology, immune-pathogenicity, applications in diagnosis, and vaccine design. *Microb Pathog*. (2017) 112:20–9. doi: 10.1016/j.micpath.2017.08.040

22. Yuk JM, Jo EK. Host immune responses to mycobacterial antigens and their implications for the development of a vaccine to control tuberculosis. *Clin Exp Vaccine Res*. (2014) 3:155–67. doi: 10.7774/cevr.2014.3.2.155

23. 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 epitopes1. *J Immunol*. (2002) 168(11):5499–506. doi: 10.4049/jimmunol.168.11.5499

24. Chao P, Zhang X, Zhang L, Yang A, Wang Y, Chen X. Proteomics-based vaccine targets annotation and design of multi-epitope vaccine against antibiotic-resistant Streptococcus gallolyticus. *Sci Rep*. (2024) 14:4836. doi: 10.1038/s41598-024-55372-3

25. Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, et al. A potential protein adjuvant derived from mycobacterium tuberculosis rv0652 enhances dendritic cells-based tumor immunotherapy. *PloS One*. (2014) 9:e104351. doi: 10.1371/journal.pone.0104351

26. Yun JS, Kim AR, Kim SM, Shin E, Ha SJ, Kim D, et al. In silico analysis for the development of multi-epitope vaccines against Mycobacterium tuberculosis. *Front Immunol*. (2024) 15:1474346. doi: 10.3389/fimmu.2024.1474346

27. Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: An in silico approach. *Infect Genet Evol*. (2017) 49:309–17. doi: 10.1016/j.meegid.2017.02.007

28. Adianingsih O, Kharisma V. Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. *J Appl Pharm Sci*. (2019) 9:98–103. doi: 10.7324/JAPS.2019.90114

29. 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

30. Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS. Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. *Science*. (1997) 276:1420–2. doi: 10.1126/science.276.5317.1420

31. Chugh S, Bahal RK, Dhiman R, Singh R. Antigen identification strategies and preclinical evaluation models for advancing tuberculosis vaccine development. *NPJ Vaccines*. (2024) 9:1–21. doi: 10.1038/s41541-024-00834-y

32. Sachdeva KS, Chadha VK. TB-vaccines: Current status & challenges. *Indian J Med Res*. (2024) 160:338–45. doi: 10.25259/IJMR_1478_2024

33. Morrison H, McShane H. BCG: past, present and future direction. In: : Christodoulides M, editor. *Vaccines for Neglected Pathogens: Strategies, Achievements and Challenges : Focus on Leprosy, Leishmaniasis, Melioidosis and Tuberculosis*. Springer International Publishing, Cham (2023). p. 171–95. doi: 10.1007/978-3-031-24355-4_8

34. Qu M, Zhou X, Li H. BCG vaccination strategies against tuberculosis: updates and perspectives. *Hum Vaccines Immunother*. (2021) 17(12):5284–95. doi: 10.1080/21645515.2021.2007711

35. McMurray DN. A coordinated strategy for evaluating new vaccines for human and animal tuberculosis. *Tuberculosis*. (2001) 81:141–6. doi: 10.1054/tube.2000.0265

36. Zhuang L, Ye Z, Li L, Yang L, Gong W. Next-generation TB vaccines: progress, challenges, and prospects. *Vaccines*. (2023) 11:1304. doi: 10.3390/vaccines11081304

37. Launois P, Drowart A, Bourreau E, Couppie P, Farber CM, Van Vooren JP, et al. T Cell Reactivity against Mycolyl Transferase Antigen 85 of M. tuberculosis in HIV-TB Coinfected Subjects and in AIDS Patients Suffering from Tuberculosis and Nontuberculous Mycobacterial Infections. *J Immunol Res*. (2011) 2011:640309. doi: 10.1155/2011/640309

38. Ghaffar SA, Tahir H, Muhammad S, Shahid M, Naqqash T, Faisal M, et al. Designing of a multi-epitopes based vaccine against Haemophilius parainfluenzae and its validation through integrated computational approaches. *Front Immunol*. (2024) 15:1380732. doi: 10.3389/fimmu.2024.1380732

39. Facciolà A, Visalli G, Laganà A, Pietro AD. An overview of vaccine adjuvants: current evidence and future perspectives. *Vaccines*. (2022) 10:819. doi: 10.3390/vaccines10050819

40. Invenção M da CV, de Macêdo LS, de Moura IA, Santos LAB de O, Espinoza BCF, de Pinho SS, et al. Design and immune profile of multi-epitope synthetic antigen vaccine against SARS-coV-2: an in silico and *in vivo* approach. *Vaccines*. (2025) 13:149. doi: 10.3390/vaccines13020149

41. Shehata MM, Mahmoud SH, Tarek M, Al-Karmalawy AA, Mahmoud A, Mostafa A, et al. In silico and *in vivo* evaluation of SARS-coV-2 predicted epitopes-based candidate vaccine. *Molecules*. (2021) 26:6182. doi: 10.3390/molecules26206182

42. Kaushik V, G SK, Gupta LR, Kalra U, Shaikh AR, Cavallo L, et al. Immunoinformatics Aided Design and In-Vivo Validation of a Cross-Reactive Peptide Based Multi-Epitope Vaccine Targeting Multiple Serotypes of Dengue Virus. *Front Immunol*. (2022) 13:865180. doi: 10.1021/acsptsci.2c00130

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 11 July 2025
DOI 10.3389/fimmu.2025.1583439

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College & Hospital, India

REVIEWED BY
Juraj Ivanyi,
King's College London, United Kingdom
Faraz Ahmad,
University of Missouri, United States

*CORRESPONDENCE
Subramanian Dhandayuthapani
pani.dhandayuthapani@utrgv.edu
Shrikanth S. Gadad
shrikanth.gadad@utrgv.edu

†PRESENT ADDRESSES
Melina J. Sedano,
South Texas Center of Excellence in Cancer Research, Department of Medicine and Oncology, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, United States
Shrikanth S. Gadad,
South Texas Center of Excellence in Cancer Research, Department of Medicine and Oncology, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, United States
Subramanian Dhandayuthapani,
South Texas Center of Excellence in Cancer Research, Department of Medicine and Oncology, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, United States

‡These authors have contributed equally to this work

RECEIVED 25 February 2025
ACCEPTED 03 June 2025
PUBLISHED 11 July 2025

CITATION
Veerapandian R, Yang B, Carmona A, Sedano MJ, Reid V, Jimenez R, Chacon J, Jagannath C, Ramos EI, Gadad SS and Dhandayuthapani S (2025) Comparative transcriptomic analysis of mouse macrophages infected with live attenuated vaccine strains of *Mycobacterium tuberculosis*. *Front. Immunol.* 16:1583439. doi: 10.3389/fimmu.2025.1583439

Comparative transcriptomic analysis of mouse macrophages infected with live attenuated vaccine strains of *Mycobacterium tuberculosis*

Raja Veerapandian[1‡], Barbara Yang[2‡], Areanna Carmona[1], Melina J. Sedano[2†], Victoria Reid[1,2], Rodrigo Jimenez[1,3], Jessica Chacon[4], Chinnaswamy Jagannath[5], Enrique I. Ramos[2,6], Shrikanth S. Gadad[2,3*†] and Subramanian Dhandayuthapani[1,3*†]

[1]Center of Emphasis in Infectious Diseases, Department of Molecular and Translational Medicine, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX, United States, [2]Center of Emphasis in Cancer, Department of Molecular and Translational Medicine, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX, United States, [3]Frederick L. Francis School of Biomedical Sciences, Texas Tech University Health Sciences Center El Paso, El Paso, TX, United States, [4]Department of Medical Education, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX, United States, [5]Department of Pathology and Genomic Medicine, Houston Methodist Research Institute & Weill Cornell Medical College, Houston, TX, United States, [6]Department of Biology, University of Texas El Paso, El Paso, TX, United States

The BCG vaccine has been used against tuberculosis (TB) for over a hundred years; however, it does not protect adults from pulmonary TB. To develop alternative vaccines against TB, we generated *Mycobacterium tuberculosis* H37Rv (Mtb)-derived vaccine strains by rationally deleting key virulent genes, resulting in single (*SKO; ΔfbpA*), double (DKO; *ΔfbpA-ΔsapM*), triple (TKO-D; *ΔfbpA-ΔsapM-ΔdosR* and TKO-Z; *ΔfbpA-ΔsapM-Δzmp1*), and quadruple (QKO; *ΔfbpA-ΔsapM-Δzmp1-dosR*) strains. To understand how macrophages, the host cells that defend against infection and process antigens for presentation to immune cells, respond to these vaccine strains, we performed transcriptomic analyses of mouse bone marrow-derived macrophages (BMDMs) infected with these strains. The transcriptomic data were compared with similar data obtained from macrophages infected with Mtb H37Rv and BCG. Our analyses revealed that genes associated with various immune and cell signaling pathways, such as NF-kappa B signaling, TNF signaling, cytokine-cytokine receptor interaction, chemokine signaling, hematopoietic cell lineage, Toll-like receptor signaling, IL-17 signaling, Th1 and Th2 cell differentiation, Th17 cell differentiation, and T cell receptor signaling were differentially expressed in BMDMs infected with our vaccine strains. Enhanced expression of cytokines and chemokines, including proinflammatory cytokines such as TNF-α, IL-6, GM-CSF, and IL-1, which are essential for the immune response against Mtb infection, was also observed in

BMDMs infected with these strains. In particular, BMDMs infected with all vaccine strains exhibited a significant upregulation of genes associated with the IL-17 pathway. These results may indicate that our vaccine strains could induce a protective immune response against TB.

KEYWORDS

Mtb-vaccines, BCG, mouse, macrophages, RNA-sequencing, transcripts, immune signaling, IL-17

Introduction

Tuberculosis (TB) is a deadly disease caused by an intracellular human pathogen, *Mycobacterium tuberculosis* (Mtb), which has coexisted with humans for approximately seventy thousand years (1, 2). According to the World Health Organization (WHO) report, nearly 10.8 million people were affected by TB, resulting in a mortality rate of 1.25 million in 2023 (3). Historically considered hereditary, TB was recognized as a contagious disease by Jean-Antoine Villemin in 1865, and Robert Koch identified the causative bacterium, Mtb, in 1882 (4). Selman Waksman developed Streptomycin, the first effective TB drug, earning him the Nobel Prize in 1952 (5). Conversely, the BCG vaccine developed by Albert Calmette and Camille Guérin remains the only licensed vaccine against TB. Unfortunately, BCG has not been regarded as an effective vaccine against TB because of the emergence of various sub-strains that produce differing levels of protective efficacy (6). The rise of multidrug-resistant TB (MDR-TB) has complicated treatment strategies even further, necessitating confirmation of bacterial infection and testing for antibiotic resistance. Alarmingly, only two out of five MDR-TB cases received treatment in 2022 (7).

In 2014, WHO launched the "End TB Strategy" (8) to significantly reduce the TB burden by 2035, emphasizing the critical role of vaccines. Various vaccine types, including live attenuated vaccines (LAV), subunit vaccines, viral vectored vaccines, DNA vaccines, whole-cell killed/inactivated vaccines, and recombinant protein-adjuvant formulations, have been developed and studied for TB prevention (9). Among these, LAV stands out for its ability to induce long-lasting immune responses, with BCG serving as a prime example. The BCG vaccine differs from the Mtb strain due to the deletion of various Mtb-specific open reading frames (ORFs) clustered in 16 genomic regions of difference (RD1–RD16) (10, 11). Mtb has also been modified to enhance its vaccine efficacy, particularly by knocking out the secretory proteins or secretory systems of mycobacteria (12).

As a first of its kind, we reported that the Mtb Δ*fbpA* strain protects mice against challenges similar to or better than BCG (13). Fibronectin-binding protein (FbpA; Rv3804c) is a secreted protein belonging to the Ag85 complex, which is highly conserved among species of the *Mycobacterium tuberculosis* complex. It has a mycolyltransferase enzyme function, catalyzing mycolic acid transfer during cell wall biogenesis (14). Gene disruption studies in Mtb demonstrated that FbpA is one of the key components necessary for intracellular survival (15). To enhance the vaccine efficacy, we additionally deleted the *sapM* gene in the Δ*fbpA* strain to create a double knockout (DKO) (16). The *sapM* gene (*Rv3310*) encodes the secreted acid phosphatase SapM, initially identified in Mtb (17). It interferes with the phagosome maturation by dephosphorylating PI-3 phosphate (18). Our DKO vaccine strain induces strong protection through enhanced antigen processing and the autophagy mechanism (19). To further enhance our DKO vaccine, we carefully deleted two additional genes, specifically *zmp1* (*Rv0198c*) and *dosR* (*Rv3133c*). Zmp1 is a ~75 kDa zinc metalloprotease secretory protein that plays a significant role in blocking phagosome maturation and impairing inflammasome activation, resulting in greater vaccine efficacy, (20, 21), whereas DosR is a dormancy survival regulator that collectively affects approximately fifty genes in the Mtb genome and is highly activated under microenvironmental conditions such as granulomas (22, 23). These new vaccine strains have shown increased immunogenicity (24), and efficacies against TB in animal models are being investigated.

This study follows up on our prior observation that a double-knockout (DKO) vaccine provides superior and longer-lasting protection compared to the BCG vaccine (19). In this study, we aimed to investigate the intricate molecular responses of macrophages to the Mtb-based live attenuated vaccines (LAVs) developed in our laboratory. Macrophages play a crucial role in defending against intracellular pathogens like Mtb and processing and presenting antigens to immune cells. An effective mycobacterial vaccine should induce key immune and cell signaling pathways that lead to effective antigen presentation and the subsequent pathogen clearance from the host. Thus, studying the molecular interactions between Mtb-derived vaccines and macrophages through RNA-seq analysis should provide important insights into vaccine efficacy. Although studies have documented genome-wide transcriptomic changes in human or mouse macrophages following Mtb infection (25–27), our study focuses on Mtb-derived vaccine strains for the first time. This approach has allowed us to identify the crucial immune and cell signaling pathways and profile the vital cytokines and chemokines for TB vaccines. Further, our findings underscore the importance of the IL-17 pathway regulated by LAV strains.

Materials and methods

Mtb strains and culture conditions

Wild-type and knockout Mtb strains were grown at 37°C in either Middlebrook 7H9 broth or 7H10 agar (BD Difco), both containing 0.05% Tween 80 (TW), 0.2% glycerol, and OADC (10%) enrichment. All mutant Mtb strains used in this study are a derivative of H37Rv. We published the single knockout, SKO (*ΔfbpA*) and double knockout, DKO (*ΔfbpA-ΔsapM*) strains used in this study earlier (13, 16). Triple knockouts such as ΔTKO-D (*ΔfbpA-ΔsapM-ΔdosR*), ΔTKO-Z (*ΔfbpA-ΔsapM-Δzmp1*), and quadruple knockout, ΔQKO (*ΔfbpA-ΔsapM-Δzmp1-dosR*) strains were made on the DKO background, and their construction was briefly reported earlier (24) and will be published elsewhere (manuscript under preparation).

Animals and ethics

C57BL/6J mice aged 4–6 weeks were purchased from Jackson Laboratories, Bar Harbor, ME. The mice received were housed with unlimited access to water and mouse chow and permitted to move without restraints within their cages at the Laboratory Animal Resource Center, Texas Tech University Health Sciences Center El Paso. The Institutional Animal Care and Use Committee (IACUC) of the Texas Tech University Health Sciences Center El Paso approved an animal protocol for this study (Protocol #17003).

Isolation of bone marrow-derived macrophages

As described previously, BMDMs were extracted from wild-type C57BL/6 mice (28). Briefly, the BMDMs were cultured from the femurs and tibias of mice in DMEM medium (DMEM, 10% fetal bovine serum, 10 ng/ml of M-CSF) and incubated at 37°C in 5% CO_2 for 7 days, with the addition of new medium containing M-CSF every 2–3 days.

Infection of BMDMs with Mtb strains

Mtb wild-type H37Rv and all mycobacterial vaccine strains were cultured in 7H9 medium with appropriate antibiotics in roller bottles at 37°C for 5–7 days. Colony-forming units (CFUs) of the bacterial suspensions were determined and stored at -80°C until use. Before infection, bacteria were pelleted, washed with PBS, and dispersed using a 23G syringe to eliminate clumps. Mouse BMDMs (10^6 cells/well) seeded in 6-well tissue culture plates (Corning, USA) were infected at a multiplicity of infection (MOI) of 1:5 in DMEM for 4 h to allow phagocytosis. Afterward, cells were washed thrice with D-PBS (Corning, USA) to remove extracellular bacteria and replaced with fresh DMEM containing 10% fetal bovine serum for further incubation at 24 and 72 h.

RNA sequencing and data analysis

Mouse BMDMs (10^6/well) were infected with the respective mycobacterial strains as described above for different time points (24 h and 72 h). After respective time points, total RNA was extracted from the infected BMDMs using the EZ-10 DNAaway RNA Mini-Preps Kit (Bio Basic, Canada) as described previously (28). RNA quantification was performed using Nanodrop (Thermo Scientific, USA). The quality of RNA was measured using TapeStation (Agilent Technologies 4200). The library preparation enriched for polyA RNA fraction was performed in house and RNA sequencing was performed at Novogene Corporation Inc. (Sacramento, CA, USA), as described elsewhere (29). Two biological replicates for each condition were performed. We employed web-based application Genialis to analyze RNA sequencing raw data using their "General RNA-Seq pipeline (featureCounts)" with default settings (Genialis, Inc., Boston, MA). RNA-Seq data were aligned using STAR aligner to the mouse transcriptome from Ensembl release version 109 with trimmed reads removing adapter sequences. Read counts were computed using featureCounts. Quality control metrics were determined, and the average quality per read was 36 (Phred score) (Supplementary files - Table S1,S2, S2). Principle component analysis (PCA) was generated within the Genialis RNA-seq pipeline visualization features. The differentially expressed genes (DEGs) were also computed using the Genialis built-in DESeq2 tool, defining the control samples of PBS or H37Rv and the case samples accordingly, and the filtering criteria for DEGs are FDR < 0.05 with $\log_2$ fold change greater than 1 for upregulated DEGs and less than 1 for downregulated DEGs. DEGs were presented in heatmaps, volcano plots, and Venn diagrams, using pheatmap, ggplot2, and Venn packages, respectively, in the R program. For downstream analysis of KEGG pathway analysis and Gene Ontology (GO) analysis for biological processes (BP), we queried the bioinformatic Database for Annotation, Visualization, and Integrated Discovery (DAVID) with default settings and plotted the top 30 KEGG pathways or BP based on ascending p-value as dot plots using ggplot2 in R, as described previously (28, 29).

cDNA synthesis and qRT-PCR

The total RNA from infected BMDM was used to synthesize cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo) according to the manufacturer's protocol. Quantitative reverse transcriptase PCR (qRT-PCR) was performed using a LightCycler® 96 Instrument (Roche). PCR was performed using PowerTrack™ SYBR Green (Thermo Fisher Scientific) according to the manufacturer's recommendations. Three biological replicates for each condition were performed. Primer details are given in the Supplementary files (Supplementary Table S3). The relative CT (ΔΔCT) method was used to quantify gene expression as described elsewhere (19). The expression levels of target genes were normalized to the house keeping gene, *actB* (ß- actin) with the H37Rv group set as the reference value 1 for comparison with all vaccine groups.

Results

Transcriptome analysis of mouse BMDMs infected with vaccine strains compared to uninfected cells

To explore the variation in transcriptional signatures of mouse BMDMs infected with various Mtb vaccine strains, we conducted a genome-wide gene expression analysis using an RNA-sequencing platform (Figure 1). The performance of Principal Component Analysis (PCA) on the transcripts from mouse BMDMs clearly distinguished the infection groups from the PBS control at both 24 and 72 h time points. At 24 h post-infection, groups of transcripts with H37Rv background knockouts clustered almost entirely together, distinctly separating from those associated with BCG. By 72 h post-infection, almost all groups were distinctly separated, regardless of their H37Rv or BCG background (Supplementary Figure S1).

For gene expression analysis, we considered fold change cut-off values of Log_2 fold change >1.0 as upregulated and <-1.0 for downregulated genes (FDR < 0.05). Our transcriptome analysis identified more than 14,000 genes exhibiting expression among Mtb-infected mouse BMDMs compared to uninfected control (Supplementary Data 1, 2). The heat map displayed a gene expression profile, showing a high disparity among infected groups compared to the PBS group, regardless of the time point (Figures 2A, D). Different clusters in the heat map indicate distinct modes of regulation, with cluster 1 being predominantly upregulated and cluster 3 downregulated across all vaccine groups. Interestingly, cluster 2 remains unchanged in the H37Rv wild-type group; however, there is a significant difference in this cluster among the vaccine groups. Notably, an additional change in cluster 2 of the vaccine groups is observed at the 72 h time point compared to 24 h. The Venn diagram illustrates both the unique and shared DEGs among the vaccine-infected groups. The total number of unique genes in the various vaccine-infected groups at 24 and 72 h post-infection are as follows: H37Rv (59 and 58), BCG (1309 and 422), SKO (52 and 126), DKO (206 and 225), TKO-D (111 and 155), TKO-Z (44 and 15), and QKO (23 and 40) (Figures 2B, E).

There are no significant percentage differences in differential gene expression between 24 and 72 h post-infection within the same group (Figures 2C, F). Compared to H37Rv background vaccine strains (~12.5-20.9%), BCG displays a higher percentage of DEGs at both time points (25.8 & 21.8%), while the H37Rv wildtype shows (14.4 & 12.4%) DEGs. Interestingly, the percentage of DEGs decreased successively as the gene deletion increased in vaccine strains such as SKO (18.7 &20.9%), DKO (18.8 &18.4%), TKO-D (19.8 &19.3%), TKO-Z (14.3 &12.5%), and QKO (13.2 &14.7%).

Transcriptome analysis of mouse BMDMs infected with Mtb vaccine strains compared to H37Rv wild-type strain

As we observed differential regulation of genes within the vaccine groups, we proceeded to determine the number of DEGs in these groups compared to the H37Rv background (Supplementary Data 3, 4). All our vaccine strains originate from the H37Rv background, where genes are sequentially deleted to create mutant strains. Thus, we compared all our vaccine strains, including BCG, with H37Rv. Unlike the previous comparison with uninfected, where three distinct

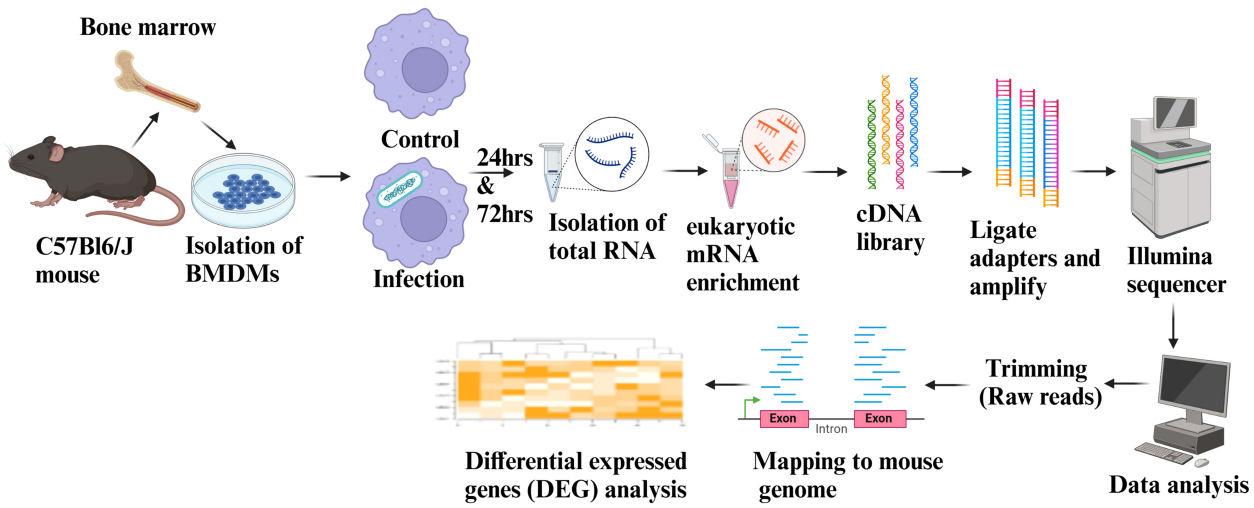


FIGURE 1
Schematics showing RNA-Seq workflow and data analysis. Fresh BMDMs were isolated from the female wild-type C57BL/6J mice and infected with respective vaccine strains or left uninfected. Following a 4-hour phagocytosis period, the BMDMs were washed with D-PBS and cultured in fresh DMEM supplemented with 10% fetal bovine serum for an additional 24 and 72 hours. RNA was isolated and subjected to eukaryotic mRNA enrichment at each time point. Subsequently, cDNA libraries were prepared, followed by adapter ligation and amplification for Illumina sequencing. The RNA-Seq data were aligned to the mouse transcriptome, and differential gene expression (DEG) analysis was performed. The figure was generated using BioRender.

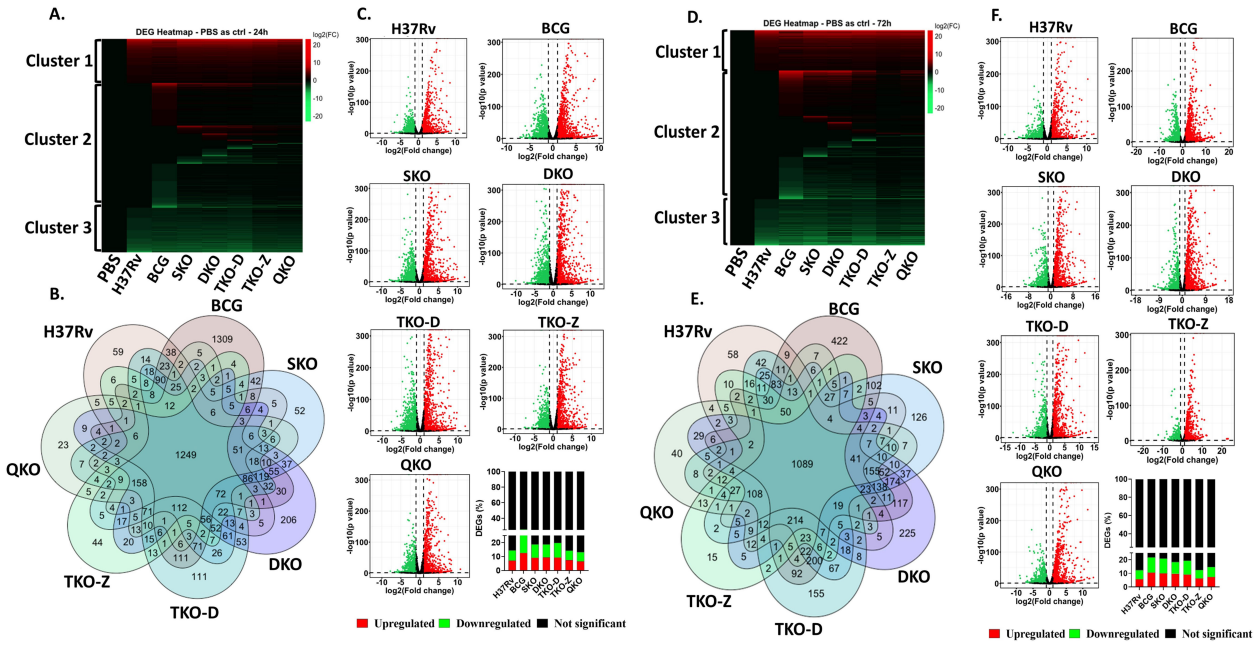


FIGURE 2
Transcriptome profiling of mouse BMDMs infected with vaccine strains in comparison to control. **(A, D)**, Heatmap of the differentially expressed genes at 24 h and 72 h; **(B, E)**, Venn diagram of the differentially expressed genes showing the number of overlapping and unique genes among groups at 24 h and 72 h; **(C, F)**, Volcano plot showing distribution of p values and $\log_2$ fold change of differentially expressed genes as green (downregulated), red (upregulated) and black (Not significant) at 24 h and 72 h with respective bar graphs in percentage.

clusters were observed, here we observed two distinct clusters. While minimal changes are noted in both clusters among the vaccine groups, with the exception of BCG and DKO, clusters 1 and 2 display a high number of upregulated and downregulated genes in the BCG group, along with only a few differentially regulated genes in the H37Rv background vaccine groups. Interestingly, certain alterations are observed in the H37Rv background vaccine groups in regions where the BCG group shows no changes. In contrast, the DKO group exhibits more drastic changes in gene expression in those regions (Figures 3A, D). The Venn diagram reveals a limited number of unique and shared genes among the vaccine groups. At 24 and 72 h post-infection, the number of unique genes in various vaccine-infected groups is as follows: BCG (597 and 512), SKO (5 and 28), DKO (199 and 331), TKO-D (20 and 22), TKO-Z (8 and 8), and QKO (11 and 29). Notably, BCG and DKO exhibit numerous unique DEGs, indicating distinct genetic responses (Figures 3B, E). According to the Volcano plot analysis, the percentage of DEGs at 24- and 72-h post-infection is as follows: BCG (5.4 &9.7%), SKO (1.1 &5.3%), DKO (3.1 & 9%), TKO-D (1 &2.6%), TKO-Z (0.4 &0.8%), and QKO (0.4 &3.2%). Greater percentage differences in differential gene expression are observed between 24- and 72 h post-infection within the same group, except for TKO-D and TKO-Z. Additionally, compared to H37Rv background vaccine strains (~0.4-9%), BCG exhibits a higher percentage of DEGs at both time points (5.4 & 9.7%) (Figures 3C, F).

KEGG pathway analysis of mouse BMDMs infected with Mtb vaccine strains versus uninfected cells

KEGG pathway analysis was performed for all vaccine groups in comparison to the transcriptome of uninfected controls to identify pathways associated with mycobacterial infection. Several top enriched pathways were unique to the vaccine groups. Pathways such as graft-versus-host disease, allograft rejection, leishmaniasis, type I diabetes mellitus, TNF signaling, rheumatoid arthritis, inflammatory bowel disease, influenza A, NF-kappa B signaling, viral protein interaction with cytokines and cytokine receptors, Epstein-Barr virus infection, NOD-like receptor signaling, phagosome, cytokine-cytokine receptor interaction, Kaposi sarcoma-associated herpesvirus infection, measles, lipid and atherosclerosis, and COVID-19 were enriched with the upregulated DEGs of each condition across all vaccine groups (Supplementary Figures S2, S3). Conversely, pathways such as DNA replication, homologous recombination, cell cycle, Fanconi anemia, progesterone-mediated oocyte maturation, p53 signaling, oocyte meiosis, focal adhesion, cellular senescence, pathways in cancer, efferocytosis, ECM-receptor interaction, small cell lung cancer, motor proteins, PI3K-Akt signaling, Rap1 signaling, and MAPK signaling exhibited with downregulated DEGs across all vaccine groups (Supplementary Figures S4, S5).

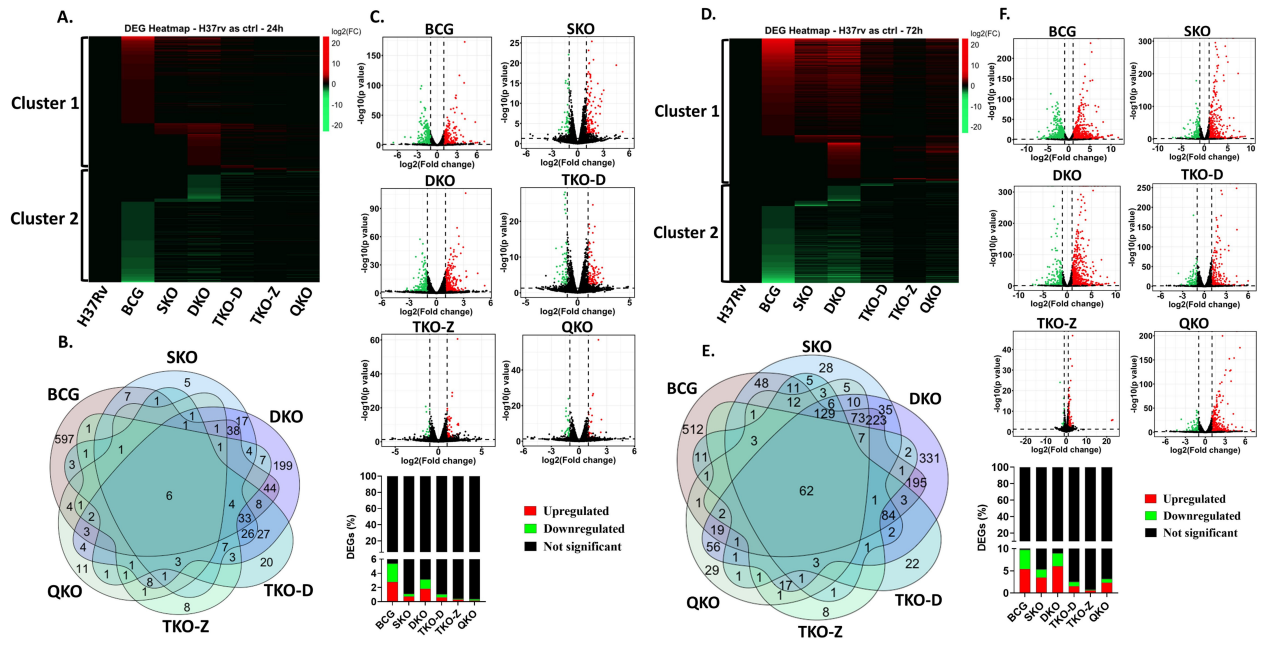


FIGURE 3
Transcriptome profiling of mouse BMDMs infected with vaccine strains in comparison to Mtb H37Rv. **(A, D)**, Heatmap of the differentially expressed genes at 24 h and 72 h; **(B, E)**, Venn diagram of the differentially expressed genes showing the number of overlapping and unique genes among groups at 24 h and 72 h; **(C, F)**, Volcano plot showing differentially expressed genes as green (downregulated), red (upregulated) and black (Not significant) at 24 h and 72 h with respective bar graphs in percentage.

KEGG pathway analysis of mouse BMDMs infected with mycobacterial vaccine strains compared to H37Rv

To understand the roles of cellular pathways in the host's response to mycobacterial infection, we performed a KEGG pathway enrichment analysis with the DEGs using Mtb H37Rv wild-type as the control. We identified multiple pathways that were uniquely and differentially dysregulated at both time points, with a similar pattern observed within the same groups across time points and between groups at each time point (Figures 4–7). Pathways including rheumatoid arthritis, viral protein interaction with cytokines and cytokine receptors, the IL-17 signaling pathway, hematopoietic cell lineage, and cytokine-cytokine receptor interaction were uniquely enriched with upregulated DEGs in all vaccine-infected macrophage groups compared to H37Rv (Figure 5). Notably, TKO-Z and QKO showed a delayed response in enriching those pathways in upregulated DEGs at 72 h. However, several pathways, such as the biosynthesis of unsaturated fatty acids, PPAR signaling, and fatty acid metabolism, exhibited overrepresentation in upregulatted DEGs (Figure 4) at 24 h. A delayed IL-17 response at 72 h was also noted in TKO-Z and QKO when compared to other vaccine groups (Figure 5).

Intriguingly, the pathways affected by downregulated DEGs varied across different vaccine groups. While BCG showed multiple pathways enriched in downregulated DEGs at both time points, our vaccine groups exhibited less pathways at 24 h (Figure 6) but pathways enriched increased in downregulated DEGs at 72 h (Figure 7). Specifically, TKO-Z had a few pathways, including cardiac muscle contraction, oxidative phosphorylation, prion disease, Alzheimer's disease, thyroid cancer, p53 signaling, and thyroid hormone synthesis overrepresented in downregulated DEGs (Figures 6, 7). Uniquely, the ferroptosis pathway, critical in mycobacterial infection, was enriched in both BCG and TKO-D downregulated DEGs at 72 h (Figure 7). Additionally, BCG, SKO, TKO-D, and QKO downregulated DEGs enriched the ECM-receptor interaction and focal adhesion pathways at 72 h, potentially limiting molecule translocation across barriers. Pathways such as protein digestion and absorption and PI3K-Akt signaling were enriched at 72 h in SKO, TKO-D, and QKO downregulated DEGs (Figure 7).

We also compared BCG with H37Rv, where pathways such as rheumatoid arthritis, viral protein interaction with cytokines and their receptors, IL-17 signaling pathway, hematopoietic cell lineage, inflammatory bowel disease, type 1 diabetes mellitus, TNF signaling pathway, NF-kappa B signaling pathway, cytokine-cytokine receptor interaction, and chemokine signaling pathway were similarly enriched with the upregulated DEGs in the BCG group at both time points (24 h and 72 h) (Figures 4, 5). When observing the pathways enriched in downregulated DEGs, we noted pathways like influenza A, measles, antigen processing and presentation, hepatitis C, parathyroid hormone synthesis, secretion, and action, osteoclast differentiation, hepatitis B, human papillomavirus infection, cell adhesion molecules, Epstein-Barr virus infection, calcium signaling pathway, COVID-19, PI3K-Akt signaling pathway, and pathways in cancer were consistent at both time

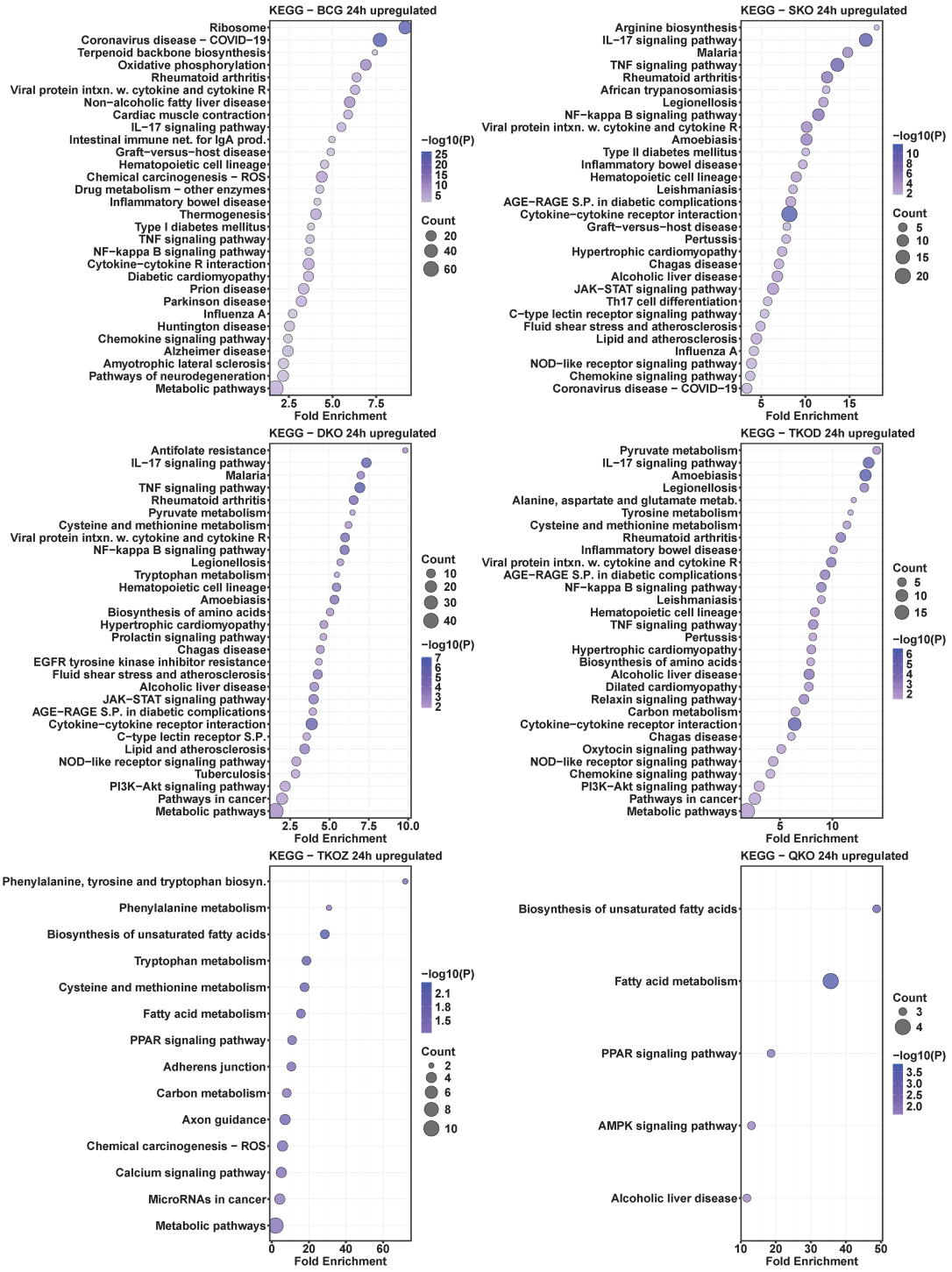


FIGURE 4
KEGG pathway analysis in upregulated differentially expressed transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 24 h post-infection. Dot plots illustrate the top 30 enriched pathways in upregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each pathway, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

points (24 h and 72 h). However, multiple pathways were uniquely dysregulated at their respective time points (Figures 6, 7).

Subsequently, we compared the pathways overrepresented in differentially regulated genes by BCG vaccine strains with those from Mtb-derived vaccines to assess how closely our newly developed Mtb-derived vaccines mimic BCG. Remarkably, similar KEGG pathways were enriched in our Mtb-derived vaccine strains compared to BCG in upregulated DEGs particularly at the 72 h time point (Supplementary Figure S6). At the 24 h time point, only a few pathways such as rheumatoid arthritis, viral protein interaction

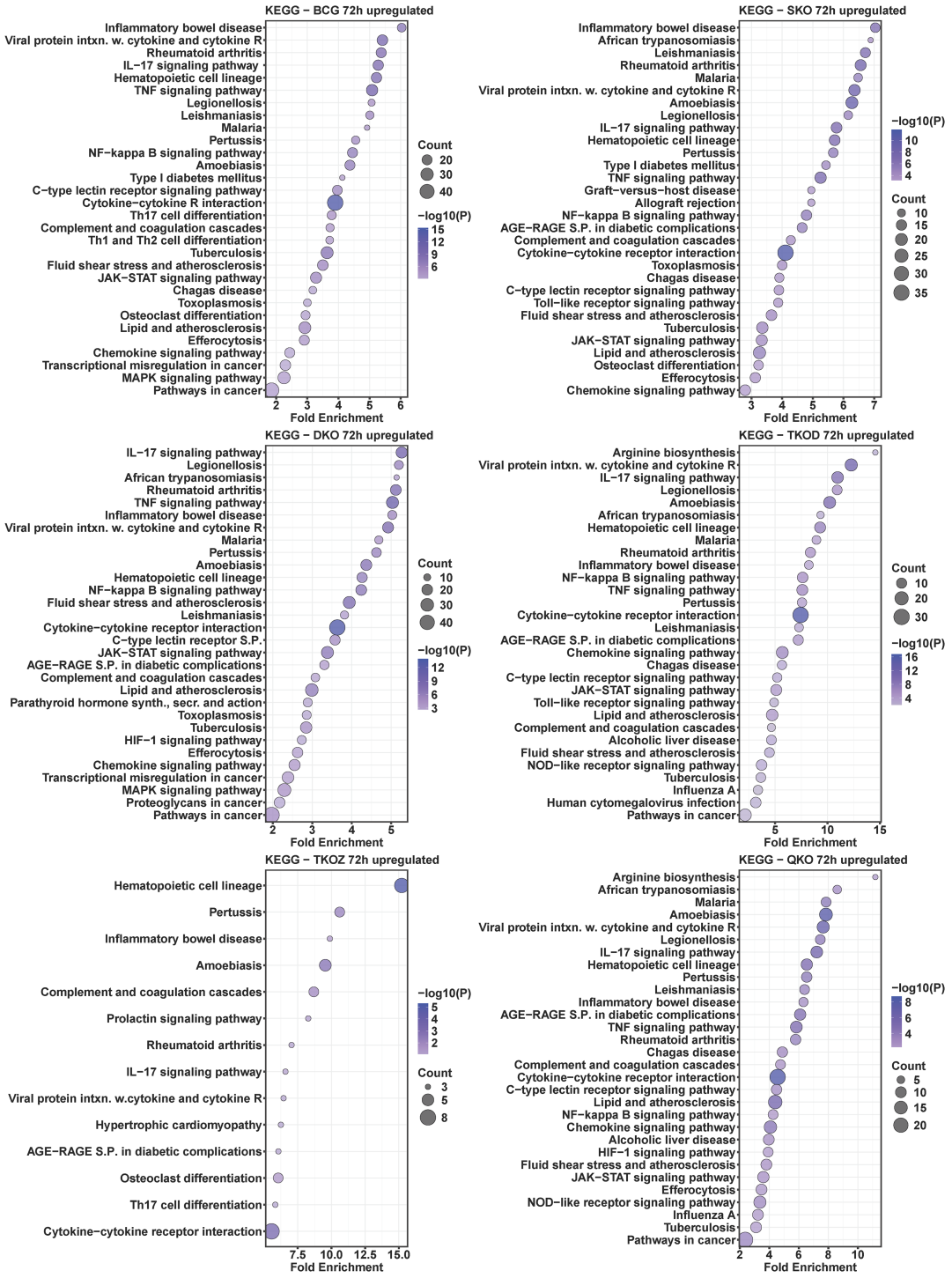


FIGURE 5
KEGG pathway analysis in upregulated differentially expressed transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 72 h post-infection. Dot plots illustrate the top 30 enriched pathways in upregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each pathway enriched, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

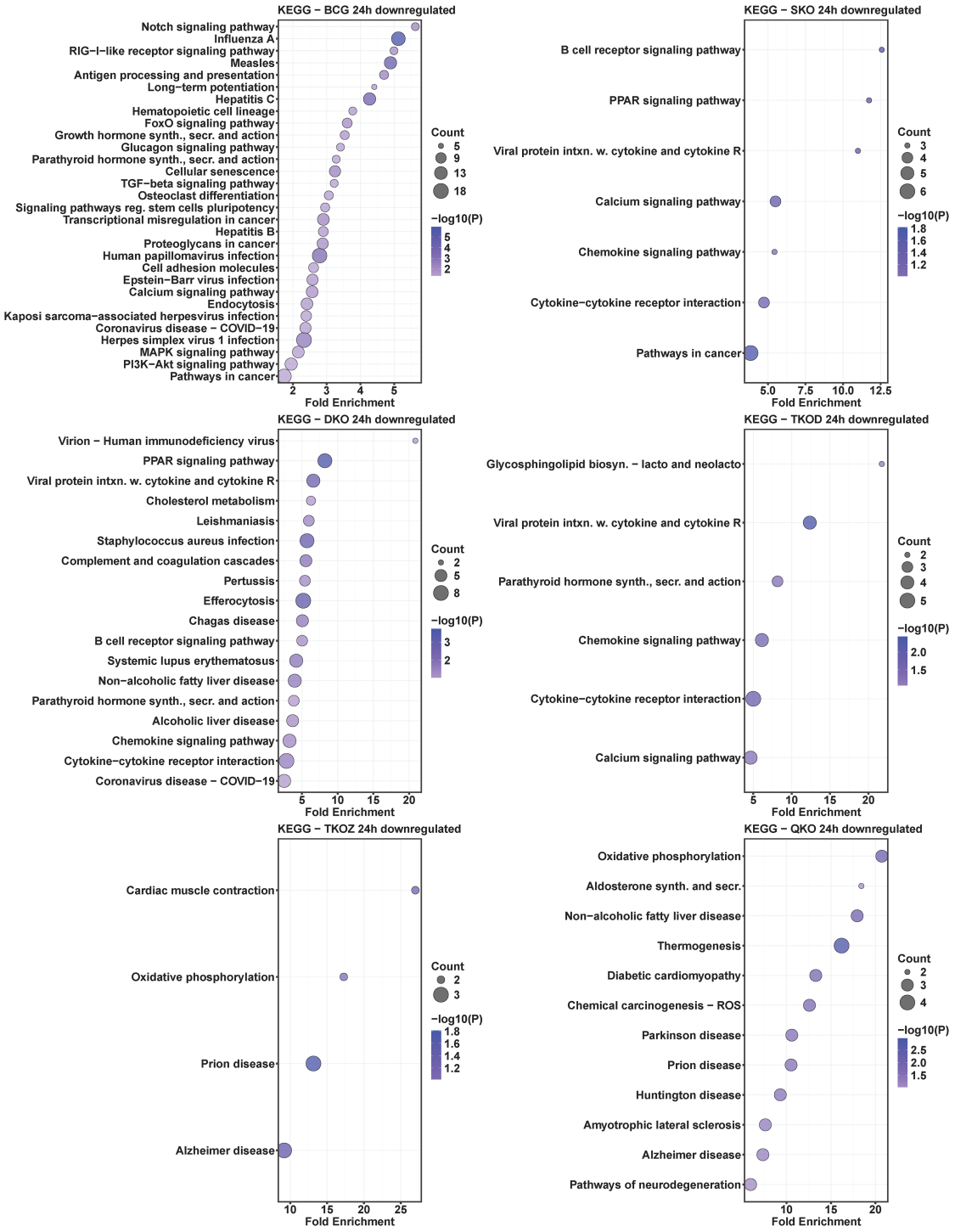


FIGURE 6
KEGG pathway analysis in downregulated differentially expressed transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 24 h post-infection. Dot plots illustrate the top 30 enriched pathways in downregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each category, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

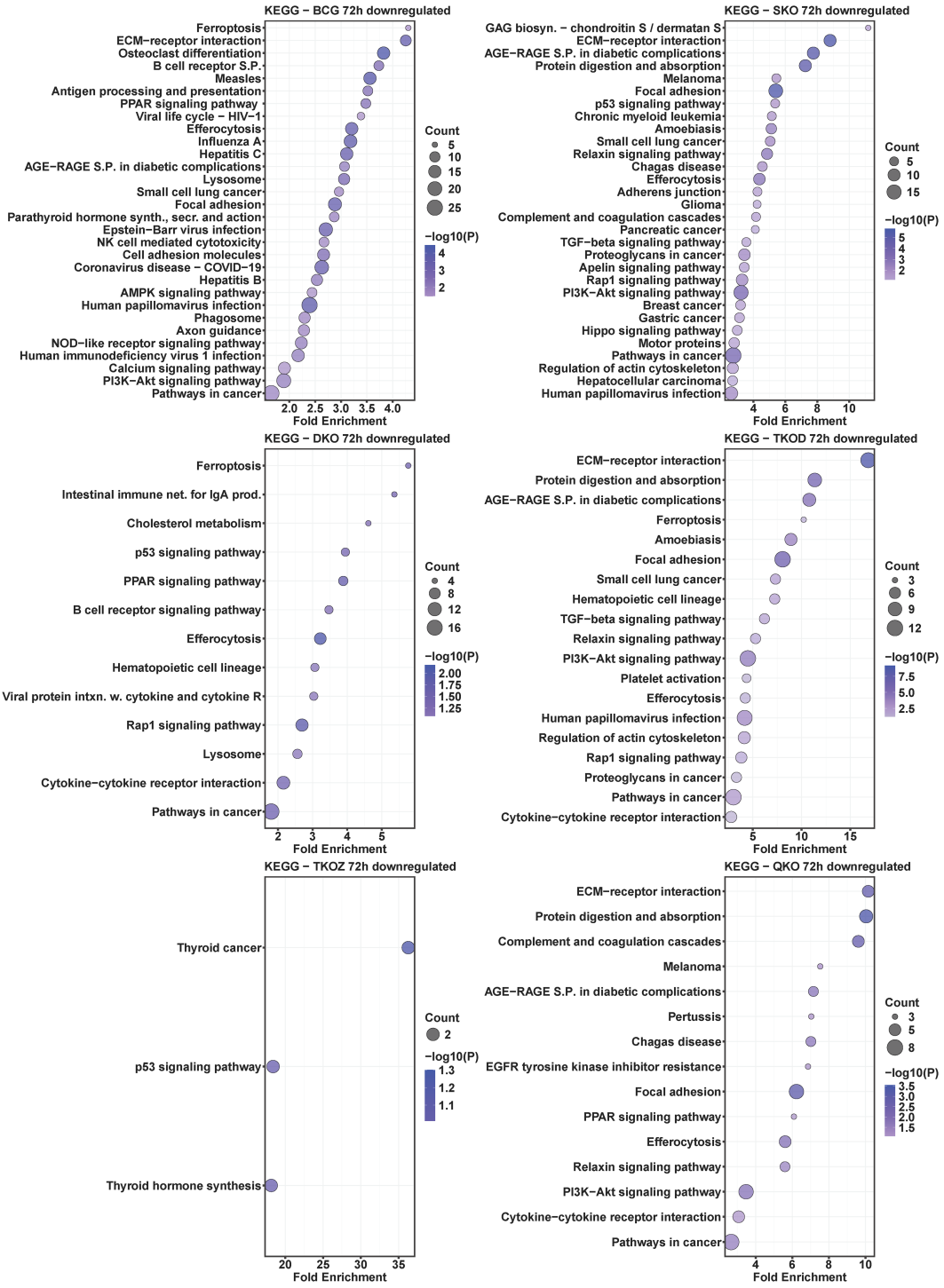


FIGURE 7
KEGG pathway analysis in downregulated differentially expressed transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 72 h post-infection. Dot plots illustrate the top 30 enriched pathways in downregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each category, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

with cytokine and cytokine receptor, IL-17 signaling pathway, hematopoietic cell lineage, TNF signaling pathway, NF-kappa B signaling pathway, and cytokine-cytokine receptor interaction were uniquely enriched in upregulated DEGs in SKO, DKO, and TKO-D; almost no pathways matched with BCG for the TKO-Z and QKO groups. Interestingly, at 72 h, nearly all the pathways were uniquely enriched with the upregulated DEGs in our Mtb-derived vaccine groups compared to BCG. In examining the pathways enriched in downregulated DEGs, very few overlapped at 72 h, and almost none at 24 h (Supplementary Figure S7). Notably, the TKO-Z group did not exhibit any common pathways overrepresented in downregulated DEGs like BCG, in contrast to many that were enriched in upregulated DEGs in this group during 72 h. The-common KEGG pathways enriched in downregulated DEGs across all groups appeared primarily due to the SKO strain (Supplementary Figure S7).

Gene ontology analysis of mouse BMDMs infected with mycobacterial vaccine strains compared to H37Rv

To investigate altered biological processes by DEGs, we performed GO analysis for biological processes for all vaccine groups compared to H37Rv. Biological processes, such as neutrophil chemotaxis, positive regulation of interleukin-6 production, inflammatory response, and immune response, were enriched with upregulated DEGs at the 24 h time point in the BCG, SKO, and TKO-D groups (Figure 8). Similar to the KEGG pathway analysis, the TKO-Z and QKO groups exhibited delayed enrichment of some common pathways in upregulated DEGs to other vaccine groups, primarily at the 72 h time point (Figure 9). At 72 h, biological processes, including neutrophil chemotaxis, positive regulation of interferon-gamma production, cytokine-mediated signaling pathway, response to lipopolysaccharide, inflammatory response, cellular response to lipopolysaccharide, negative regulation of cell proliferation, positive regulation of the ERK1 and ERK2 cascade, immune system process, immune response, and response to xenobiotic stimulus, were consistently overrepresented in the upregulated DEGs across all vaccine groups (Figure 9).

Similar to the KEGG pathway analysis, we observed a comparable pattern here, with a greater number of the enriched biological processes overlapping in the upregulated DEGs than in the downregulated DEGs. The TKO-Z group was particularly distinct, exhibiting overrepresented biological processes such as the positive regulation of endothelial cell proliferation and cell adhesion, in downregulated DEGs. Interestingly, most regulation of pathways appears to be linked to the deletion of *fbpA*, as indicated by the downregulated DEGs-enriched processes seen in the SKO group. These include the phospholipase C-activating G-protein coupled receptor signaling pathway, positive regulation of angiogenesis, positive regulation of cytosolic calcium ion concentration, response to hypoxia, gene expression, inflammatory response, positive regulation of transcription from the RNA polymerase II promoter, response to dietary excess, positive regulation of stress fiber assembly, immune system processes, positive regulation of the MAPK cascade, positive regulation of the ERK1 and ERK2 cascade, and cell adhesion (Figures 10, 11).

Key immune and cell signaling pathways differentially regulated among vaccine groups

KEGG pathway analysis was performed on immune and cell signaling pathways to investigate the mechanisms underlying vaccine-induced immune responses in macrophages infected with various vaccine strains (Supplementary Data 5, 6). Importantly, gene deletions in our vaccine strains resulted in differential modulation of several signal transduction pathways, including the NF-kappa B signaling pathway (mmu04064) and TNF signaling pathway (mmu04668). The signaling molecule and interaction pathway, particularly Cytokine-cytokine receptor interaction (mmu04060), showed significant upregulation, while the ECM-receptor interaction pathway (mmu04512) demonstrated downregulation. Several cytokine and chemokine genes, including *tnf, il6, il1α, il1β, il1r1, il1r2, il12a, il12b, il23, cxcl1, cxcl2, cxcl3, ccl22, ccl2, ccl3, ccl4, ccl6, and ccl7*, were significantly upregulated across most vaccine groups. In contrast, genes such as *cxcr1, cxcr3, cxcl9, cxcl12, and ccl8* were downregulated in the majority of vaccine groups. Furthermore, as previously noted, the ECM-receptor interaction pathway was significantly enriched in downregulated DEGs in specific vaccine strain-infected BMDMs, including SKO, TKO-D, and QKO, at 72 h (Figure 7).

Key immune system pathways were also affected, including Hematopoietic cell lineage (mmu04640), Chemokine signaling pathway (mmu04062), Toll-like receptor signaling pathway (mmu04620), IL-17 signaling pathway (mmu04657), Th1 and Th2 cell differentiation (mmu04658), Th17 cell differentiation (mmu04659), and T cell receptor signaling pathway (mmu04660) (Supplementary Data 5, 6). Notably, only a small number of genes in the B cell receptor signaling pathway (mmu04662) were differentially regulated across all vaccine strains, including BCG.

Among these pathways, the IL-17 signaling pathway exhibited the most pronounced differential regulation of DEGs across all vaccine groups. Heatmap analysis confirmed the list of genes with differential expression within the IL-17 signaling pathway (Figure 12). A few of the upregulated genes, such as *csf2, csf3, il1β, ptgs2, and lcn2*, were further confirmed by qRT-PCR, corroborating the transcriptome findings (Figure 12).

Further confirmation of DEGs in vaccine-infected BMDMs using qRT-PCR

To further confirm the findings of our study, we performed qRT-PCR on macrophages infected with our vaccine strains. For this analysis, we randomly selected DEGs from various pathways. We examined genes such as *csf1, TNF, slc7a2, lta, ddit4*, and *dapk2*

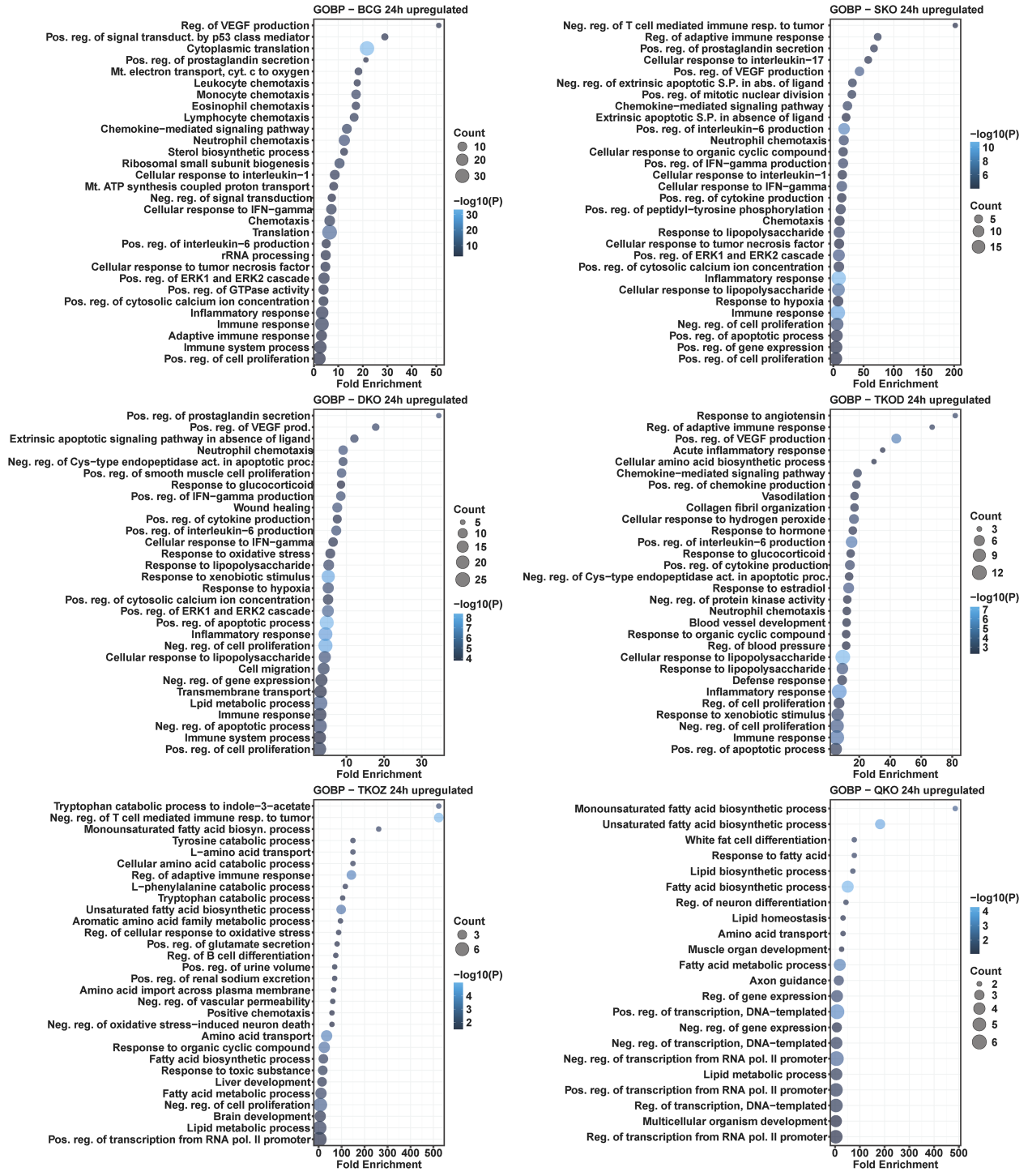


FIGURE 8
Gene ontology analysis for biological processes in the differentially upregulated transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 24 h post-infection. Dot plots illustrate the top 30 enriched biological processes in upregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where dot size reflects the total number of genes in each biological process, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

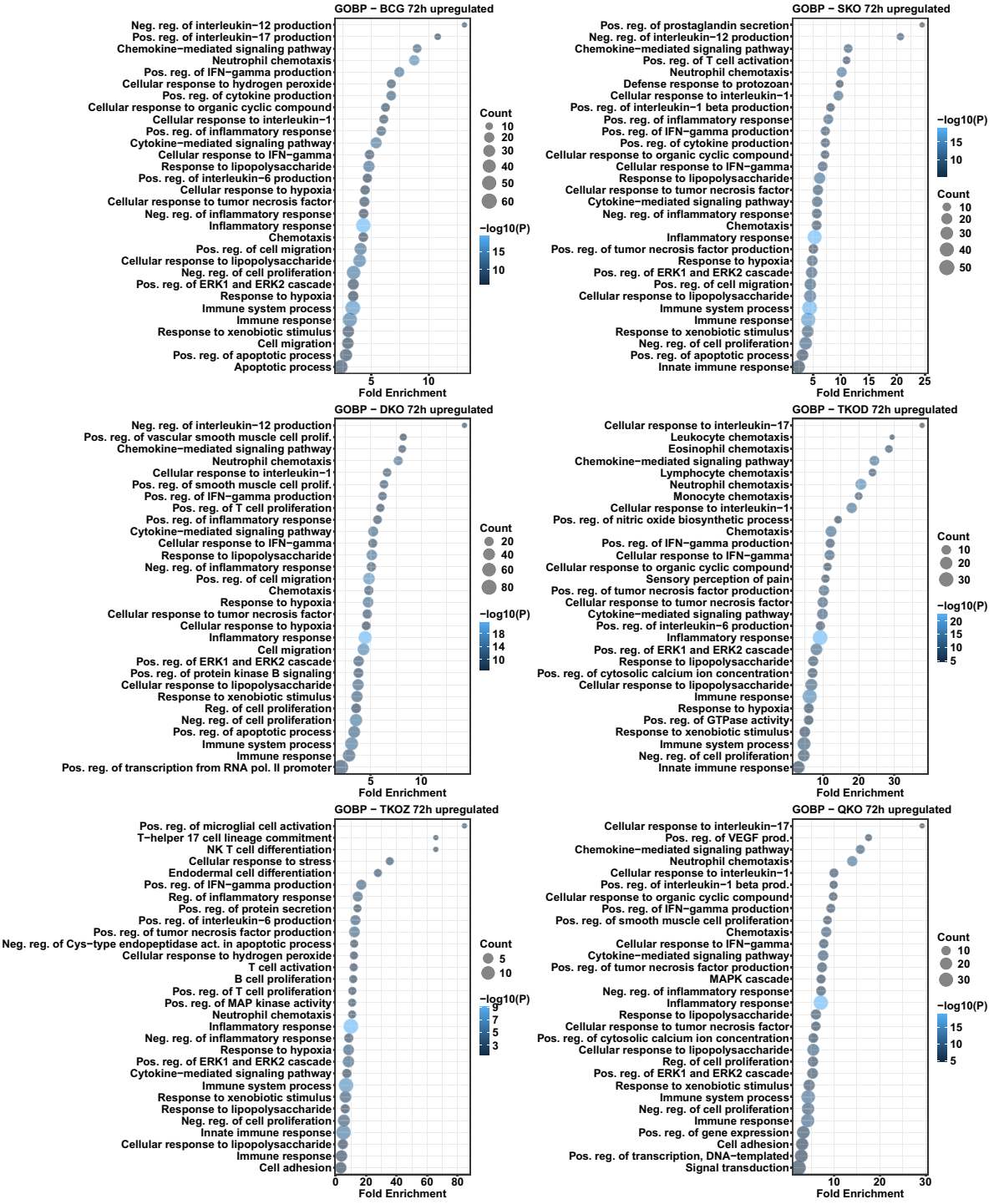


FIGURE 9
Gene ontology analysis for biological processes in the differentially upregulated transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 72 h post-infection. Dot plots illustrate the top 30 enriched biological processes in upregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each biological process, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

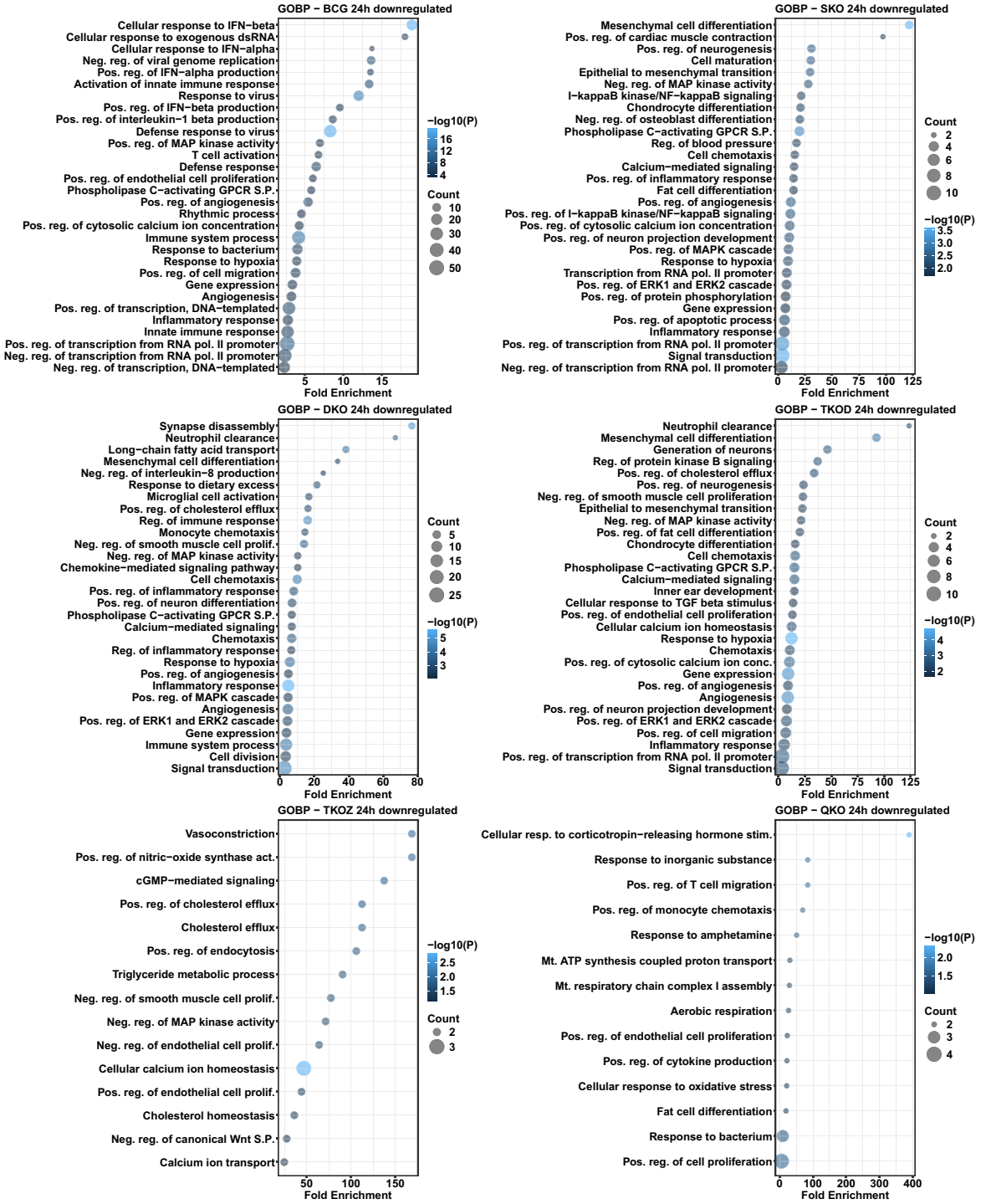


FIGURE 10
Gene ontology analysis for biological processes in the differentially downregulated transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 24 h post-infection. Dot plots illustrate the top 30 enriched biological processes of downregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each biological process, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

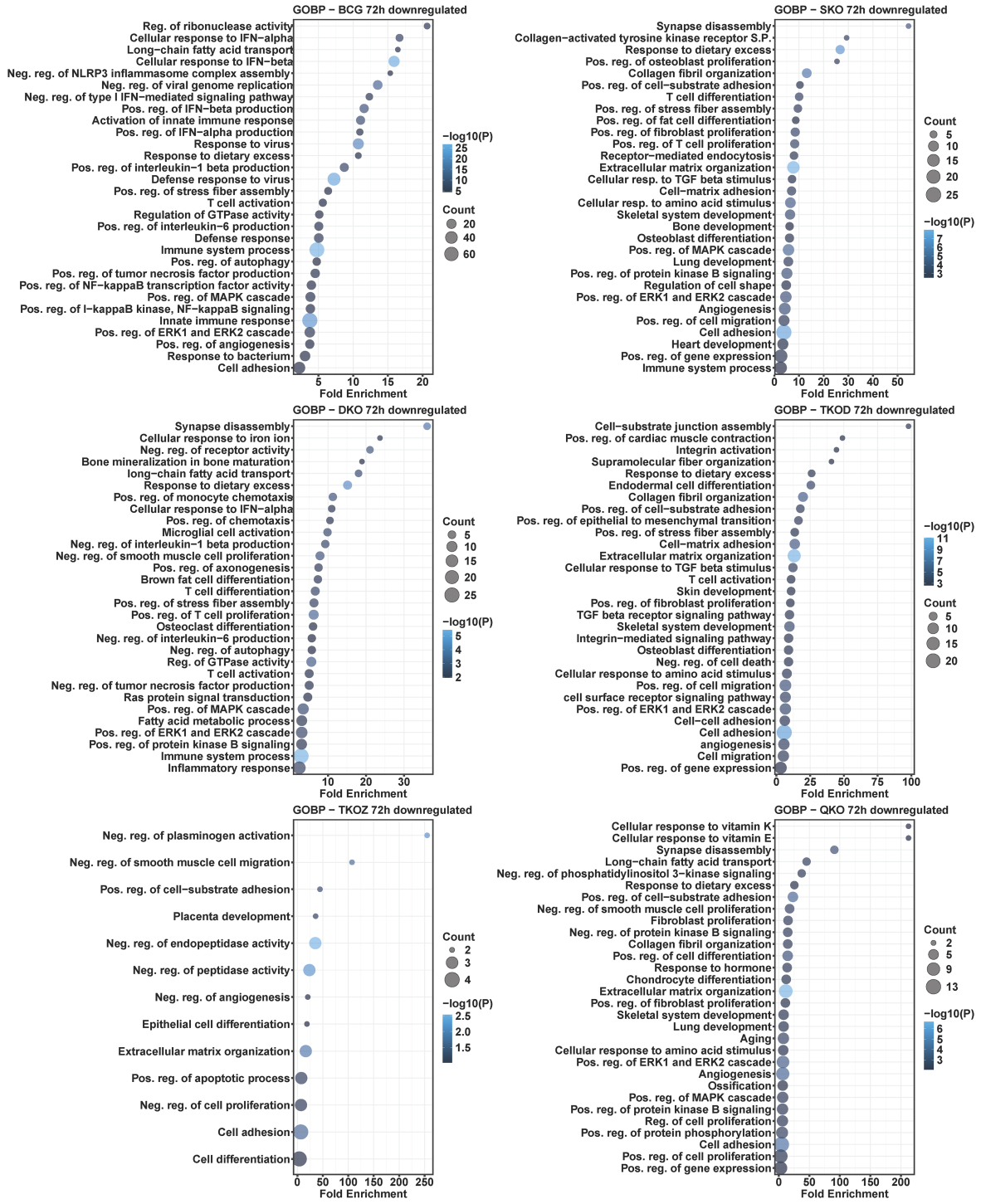


FIGURE 11
Gene ontology analysis for biological processes in the differentially downregulated transcripts of mouse BMDMs infected with vaccine strains versus H37Rv at 72 h post-infection. Dot plots illustrate the top 30 enriched biological processes in downregulated DEGs in BCG, SKO, DKO, TKOD, TKOZ, and QKO compared to H37Rv control. Dot plots measure fold enrichment, where the dot size reflects the total number of genes in each biological process, and the gradient color indicates statistical significance expressed as $-\log_{10}$ (P).

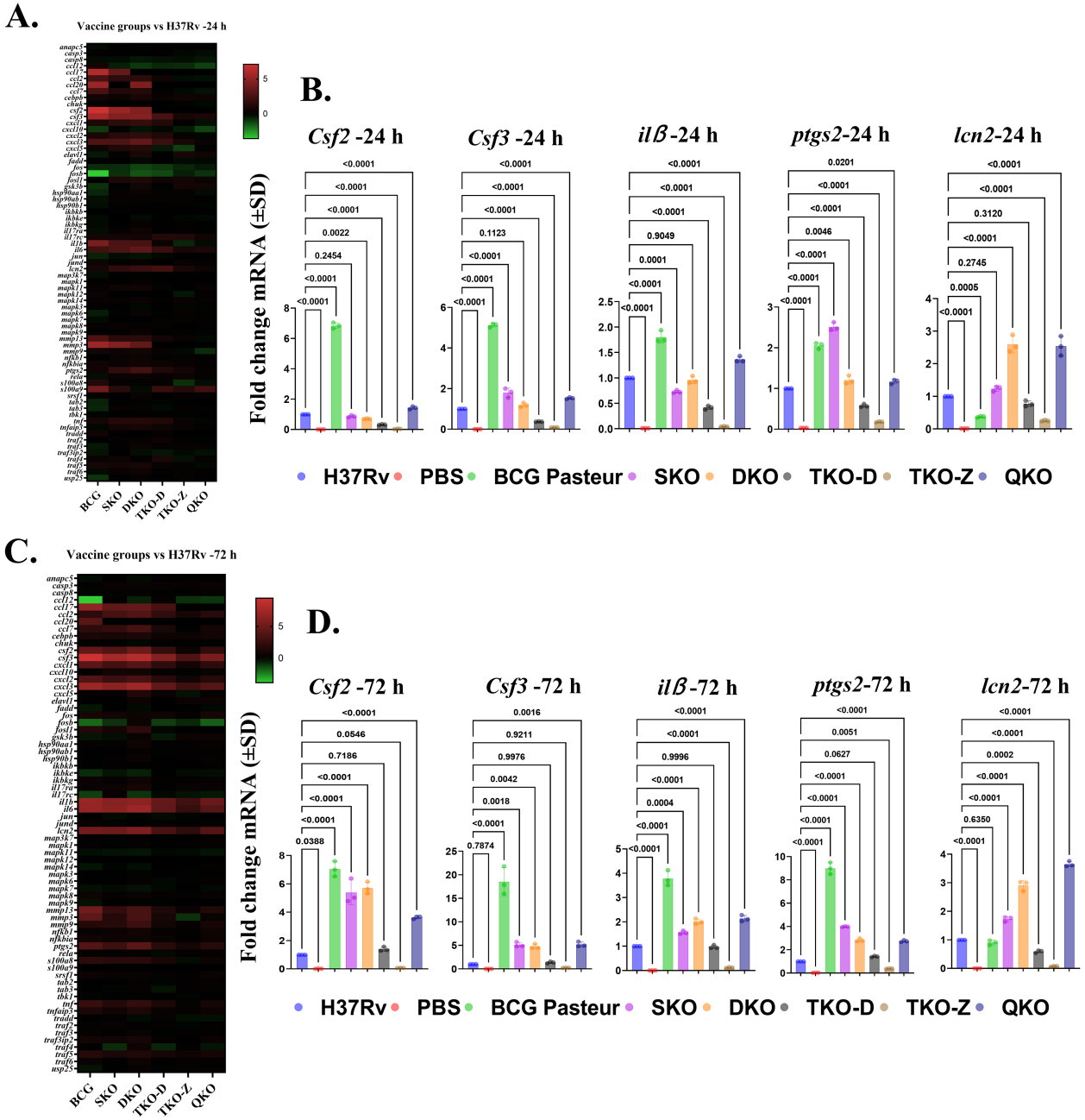


FIGURE 12
IL17 signaling pathway is differentially regulated in vaccine groups in comparison with H37Rv. **(A, C)** Heat map depicts the green–red gradient that reflects relative gene expression among vaccine groups at 24 h and 72 h. **(B, D)** Gene expression levels of Csf2, Csf3, ilß, ptgs2, and lcn2 in BMDMs infected with vaccine strains at 24 and 72 hours as determined by qRT-PCR. Data were analyzed by one-way ANOVA followed by Dunnett's multiple comparisons test.

at the 24 h and 72 h time points (Figure 13). The results demonstrated strong alignment with the transcriptome data. Interestingly, we observed an increased expression of the solute carrier family 7-member 2 (*slc7a2*) gene across all our vaccine groups in both transcriptome and qPCR analyses (Figure 13). Notably, this gene is reported to be highly expressed in macrophages infected with avirulent Mtb strain H37Ra (25), indicating that our vaccine strains exhibit reduced virulence compared to the wild-type strain.

Discussion

In this study, we performed genome-wide transcriptome analyses of mouse macrophages after infection with our Mtb-derived vaccine strains. We also included the BCG vaccine, as it is an established vaccine against TB. While multiple studies have reported transcriptome data of mouse macrophages infected with either BCG or Mtb, (25–27, 30, 31); our study focused on vaccine strains deficient in genes such as *fbpA*, *sapM*, *zmp1*, and *dosR*, either

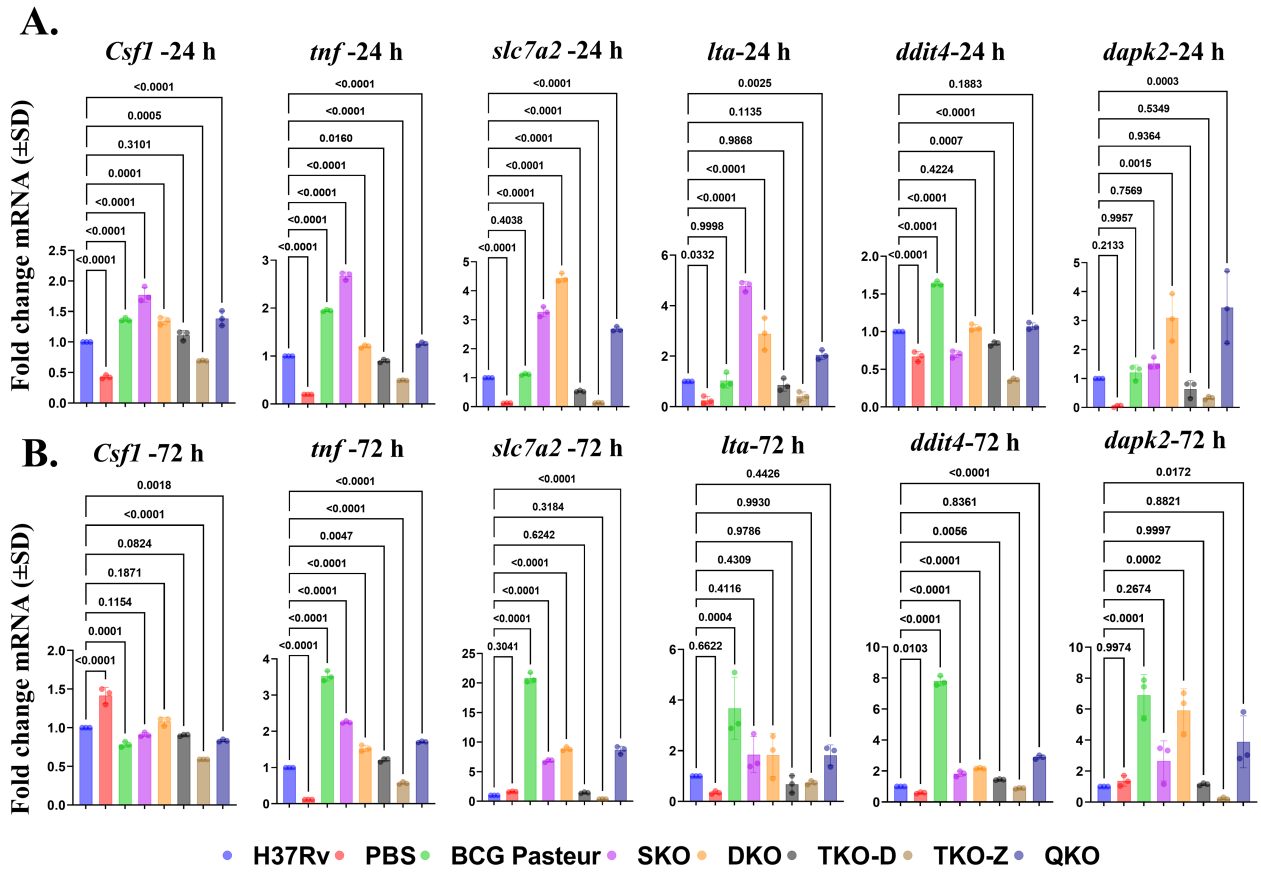


FIGURE 13
qPCR Validation of differentially regulated pathways in vaccine groups comparison with H37Rv. **(A, B)** Gene expression levels of Csf1, tnf, slc7a2, lta, ddit4, and dapk2 in vaccine strains infected BMDMs at 24 h and 72 h as determined by qRT-PCR. Data were analyzed by one-way ANOVA followed by Dunnett's multiple comparisons test.

individually or in combination. Transcriptomic analysis of our vaccine-infected macrophages was conducted using a heatmap, volcano plot, and venn diagram, demonstrating a distinct difference in DEGs expressed among groups. The number of DEGs in BCG-infected macrophages was greater than in macrophages infected with other vaccine groups and Mtb H37Rv. This disparate response between BCG and our Mtb H37Rv-derived strains may be due to the deletion of several ORFs within RD1-RD16 regions in BCG.

KEGG pathway analysis of DEGs identified several immune pathways that are implicated in vaccine-infected macrophages. Some prominent pathways include infections with intracellular bacteria, as well as immune, viral, cancer, and disease-related components. Type I interferon-related pathways, such as cytosolic DNA-sensing, NOD-like receptor signaling, NF-kappa B signaling, and C-type lectin receptor signaling, were significantly enriched with upregulated DEGs in our vaccine strains compared to the naïve group, similar to the previous reports (30, 32, 33). Also, consistent with earlier reports, TNF signaling was enriched with upregulated DEGs in all vaccine groups (25, 34). Notably, pathways such as the cell cycle, DNA replication, p53 signaling, progesterone-mediated oocyte maturation, focal adhesion, and efferocytosis were enriched with downregulated DEGs in all vaccine groups, with some of the pathways aligning with earlier findings (30, 35).

Significant variations in upregulated DEGs-enriched pathways were noted across the vaccine groups between 24 and 72 h. Notably, TKO-Z and QKO exhibited a few pathways unique to other vaccine strains at 24 h, but by 72 h, they aligned with strains like SKO, DKO, and TKO-D. This change in pathways in TKO-Z and QKO may result from the deletion of zmp1. The zmp1 gene encodes a crucial enzyme for *M. tuberculosis* pathogenicity, playing various roles such as inhibiting phagosome maturation, suppressing inflammasome activation, mediating necrosis, and providing protection in guinea pig models (20, 21, 36). Interestingly, intracellular bacterial pathways, including legionellosis, leishmaniasis, and tuberculosis, were not enriched with upregulated DEGs in the TKO-Z group compared to H37Rv at either time point. However, these pathways were significantly enriched in the TKO-Z group when compared to the uninfected control. Due to the lack of transcriptome data for a *zmp1* mutant infected macrophages, direct comparisons with previous findings cannot be made. In comparison to H37Rv, the upregulation of genes in the H37Rv-derived vaccine groups ranged from approximately 0.2 -6.1%, while the downregulated DEGs showed minimal changes, with only about 0.1-2.9% of genes

being downregulated. This highlights the need for transcriptome data from Mtb-derived vaccine candidates currently in clinical trials for a more comprehensive comparison. Similar to BCG, pathways such as ECM receptor interaction, efferocytosis, focal adhesion, and PI3K-Akt signaling were enriched with downregulated DEGs in some vaccine groups in our study.

Despite developing numerous vaccines against TB, we still lack exact knowledge of the immune correlates of protection (CoPs) for TB. However, data from animal and human studies provide insights into the immune cells that may be crucial for controlling TB, including Th1, Th17, CD8+ T cells, B cells, tissue-resident memory T cells, trained immunity, and tissue-resident alveolar macrophages (37). Interestingly, genes from the major immune and cell signaling pathways like Cytokine-cytokine receptor interaction, Chemokine signaling, NF-kappa B signaling pathway, Toll-like receptor signaling, IL-17 signaling, Th1 and Th2 cell differentiation, Th17 cell differentiation, T cell receptor signaling, and TNF signaling pathway were differentially regulated in our vaccine strains compared to wildtype H37Rv. Numerous vaccine studies have underscored the crucial role of T cell-mediated protection against Mtb infection (19, 38–41) and the limited significance of B cell-mediated responses in TB vaccines (42). Similarly, our vaccines mainly boost T cell-mediated immune pathways instead of B cell receptor signaling pathways. These findings offer a hopeful perspective for the development of more effective TB vaccines.

Cytokines and chemokines are essential in coordinating the immune response to mycobacterial infection (43). TB vaccine candidates like VPM1002 and MTBVAC have shown increased cytokine responses (44, 45). Our vaccine strains similarly showed different cytokine and chemokine expression profiles. Pro-inflammatory cytokines, including TNF-α, IL-6, GM-CSF, and IL-1, are crucial for the immune response against Mtb infection and play a vital role in host survival (46–48). Consistent with previous findings, our vaccine-infected macrophages exhibit increased expression of TNF-α, IL-6, and GM-CSF. Furthermore, our vaccine strains induce higher levels of IL-1 family cytokines, including IL-1α, IL-1β, IL-1R1, and IL-1R2. Chemokines such as CXCL1, CXCL2, and CXCL3 promote the recruitment of neutrophils and natural killer cells, while CCL3 and CCL4 aid in T-cell recruitment. Additionally, CCL7 is vital for recruiting monocytes, dendritic cells, T cells, and natural killer cells (49). Notably, our vaccine strains show strong expression of these chemokines, highlighting their potential immunomodulatory role effects.

Emerging evidence underscores the essential role of IL-17 in TB control across various species, including mice (50, 51), non-human primates (52, 53), and humans (54). Initially, IL-17 was thought to primarily mediate responses against extracellular pathogens rather than intracellular bacteria like Mtb. However, recent findings underscore its essential role in TB control. Studies have revealed that IL-17 levels are significantly lower in individuals who progress to active TB compared to non-progressors (55). Reports indicate that CD4+ T cells producing IL-17 are primarily localized in the lungs compared to TNF-α and IL-2. Furthermore, administering exogenous IL-17 in human granuloma models has shown effectiveness in controlling Mtb (56). Moreover, IL-17 has been identified as essential in mice for providing early protective immunity against Mtb HN878 infection (51). Mice that lack IL-17 receptors show reduced long-term control of Mtb infection (57). In our study, IL-17 signaling was significantly upregulated across all vaccine strains, including SKO, DKO, TKO-D, TKO-Z, and QKO. Gene ontology analysis revealed upregulated cellular responses to IL-17 in TKO-D and QKO, T-helper 17 cell lineage commitment in TKO-Z, and positive regulation of IL-17 production in the BCG vaccine group. These findings underscore the pivotal role of IL-17 in TB control and the effectiveness of our vaccine strains in eliciting an appropriate immune response.

Recent studies have reported that the upregulation of *slc7a2* in macrophages plays a critical role in controlling the intracellular survival of Mtb (25). Notably, *slc7a2* expression is higher in macrophages infected with the avirulent strain H37Ra compared to the virulent H37Rv strain. Consistent with these findings, our DEGs analysis revealed increased expression of *slc7a2* transcripts in our vaccine strains compared to the wild-type H37Rv, a result further validated through qPCR. While safety studies in SCID mice are still required to establish the safety profile of our vaccine strains, these findings suggest an improved safety profile for the vaccine strains used in this study.

One of the major limitations of the present study is that the comparative transcriptomic analysis was performed under *in vitro* conditions and not *in vivo*. While our experimental design provides us with the controlled environment to study BMDMs' responses after infection with our vaccine strains, it lacks the *in vivo* conditions like interactions with other cell types, location-specific cell signals, etc. However, our study offers valuable comparative transcriptomic analysis datasets among our vaccine strains along with BCG, which offer insights that help enhance our understanding. This study focuses exclusively on comparing the transcriptomes of vaccine strains derived from the H37Rv Mtb strain. Further research is required to understand the relationship between the immune and cell signaling pathways activated by these vaccines and their actual protective efficacy. Moreover, this study emphasizes the importance of performing comparative transcriptomic analyses for vaccine candidates such as VPM1002 and MTBVAC, currently undergoing clinical trials, to gain deeper insights into the host immune response. Simultaneously, we recognize the importance of 'decoy' immune responses in TB infection (58). While certain host immune responses may appear promising, they indeed support the pathogen by promoting its persistence within the host. Thus, we strongly underscore the importance of performing protection studies in animal models and correlating these immune responses to actual protection, rather than relying solely on the statement that heightened proinflammatory cytokine production alone is beneficial. Overall, our study provides a thorough comparative transcriptome analysis of Mtb-derived vaccine strains alongside BCG, highlighting key immune pathways that

play a crucial role in modulating immune and cell signaling events in the fight against the Mtb pathogen.

Data availability statement

The new RNA-seq genomic datasets presented here can be accessed from the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) with the following accession number: GSE290804.

Ethics statement

The animal study was approved by The Institutional Animal Care and Use Committee (IACUC) of the Texas Tech University Health Sciences Center El Paso. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

RV: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. BY: Data curation, Formal analysis, Methodology, Software, Validation, Writing – review & editing. AC: Methodology, Validation, Writing – review & editing. MS: Methodology, Validation, Writing – review & editing. VR: Formal analysis, Software, Writing – review & editing. RJ: Methodology, Writing – review & editing. JC: Methodology, Writing – review & editing. CJ: Supervision, Writing – review & editing. ER: Data curation, Formal Analysis, Methodology, Software, Writing – review & editing. SG: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. SD: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by NIH 1R15AI156647 and NIH 1RO1AI175837 grants.

Acknowledgments

The authors thank SD and SG lab members for their careful review and helpful suggestions on this work. SG was supported by a First-time faculty recruitment award from the Cancer Prevention and Research Institute of Texas (CPRIT; RR170020). SG is also supported by the American Cancer Society (RSG-22-170-01-RMC), NIH 1R16GM149497-01, and CPRIT-TREC (RP230420) grants. Figure 1 was created using www.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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1583439/full#supplementary-material

References

1. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet*. (2013) 45:1176–82. doi: 10.1038/ng.2744

2. Brites D, Gagneux S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunol Rev*. (2015) 264:6–24. doi: 10.1111/imr.2015.264.issue-1

3. World Health Organization. *Global tuberculosis report 2024*. Geneva, Switzerland (2024).

4. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J Prev Med Hyg*. (2017) 58: E9–e12.

5. Woodruff HB. Selman A. Waksman, winner of the 1952 Nobel Prize for physiology or medicine. *Appl Environ Microbiol*. (2014) 80:2–8. doi: 10.1128/AEM.01143-13

6. Dockrell HM, Smith SG. What have we learnt about BCG vaccination in the last 20 Years? *Front Immunol*. (2017) 8:1134. doi: 10.3389/fimmu.2017.01134

7. World Health Organization. *Global tuberculosis report 2022*. Geneva, Switzerland (2023).

8. World Health Organization. *The End TB Strategy*. Geneva (2015).

9. Zhuang L, Ye Z, Li L, Yang L, Gong W. Next-generation TB vaccines: progress, challenges, and prospects. *Vaccines (Basel)*. (2023) 11(8):1304. doi: 10.3390/vaccines11081304

10. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*. (1999) 284:1520–3. doi: 10.1126/science.284.5419.1520

11. Jungblut PR, Müller EC, Mattow J, Kaufmann SH. Proteomics reveals open reading frames in Mycobacterium tuberculosis H37Rv not predicted by genomics. *Infect Immun*. (2001) 69:5905–7. doi: 10.1128/IAI.69.9.5905-5907.2001

12. Veerapandian R, Gadad SS, Jagannath C, Dhandayuthapani S. Live attenuated vaccines against tuberculosis: targeting the disruption of genes encoding the secretory proteins of mycobacteria. *Vaccines (Basel)*. (2024) 12(5):530. doi: 10.3390/vaccines12050530

13. Copenhaver RH, Sepulveda E, Armitige LY, Actor JK, Wanger A, Norris SJ, et al. A mutant of Mycobacterium tuberculosis H37Rv that lacks expression of antigen 85A is attenuated in mice but retains vaccinogenic potential. *Infect Immun*. (2004) 72:7084–95. doi: 10.1128/IAI.72.12.7084-7095.2004

14. Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS. Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. *Science*. (1997) 276:1420–2. doi: 10.1126/science.276.5317.1420

15. Armitige LY, Jagannath C, Wanger AR, Norris SJ. Disruption of the genes encoding antigen 85A and antigen 85B of Mycobacterium tuberculosis H37Rv: effect on growth in culture and in macrophages. *Infect Immun*. (2000) 68:767–78. doi: 10.1128/IAI.68.2.767-778.2000

16. Saikolappan S, Estrella J, Sasindran SJ, Khan A, Armitige LY, Jagannath C, et al. The fbpA/sapM double knock out strain of Mycobacterium tuberculosis is highly attenuated and immunogenic in macrophages. *PLoS One*. (2012) 7:e36198. doi: 10.1371/journal.pone.0036198

17. Saleh MT, Belisle JT. Secretion of an acid phosphatase (SapM) by Mycobacterium tuberculosis that is similar to eukaryotic acid phosphatases. *J Bacteriol*. (2000) 182:6850–3. doi: 10.1128/JB.182.23.6850-6853.2000

18. Vergne I, Chua J, Lee HH, Lucas M, Belisle J, Deretic V. Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*. (2005) 102:4033–8. doi: 10.1073/pnas.0409716102

19. Mishra A, Khan A, Singh VK, Glyde E, Saikolappan S, Garnica O, et al. The ΔfbpAΔsapM candidate vaccine derived from Mycobacterium tuberculosis H37Rv is markedly immunogenic in macrophages and induces robust immunity to tuberculosis in mice. *Front Immunol*. (2024) 15. doi: 10.3389/fimmu.2024.1321657

20. Master SS, Rampini SK, Davis AS, Keller C, Ehlers S, Springer B, et al. Mycobacterium tuberculosis prevents inflammasome activation. *Cell Host Microbe*. (2008) 3:224–32. doi: 10.1016/j.chom.2008.03.003

21. Sander P, Clark S, Petrera A, Vilaplana C, Meuli M, Selchow P, et al. Deletion of zmp1 improves Mycobacterium bovis BCG-mediated protection in a Guinea pig model of tuberculosis. *Vaccine*. (2015) 33:1353–9. doi: 10.1016/j.vaccine.2015.01.058

22. Hu Y, Movahedzadeh F, Stoker NG, Coates AR. Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in *vivo*. *Infect Immun*. (2006) 74:861–8. doi: 10.1128/IAI.74.2.861-868.2006

23. Roupie V, Romano M, Zhang L, Korf H, Lin MY, Franken KL, et al. Immunogenicity of eight dormancy regulon-encoded proteins of Mycobacterium tuberculosis in DNA-vaccinated and tuberculosis-infected mice. *Infect Immun*. (2007) 75:941–9. doi: 10.1128/IAI.01137-06

24. Garnica O, Das K, Ouellet H, Caad B, Dhandayuthapani S. Genetically altered Mycobacteriunm tuberculosis vaccine strains show increased antigen presentation In: *Tuberculosis: Translating Scientific Findidngs for Clinical and Public Health Impact*. Keystone Symposia, Whistler, BC, Canada (2018).

25. Lee J, Lee SG, Kim KK, Lim YJ, Choi JA, Cho SN, et al. Characterisation of genes differentially expressed in macrophages by virulent and attenuated Mycobacterium tuberculosis through RNA-Seq analysis. *Sci Rep*. (2019) 9:4027. doi: 10.1038/s41598-019-40814-0

26. Pisu D, Huang L, Grenier JK, Russell DG. Dual RNA-Seq of Mtb-infected macrophages *in vivo* reveals ontologically distinct host-pathogen interactions. *Cell Rep*. (2020) 30:335–50.e334. doi: 10.1016/j.celrep.2019.12.033

27. Pu W, Zhao C, Wazir J, Su Z, Niu M, Song S, et al. Comparative transcriptomic analysis of THP-1-derived macrophages infected with Mycobacterium tuberculosis H37Rv, H37Ra and BCG. *J Cell Mol Med*. (2021) 25:10504–20. doi: 10.1111/jcmm.v25.22

28. Veerapandian R, Ramos EI, Vijayaraghavan M, Sedano MJ, Carmona A, Chacon JA, et al. Mycobacterium smegmatis secreting methionine sulfoxide reductase A (MsrA) modulates cellular processes in mouse macrophages. *Biochimie*. (2023) 211:1–15. doi: 10.1016/j.biochi.2023.02.010

29. Ramos EI, Das K, Harrison AL, Garcia A, Gadad SS, Dhandayuthapani S. Mycoplasma genitalium and M. pneumoniae Regulate a Distinct Set of Protein-Coding Genes in Epithelial Cells. *Front Immunol*. (2021) 12:738431. doi: 10.3389/fimmu.2021.738431

30. Ding Y, Bei C, Xue Q, Niu L, Tong J, Chen Y, et al. Transcriptomic analysis of mycobacterial infected macrophages reveals a high MOI specific type I IFN signaling. *Infect Immun*. (2023) 91:e0015523. doi: 10.1128/iai.00155-23

31. Yabaji SM, Rukhlenko OS, Chatterjee S, Bhattacharya B, Wood E, Kasaikina M, et al. Cell state transition analysis identifies interventions that improve control of Mycobacterium tuberculosis infection by susceptible macrophages. *Sci Adv*. (2023) 9:eadh4119. doi: 10.1126/sciadv.adh4119

32. Wiens KE, Ernst JD. The mechanism for type I interferon induction by mycobacterium tuberculosis is bacterial strain-dependent. *PloS Pathog*. (2016) 12:e1005809. doi: 10.1371/journal.ppat.1005809

33. Jiang H, Tsang L, Wang H, Liu C. IFI44L as a forward regulator enhancing host antituberculosis responses. *J Immunol Res*. (2021) 2021:5599408. doi: 10.1155/2021/5599408

34. Hinman AE, Jani C, Pringle SC, Zhang WR, Jain N, Martinot AJ, et al. Mycobacterium tuberculosis canonical virulence factors interfere with a late component of the TLR2 response. *Elife*. (2021) 10:e73984. doi: 10.7554/eLife.73984

35. Li P, Li Y, Wang CC, Xia LG. Comparative transcriptomics reveals common and strain-specific responses of human macrophages to infection with Mycobacterium tuberculosis and Mycobacterium bovis BCG. *Microb Pathog*. (2024) 189:106593. doi: 10.1016/j.micpath.2024.106593

36. Johansen P, Fettelschoss A, Amstutz B, Selchow P, Waeckerle-Men Y, Keller P, et al. Relief from Zmp1-mediated arrest of phagosome maturation is associated with facilitated presentation and enhanced immunogenicity of mycobacterial antigens. *Clin Vaccine Immunol*. (2011) 18:907–13. doi: 10.1128/CVI.00015-11

37. Wang J, Fan XY, Hu Z. Immune correlates of protection as a game changer in tuberculosis vaccine development. *NPJ Vaccines*. (2024) 9:208. doi: 10.1038/s41541-024-01004-w

38. Henao-Tamayo M, Junqueira-Kipnis AP, Ordway D, Gonzales-Juarrero M, Stewart GR, Young DB, et al. A mutant of Mycobacterium tuberculosis lacking the 19-kDa lipoprotein Rv3763 is highly attenuated *in vivo* but retains potent vaccinogenic properties. *Vaccine*. (2007) 25:7153–9. doi: 10.1016/j.vaccine.2007.07.042

39. Roche CM, Smith A, Lindsey DR, Meher A, Schluns K, Arora A, et al. The ΔfbpA attenuated candidate vaccine from Mycobacterium tuberculosis, H37Rv primes for a stronger T-bet dependent Th1 immunity in mice. *Tuberculosis (Edinb)*. (2011) 91 Suppl 1:S96–104. doi: 10.1016/j.tube.2011.10.018

40. Martinot AJ, Blass E, Yu J, Aid M, Mahrokhian SH, Cohen SB, et al. Protective efficacy of an attenuated Mtb ΔLprG vaccine in mice. *PloS Pathog*. (2020) 16:e1009096. doi: 10.1371/journal.ppat.1009096

41. Liu X, Li H, Li S, Yuan J, Pang Y. Maintenance and recall of memory T cell populations against tuberculosis: Implications for vaccine design. *Front Immunol*. (2023) 14:1100741. doi: 10.3389/fimmu.2023.1100741

42. Ashhurst AS, Parumasivam T, Chan JGY, Lin LCW, Flórido M, West NP, et al. PLGA particulate subunit tuberculosis vaccines promote humoral and Th17 responses but do not enhance control of Mycobacterium tuberculosis infection. *PloS One*. (2018) 13:e0194620. doi: 10.1371/journal.pone.0194620

43. Domingo-Gonzalez R, Prince O, Cooper A, Khader SA. Cytokines and chemokines in mycobacterium tuberculosis infection. *Microbiol Spectr*. (2016) 4(5). doi: 10.1128/microbiolspec.TBTB2-0018-2016

44. Grode L, Ganoza CA, Brohm C, Weiner J3rd, Eisele B, Kaufmann SH. Safety and immunogenicity of the recombinant BCG vaccine VPM1002 in a phase 1 open-label randomized clinical trial. *Vaccine*. (2013) 31:1340–8. doi: 10.1016/j.vaccine.2012.12.053

45. Dijkman K, Aguilo N, Boot C, Hofman SO, Sombroek CC, Vervenne R, et al. Pulmonary MTBVAC vaccination induces immune signatures previously correlated with prevention of tuberculosis infection. *Cell Rep Med*. (2021) 2:100187. doi: 10.1016/j.xcrm.2020.100187

46. Etna MP, Giacomini E, Severa M, Coccia EM. Pro- and anti-inflammatory cytokines in tuberculosis: a two-edged sword in TB pathogenesis. *Semin Immunol*. (2014) 26:543–51. doi: 10.1016/j.smim.2014.09.011

47. Vierboom MPM, Dijkman K, Sombroek CC, Hofman SO, Boot C, Vervenne R, et al. Stronger induction of trained immunity by mucosal BCG or MTBVAC vaccination compared to standard intradermal vaccination. *Cell Rep Med*. (2021) 2:100185. doi: 10.1016/j.xcrm.2020.100185

48. Mishra A, Singh VK, Jagannath C, Subbian S, Restrepo BI, Gauduin MC, et al. Human Macrophages Exhibit GM-CSF Dependent Restriction of Mycobacterium tuberculosis Infection via Regulating Their Self-Survival, Differentiation and Metabolism. *Front Immunol*. (2022) 13:859116. doi: 10.3389/fimmu.2022.859116

49. Monin L, Khader SA. Chemokines in tuberculosis: the good, the bad and the ugly. *Semin Immunol*. (2014) 26:552–8. doi: 10.1016/j.smim.2014.09.004

50. Khader SA, Bell GK, Pearl JE, Fountain JJ, Rangel-Moreno J, Cilley GE, et al. IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge. *Nat Immunol*. (2007) 8:369–77. doi: 10.1038/ni1449

51. Gopal R, Monin L, Slight S, Uche U, Blanchard E, Fallert Junecko BA, et al. Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection. *PLoS Pathog*. (2014) 10:e1004099. doi: 10.1371/journal.ppat.1004099

52. Gideon HP, Hughes TK, Tzouanas CN, Wadsworth MH 2nd, Tu AA, Gierahn TM, et al. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity*. (2022) 55:827–46.e810. doi: 10.1016/j.immuni.2022.04.004

53. Darrah PA, Zeppa JJ, Wang C, Irvine EB, Bucsan AN, Rodgers MA, et al. Airway T cells are a correlate of i.v. Bacille Calmette-Guerin-mediated protection against tuberculosis in rhesus macaques. *Cell Host Microbe*. (2023) 31:962–77.e968. doi: 10.1016/j.chom.2023.05.006

54. Ogongo P, Tran A, Marzan F, Gingrich D, Krone M, Aweeka F, et al. High-parameter phenotypic characterization reveals a subset of human Th17 cells that preferentially produce IL-17 against M. tuberculosis antigen. *Front Immunol*. (2024) 15:1378040. doi: 10.3389/fimmu.2024.1378040

55. Scriba TJ, Penn-Nicholson A, Shankar S, Hraha T, Thompson EG, Sterling D, et al. Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. *PLoS Pathog*. (2017) 13:e1006687. doi: 10.1371/journal.ppat.1006687

56. Ogongo P, Tezera LB, Ardain A, Nhamoyebonde S, Ramsuran D, Singh A, et al. Tissue-resident-like CD4+ T cells secreting IL-17 control Mycobacterium tuberculosis in the human lung. *J Clin Invest*. (2021) 131(10):e142014. doi: 10.1172/JCI142014

57. Freches D, Korf H, Denis O, Havaux X, Huygen K, Romano M. Mice genetically inactivated in interleukin-17A receptor are defective in long-term control of Mycobacterium tuberculosis infection. *Immunology*. (2013) 140:220–31. doi: 10.1111/imm.2013.140.issue-2

58. Ivanyi J. Tuberculosis vaccination needs to avoid 'decoy' immune reactions. *Tuberculosis (Edinb)*. (2021) 126:102021. doi: 10.1016/j.tube.2020.102021

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 25 July 2025
DOI 10.3389/fimmu.2025.1561572

OPEN ACCESS

EDITED BY
Dongqing Wei,
Shanghai Jiao Tong University, China

REVIEWED BY
Heling Bao,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Basem Fares,
Independent Researcher, Haifa, Israel

*CORRESPONDENCE
Yanling Meng
mylbycmu@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 16 January 2025
ACCEPTED 16 June 2025
PUBLISHED 25 July 2025

CITATION
Cai Q, Feng Y, Dong W and Meng Y (2025) Integrating bioinformatics to explore HPV-31 and HPV-52 E6/E7 proteins: from structural analysis to antigenic epitope prediction. *Front. Immunol.* 16:1561572. doi: 10.3389/fimmu.2025.1561572

Integrating bioinformatics to explore HPV-31 and HPV-52 E6/E7 proteins: from structural analysis to antigenic epitope prediction

Qixue Cai[1†], Yifan Feng[2†], Wenbo Dong[3†] and Yanling Meng[1*]

[1]Department of Pulmonary and Critical Care Medicine, Institute of Respiratory Disease, The First Hospital of China Medical University, Shenyang, Liaoning, China, [2]Department of Gastrointestinal Surgery, The First Hospital of China Medical University, Shenyang, Liaoning, China, [3]The First Clinical College, China Medical University, Shenyang, Liaoning, China

Introduction: Cervical cancer is the most common malignant neoplasm of the female reproductive tract. Infection with human papillomavirus (HPV) has been strongly associated with cervical cancer. Previous bioinformatics studies have examined the E6 and E7 proteins of high-risk HPV types; however, subtype-specific analyses for HPV-31 and HPV-52 remain limited. Understanding the structure and properties of the E6 and E7 proteins of HPV-31 and HPV-52 is crucial to elucidating their functions and advancing vaccine development.

Methods: A bioinformatics approach was employed to predict the physicochemical properties, hydrophilicity, protein structure, glycosylation sites, phosphorylation sites, terminal positions, signal peptide cleavage sites, transmembrane regions, homology, and dominant epitopes of the E6 and E7 proteins of HPV-31 and HPV-52.

Results: For HPV-31 E6, an instability index (II) of 43.93 indicated that the protein is unstable; potential B-cell epitopes were identified at residues 55–61 (RDDTPYG), 112–116 (PEEKQ), and 125–131 (FHNIGGR), while T-cell epitopes were predicted at residues 45–53 (FAFTDLTIV) and 72–80 (KVSEFRWYR). HPV-52 E6 exhibited an instability index (II) of 55.57, with B-cell epitopes at residues 110–119 (LCPEEKERHV) and 129–141 (MGRWTGRCSECWR), and T-cell epitopes at residues 45–53 (FLFTDLRIV) and 82–87 (SLYGKT). HPV-31 E7, with an instability index (II) of 51.05, exhibited B-cell epitopes at residues 8–17 (QDYYLDLQP), 16–20 (QPEAT), 29–41 (PDSSDEEDVIDEP), and 42–48 (AGQAKPDT), and T-cell epitopes at residues 7–15 (TLQDYVLDL) and 82–90 (LLMGSFGIV). HPV-52 E7, with an instability index (II) of 49.15, exhibited B-cell epitopes at residues 11–19 (YILDLQPET), 23–27 (HCYEQ), 29–38 (GDSSDEEDTD), and 36–48 (DTDGVDRPDGQAE), and T-cell epitopes at residues 53–59 (NYYIVTY) and 84–90 (MLLGTLQ).

Discussion: In summary, the E6 and E7 proteins of HPV-31 and HPV-52 contain dominant epitopes for both T cells and B cells. These findings delineate subtype-specific immunogenic regions and establish a foundation for experimental validation and vaccine design.

KEYWORDS

E6/E7, human papillomavirus 31, human papillomavirus 52, bioanalysis, antigen epitope, oncoprotein

1 Introduction

Human papillomavirus (HPV) is among the most prevalent sexually transmitted viruses worldwide, and infection with HPV has been strongly associated with the development of various cancers, particularly cervical cancer (1). Since the landmark identification of HPV's role in cervical carcinogenesis in the early 1980s (2, 3), the mechanisms by which specific HPV oncoproteins disrupt cellular pathways have been extensively elucidated. HPV types are classified as low-risk or high-risk based on their oncogenic potential (4). While HPV-16 and HPV-18 have been extensively studied, recent epidemiological and molecular studies have underscored the significance of HPV-31 and HPV-52 in cervical cancer incidence, particularly in East Asia and specific regions of Europe (5–8). However, the structural and functional characteristics of the E6 and E7 proteins of HPV-31 and HPV-52 remain poorly characterized.

The oncogenic potential of HPV largely depends on its early proteins, E6 and E7, which facilitate malignant transformation by targeting tumor suppressor pathways (9, 10). E6 binds the p53 tumor suppressor, promoting ubiquitin-mediated degradation and inhibiting apoptosis, while E7 disrupts the retinoblastoma (Rb) pathway to release E2F transcription factors and deregulate cell cycle progression (11–14). Although these mechanisms are conserved among high-risk HPV types, sequence variations in E6 and E7 can lead to differential binding affinities and functional outcomes (15). Recent structural studies have begun to resolve the atomic-level details of HPV-31 and HPV-52 E6 and E7, revealing subtype-specific conformational features that may influence oncogenic potency (16, 18, 19). Nevertheless, a gap remains in the comprehensive bioinformatics characterization of the E6 and E7 proteins of HPV-31 and HPV-52, particularly regarding antigenic epitope prediction—an essential step in vaccine design.

Advances in high-throughput sequencing and computational biology have enabled multidimensional bioinformatics analyses of HPV oncoproteins (16–19). Specifically, homology modeling, molecular docking, epitope mapping, and phylogenetic profiling have uncovered key insights into structural motifs and functional domains of E6 and E7. For instance, Conrady et al. resolved the HPV-31 E6 crystal structure and characterized its interactions with E6AP and p53 (19), whereas Ferenczi et al. conducted phylogenetic and functional analyses of HPV-31 E6 and E7 variants (18). Recent work by Kogure et al. revealed significant intra-patient genomic variability of HPV-31 in cervical cancer and precancer, underscoring the importance of considering viral quasispecies diversity when predicting E6 and E7 epitope profiles (20). Song et al. characterized the genetic variability and phylogeny of HPV-52 E6 and E7 in Sichuan, China, underscoring subtype-specific functional differences relevant to epitope selection (17). Pinheiro et al. conducted a large-scale phylogenomic analysis of HPV-31 across 2,093 genomes, linking specific viral clades to cervical carcinogenesis risk and thereby supporting targeted epitope selection based on subtype phylogeny (21). In summary, prior research has addressed HPV-31 and HPV-52 from various perspectives—sequence diversity (17, 18, 21), structural elucidation (19), and L1 protein-based VLP design (22, 23)—yet none has integrated physicochemical profiling, secondary and tertiary structure modeling, post-translational modification predictions, and B- and T-cell epitope mapping into a single, multilayered framework. Bioinformatics profiling of both subtypes remains incomplete, particularly concerning immunogenic epitope prediction, which is critical for next-generation vaccine design (24).

In this study, the E6 and E7 proteins of HPV-31 and HPV-52 were systematically analyzed using a combination of bioinformatics tools to predict physicochemical properties, post-translational modification sites, secondary and tertiary structures, and to identify potential T-cell and B-cell epitopes. The following hypotheses were tested:

1. HPV-31 and HPV-52 E6 and E7 proteins exhibit subtype-specific sequence and structural variations that lead to distinct distributions of immunogenic epitopes.
2. The simultaneous application of multiple bioinformatics tools to identical sequences was hypothesized to enhance the accuracy of predicting dominant T-cell and B-cell epitopes in HPV-31 and HPV-52 E6 and E7 proteins.
3. By comparing predicted post-translational modification (PTM) sites with conserved regions, immunogenic regions that may be cross-reactive between subtypes were expected to be uncovered.

Further, it was hypothesized that structural disparities between HPV-31 and HPV-52 E6 and E7 proteins correlate with unique antigenic epitope landscapes, thereby informing the design of future peptide-based vaccines.

2 Materials and methods

2.1 Amino acid sequence

The complete sequence of E6 and E7 oncoproteins of HPV-31 and HPV-52 was available from the National Center for Biotechnology Information(NCBI) database (accession numbers: HPV31 E6 [WAB53637], HPV31 E7 [WAB53638], HPV52 E6 [WAB54303], HPV52 E7 [WAB54304]).

2.2 Prediction of protein physicochemical parameters

2.2.1 Rationale for tool selection and distinctions

To assess basic physicochemical properties of HPV-31/52 E6/E7 proteins, we employed two ExPASy tools:

ProtParam (ExPASy ProtParam v2023.1): We used ProtParam to compute molecular weight, theoretical isoelectric point (pI), extinction coefficient, instability index (II), aliphatic index, and GRAVY (grand average of hydropathicity) in a single run. ProtParam is widely used in viral protein studies because its predictions correlate well with experimentally determined parameters. The instability index (II) quantifies the likelihood of a protein's stability *in vitro*, where a value of II > 40 indicates predicted instability (25).

ProtScale (ExPASy ProtScale v2023.1): While ProtParam provides global physicochemical metrics, ProtScale generates residue-level hydrophobicity (Kyte–Doolittle) and hydrophilicity (Hopp–Woods) plots, allowing us to identify local peaks or valleys that may correspond to linear B-cell epitopes. ProtScale employs a sliding-window approach (window size = 7) to generate a continuous hydropathy profile, which ProtParam does not offer (26).

2.2.2 Procedure and statistical processing

The ProtParam calculations were performed in triplicate, and the reported values represent the mean ± standard deviation (SD) of three independent runs.

For the ProtScale analysis, the window width was set to 7 with a default threshold of 0.5. We identified the top three hydrophilicity peaks (using the Hopp–Woods scale) and the deepest hydrophobic valleys (using the Kyte–Doolittle scale) for each protein.

No statistical tests, such as t-tests or ANOVA, were applied because this study is purely predictive, without experimental group comparisons. The results are presented as raw means ± SD for ProtParam values and qualitative hydropathy profiles for ProtScale.

2.3 Post-translational modification site prediction

2.3.1 Rationale for tool selection

NetPhos 3.1 (threshold 0.5): A neural-network–based tool that predicts Ser/Thr/Tyr phos-phorylation sites. We chose NetPhos because it has been benchmarked on short viral proteins with ≥70% accuracy (27). Compared to other open-source servers (e.g., PhosphoSite), NetPhos offers a user-friendly batch interface and provides clear residue-level confidence scores.

MotifScan v2022 (threshold 0.5): Identifies kinase-specific motifs (CK2, PKC, TK, etc.) by searching against curated motif databases (28). We selected MotifScan because it integrates multiple kinase-motif libraries and is particularly suited for mapping short linear motifs adjacent to known functional domains (e.g., LxxLL, LxCxE).

NetNGlyc 1.0 (threshold 0.5): Predicts N-linked glycosylation sites (N-X-S/T motifs) (29). Although E6/E7 proteins rarely undergo glycosylation, we included NetNGlyc to confirm the absence of glycosylation sites—a negative result that supports the cytosolic/nuclear localization of these oncoproteins.

2.3.2 Procedure and output

2.3.2.1 NetPhos 3.1

Submitted each E6/E7 sequence (single sequence mode), extracted residues with score > 0.5.

2.3.2.2 MotifScan v2022

Used default scoring matrices to detect CK2, PKC, TK motifs; only motifs with score > 0.5 were retained.

2.3.2.3 NetNGlyc 1.0

Confirmed that none of the four proteins contained an N-linked glycosylation motif above threshold 0.5.

2.4 Signal peptide and transmembrane helix prediction

SignalP 4.1 (D-score 0.45): Uses a neural network model to predict signal peptide cleavage sites (30). We chose SignalP 4.1 instead of older versions because it offers improved accuracy for proteins lacking obvious signal partners. Its published D-score threshold of 0.45 is recommended for viral oncoproteins.

TMHMM 2.0 (probability threshold 0.5): Predicts transmembrane helices using a hidden Markov model (30). We used TMHMM to verify that E6/E7 do not contain any transmembrane segments, confirming their expected nuclear/cytoplasmic localization.

2.5 Secondary structure prediction

SOPMA v3.0 predicted secondary structure elements (α-helix, β-sheet, β-turn, and random coil) using the default threshold (8%

difference, window width = 17). SOPMA's reported accuracy for viral proteins is ≥70% (31). Compared to alternatives such as PSIPRED, SOPMA provides a residue-level map that can be directly aligned with predicted epitope regions.

2.6 Tertiary structure prediction

Phyre2 v2.0 (Protein Homology/analogY Recognition Engine) (32) was used for homology modeling of E6/E7 proteins. It leverages experimentally resolved PDB templates and generates high-confidence models for proteins with known homologues (33). Although AlphaFold v3 (2024) can produce *de novo* predictions, Phyre2's reliance on validated templates ensures that our HPV E6/E7 models remain directly comparable to prior structural studies (19, 34). This consistency is crucial for accurately mapping predicted epitopes onto known functional domains.

We accepted templates only if they exhibited ≥ 90% sequence coverage and ≥ 99% confidence. Each E6/E7 sequence was submitted in single-sequence mode. For HPV-31 E6, templates c4gizC (coverage 93%, confidence 100%) were chosen; for HPV-31 E7, template d2ewla1 (coverage 50%, confidence 99.8%) was used; for HPV-52 E6, c4gizC (coverage 94%, confidence 100%); for HPV-52 E7, d2b9da1 (coverage 47%, confidence 99.8%).

Template Selection Rationale:

c4gizC: High sequence identity (≥ 90%) with HPV-31/52 E6 in residues 2–144/2–142, respectively (19, 33).

d2ewla1/d2b9da1: Best available templates for E7 with ≥ 99.8% confidence.

Although AlphaFold v3 could produce end-to-end predictions, Phyre2's reliance on experimentally validated templates (e.g., c4gizC) provides clear alignment evidence and facilitates comparability with existing HPV structural literature (18, 19, 33, 35).

2.7 Sequence homology and phylogenetic analysis

Clustal X 2.0 was chosen for multiple sequence alignment (MSA) because it provides a graphical user interface and allows manual inspection of alignment gaps and conserved motifs. Although other aligners exist (e.g., MUSCLE), Clustal X is widely cited in HPV research and facilitates identification of conserved blocks (≥70% identity).

MEGA 7.0.20 (Molecular Evolutionary Genetics Analysis) was used to construct a Neighbor-Joining phylogenetic tree with 1,000 bootstrap replicates, providing statistical support for each branch. MEGA's integrated alignment viewer and tree-editing capabilities streamline the generation of publication-quality phylograms.

We aligned full-length E6/E7 protein sequences from HPV types 16, 18, 31, 33, 35, 45, 52, 56, 58, and 61 using Clustal X 2.0 (gap open penalty = 10; gap extension = 0.1). Evolutionary trees (Neighbor-Joining method, bootstrap = 1,000) were constructed in MEGA 7.0.20 (v7.0.20) to infer phylogenetic relationships. Conserved regions were identified based on ≥ 70% identity across aligned sequences.

2.8 Linear epitope analysis of B cells oncoproteins

We employed four servers to predict linear B-cell epitopes, then selected overlapping regions as dominant candidates:

ABCpred v2.0 (threshold 0.51; peptide length = 16) uses an artificial neural network trained on known linear epitopes (36). We included ABCpred because it has been validated on viral proteins, achieving ~65.9% accuracy (37).

BepiPred 1.0 (threshold 0.35; window = 20) combines hidden Markov models and propensity scales to predict epitopes with a balanced trade-off between specificity and sensitivity (38).

BCPREDS 1.0 (epitope length = 20; specificity = 75%) uses subsequence kernels to identify linear B-cell epitopes; it excels in reducing false positives among random coil regions (39).

SVMTrip v1.0 (threshold 0.51; peptide length = 20) employs a support vector machine algorithm combined with amino acid pair propensity; it outperforms many single-algorithm tools in independently benchmarked tests (40).

Each E6/E7 sequence was submitted to all four servers in single-sequence mode. We recorded all predicted peptide segments that surpassed each server's threshold. Only peptides predicted by ≥ 2 servers were considered for final selection.

2.9 Prediction of T-cell epitopes

$CD4^+$ T cell epitopes were predicted using both SYFPEITHI v1.0 (41) and the IEDB MHC II module (42) with HLA-DRB1*15:01 as the reference allele, selected for its 20% frequency in the Chinese population (43). SYFPEITHI is a motif-based predictor that assigns quantitative scores based on known anchor-residue preferences; peptides scoring ≥ 20 were considered strong binders. The IEDB MHC II module generates consensus predictions by integrating multiple algorithms (e.g., NN-align, SMM-align) and has outperformed standalone tools such as TEPITOPE in benchmark studies; $CD4^+$ epitopes with a percentile rank ≤ 10 were deemed strong binders.

$CD8^+$ T cell epitopes were predicted using the IEDB MHC I module (NetMHCpan 4.1) with HLA-A*11:01 and HLA-A*02:01—alleles occurring at 18.0% and 15.3% frequency in Chinese individuals, respectively (43). NetMHCpan 4.1 employs a pan-specific neural network to predict peptide binding across diverse HLA-A and HLA-B alleles, consistently outperforming earlier NetMHC versions, especially for less common alleles; $CD8^+$ epitopes with a percentile rank ≤ 1 were classified as strong binders. All alleles were chosen based on high-frequency HLA data in the Chinese population (44, 45). The aforementioned methods and corresponding software are summarized in Table 1.

TABLE 1 Methods summary table.

Step	Tool/Method used	Purpose	Key parameters
Amino Acid Sequence	NCBI Database	Retrieve full-length protein sequences	HPV31 E6/E7, HPV52 E6/E7
Physicochemical Parameters	ProtParam, ProtScale	Calculate molecular weight, pI, hydrophobicity, etc.	ProtParam: instability index, GRAVY; ProtScale: hydrophobicity (Kyte–Doolittle), hydrophilicity (Hopp–Woods)
PTM Site Prediction	NetPhos 3.1, MotifScan, NetNGlyc	Predict phosphorylation, kinase motifs, glycosylation	Threshold: 0.5
Signal Peptide Prediction	SignalP 4.1	Predict signal peptide cleavage	D-score ≥ 0.45
Transmembrane Helix	TMHMM 2.0	Predict transmembrane regions	Probability ≥ 0.5
Secondary Structure	SOPMA v3.0	Predict secondary structure (α-helix, β-sheet, etc.)	Threshold: 8% difference, window width = 17
Tertiary Structure	Phyre2 v2.0	Homology modeling	Templates: ≥ 90% coverage, ≥ 99% confidence
Sequence Homology and Phylogenetic Analysis	Clustal X 2.0, MEGA 7.0.20	Align sequences and infer phylogenetic tree	Gap open penalty = 10, gap extension = 0.1, Bootstrap=1,000
B-cell Epitope Prediction	ABCpred v2.0, BepiPred 1.0, BCPREDS 1.0, SVMTrip v1.0	Predict linear epitopes for B-cells	ABCpred v2.0: threshold 0.51; peptide length = 16, BepiPred 1.0: threshold 0.35; window = 20, BCPREDS 1.0: epitope length = 20; specificity = 75%, SVMTrip v1.0: threshold 0.51; peptide length = 20
T-cell Epitope Prediction	SYFPEITHI v1.0, IEDB MHC II, IEDB MHC I	Predict $CD4^+$ and $CD8^+$ T-cell epitopes	$CD4^+$: SYFPEITHI score ≥ 20; $CD8^+$: NetMHCpan percentile rank ≤ 1

3 Results

3.1 Primary structure of HPV-31 and 52 E6 and E7 proteins

The complete amino acid sequences retrieved from NCBI (HPV-31 E6: 149 AA; HPV-31 E7: 98 AA; HPV-52 E6: 148 AA; HPV-52 E7: 99 AA) are listed below:

HPV-31 E6 (149 AA):

MFKNPAERPRKLHELSSALEIPYDELRLNCVYCKGQLT E T E V L D F A F T D L - T I V Y R D D T P Y G V C T K C LRFYSKVSEFRWYRYSVYGT TLEKLTNKGICDLLIR-C I T C Q R P L C P E E K Q R H L DKKKRFHNIGGRWTGRCIVCWRRPRTETQV

HPV-31 E7 (98 AA):

MRGETPTLQDYVLDLQPEATDLYCYEQLPDSSDEEDVID-S P A G Q A K P D T S N Y N I V T F C C Q C E S T L R L C V Q S TQVDIRILQELLMGS F GIVCPNCSTRL

HPV-52 E6 (148 AA):

M F E D P A T R P R T L H E L C E V L E E S V H E I R L Q C V Q C K K E L Q R R E V Y K F L F T D L R I V Y R DNNPYGVCIMCLRFLSKISEYRHYQYSLYGKTLEERV R K P L S E I T I R C I I C Q T P L C P E E K E R H VNANKRFHNIMGRWTGRCSECWRPRPVTQV

HPV-52 E7 (99 AA):

MRGDKATIKDYILD LQPETTDLHCYEQLGDSSDEEDTD GVDRPDGQAEQATSNYYIVTYCHSCDSTLRLCIHSTAT DLRTLQQMLLGTLQVVCPGCAR

3.2 The physicochemical parameters of the proteins

3.2.1 Methods brief

ProtParam v2023.1 was used to compute the length, molecular weight, theoretical pI, instability index (II), aliphatic index, and GRAVY. Each value is the mean ± SD of three independent runs.

ProtScale v2023.1 (window size = 7, threshold = 0.5) was used to generate Hopp–Woods hydrophilicity and Kyte–Doolittle hydrophobicity plots to localize potential B-cell epitopes.

All four proteins have a molecular weight >10 kDa, consistent with the reported immunogenic thresholds (46). Instability indices >40 suggest they are intrinsically unstable, potentially influencing antigen processing (37, 46). Negative GRAVY values classify them as hydrophilic, favoring solubility and surface exposure.

Hydrophilicity/hydrophobicity plots (ProtScale) indicate several predicted hydrophilic peaks in the protein sequences (Figure 1). The physicochemical parameters for all four proteins are summarized in Table 2.

3.3 Post-translational modification and subcellular localization predictions

3.3.1 Methods brief

NetPhos 3.1 (threshold 0.5) was used to predict Ser/Thr/Tyr phosphorylation sites.

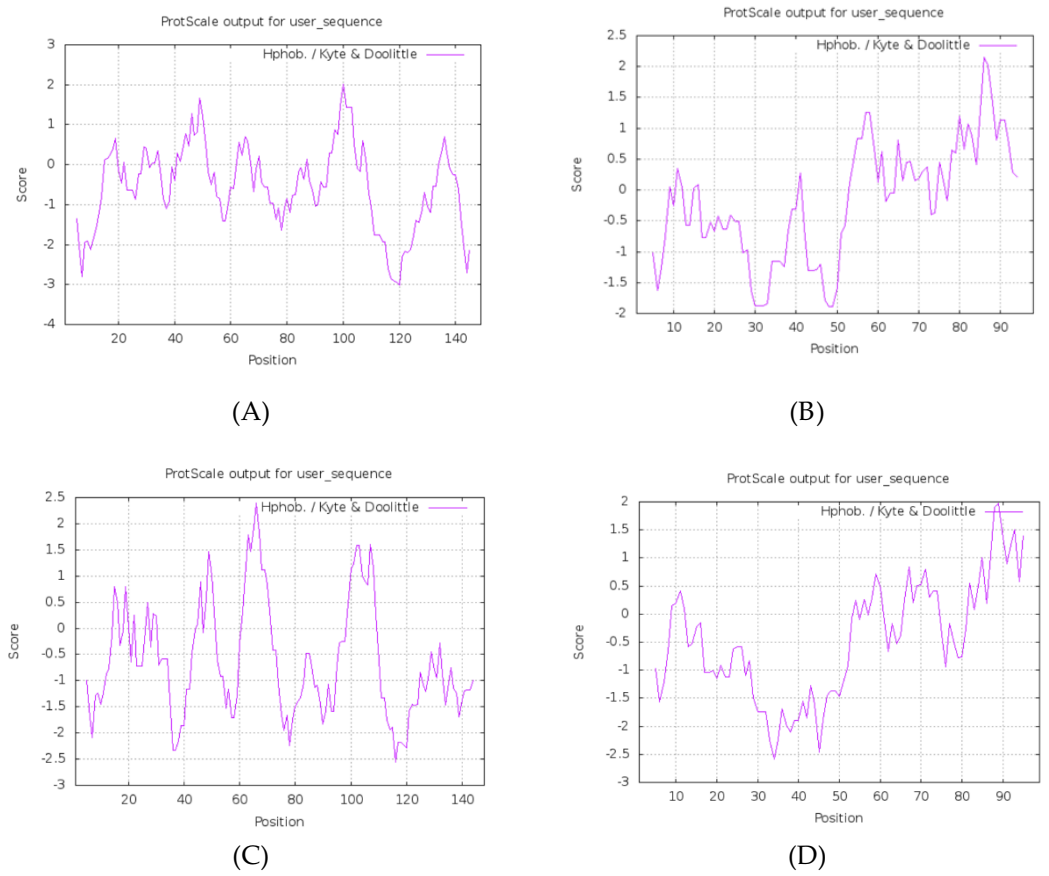


FIGURE 1
Phosphorylation sites: **(A)** HPV31 E6 **(B)** HPV31 E7 **(C)** HPV52 E6 **(D)** HPV52 E7.

MotifScan v2022 (threshold 0.5) was used to identify CK2, PKC, and tyrosine kinase (TK) motifs.

NetNGlyc 1.0 (threshold 0.5) was used to examine possible N-glycosylation sites.

SignalP 4.1 (D-score 0.45) and TMHMM 2.0 (probability 0.5) were used to check for signal peptides and transmembrane helices.

3.3.2 Key findings

The post-translational modification sites and membrane localization of the four proteins are summarized in Table 3. Both E6 proteins have Ser/Thr phosphorylation sites clustered around LxxLL motifs (e.g., S82), suggesting potential regulation of E6AP/p53 binding.

TABLE 2 Summarizes physicochemical parameters for all four proteins.

Protein	AA Length	Molecular Mass (Da)	Theoretical pI	Basic (K,R)	Acidic (D,E)	Instability Index (II)	GRAVY	Classification (II > 40 = unstable; GRAVY < 0 = hydrophilic)
HPV-31 E6	149	17,767.61	9.13	27	18	43.93	−0.567	Unstable; Hydrophilic (Figure 1A)
HPV-31 E7	98	10,944.27	3.90	5	16	51.05	−0.235	Unstable; Hydrophilic (Figure 1B)
HPV-52 E6	148	17,925.85	8.96	26	19	55.57	−0.599	Unstable; Hydrophilic (Figure 1C)
HPV-52 E7	99	11,032.24	4.33	7	17	49.15	−0.459	Unstable; Hydrophilic (Figure 1D)

TABLE 3 Summary of predicted PTM sites and membrane localization (NetPhos 3.1; MotifScan v2022; NetNGlyc 1.0; SignalP 4.1; TMHMM 2.0).

Protein	Phosphorylation (NetPhos > 0.5)	CK2 (CK2 motif > 0.5)	PKC (PKC motif > 0.5)	TK (TK motif > 0.5)	N-Glycosylation (NetNGlyc > 0.5)	Signal Peptide (SignalP D > 0.45)	Transmembrane (TMHMM > 0.5) (Figure 3)
HPV-31 E6	S16, S17, S71, S74, S82; T38, T40, T58, T64, T86, T133, T145, T147; Y60 (Figure 2A)	17–20, 38–42, 86–89	92–94, 133–135	72–79	None	None	None
HPV-52 E6	S22, S71, S74, S82, S97; T11, T48, T108, T133, T146; Y60 (Figure 2C)	11–14, 22–25, 87–90	100–102, 133–135	72–79	None	None	None
HPV-31 E7	S31, S32, S40, S50, S86; T5, T20, T64, T72; Y52 (Figure 2B)	7–10, 31–34, 72–75	64–66, 95–97	None	None	None	None
HPV-52 E7	S31, S32; T7, T19, T20, T37, T58, T66, T76; Y11 (Figure 2D)	7–10, 31–34, 74–77	7–9, 66–68	None	None	None	None

E7 proteins of both subtypes have CK2 sites near the LxCxE Rb-binding motif, suggesting modulation of Rb interaction.

No N-glycosylation, signal peptides, or transmembrane helices were predicted for any of the four proteins, consistent with their known nuclear/cytosolic localization (Figures 2, 3).

3.4 Secondary structure predictions

3.4.1 Methods brief

SOPMA v3.0 (window size = 17, threshold = 8%) was used to determine the percentages of α-helix, β-sheet, β-turn, and random coil.

According to the spatial characteristics of secondary structure, α-helix and β-sheet are not easily disrupted due to hydrogen bonding and are mostly located in the interior of the protein, making them less suitable as antigen-recognizing sites. In contrast, β-turns and irregular curls are primarily protruding structures on the protein surface (47). The specific details of the secondary structures of the four proteins are presented in Table 4. The secondary structure of the HPV-31 E6 protein was analyzed online using SOPMA (Figure 4A). The analysis showed that α-helix accounted for 49.66%, β-sheet for 14.56%, β-turn for 4.43%, and irregular curl for 35.44%. The results indicated that the HPV-31 E6 protein structure is relatively compact (34).

The results for the HPV-31 E7 protein showed that α-helix accounted for 25.51%, β-sheet for 22.45%, β-turn for 0%, and irregular curl for 52.04%, as shown in Figure 4B. The results indicated that the HPV-31 E7 protein structure is relatively loose.

For the HPV-52 E6 protein (Figure 4C), α-helix accounted for 54.05%, β-sheet for 10.81%, β-turn for 1.35%, and irregular curl for 33.78%, indicating that the protein structure is relatively compact.

For the HPV-52 E7 protein (Figure 4D), α-helix accounted for 27.27%, β-sheet for 21.21%, β-turn for 0%, and irregular curl for 51.52%, indicating that the protein structure is relatively loose.

3.5 Tertiary structure prediction (Phyre2 v2.0)

Based on Phyre2 outputs (33), high-confidence homology models were obtained for all four proteins (confidence ≥ 99.8%) (Figures 5A–D).

HPV-31 E6: The model is based on c4gizC (93% coverage, 100% confidence) (Figure 5A).

HPV-31 E7: The model is based on d2ewla1 (50% coverage, 99.8% confidence) (Figure 5B).

HPV-52 E6: The model is based on c4gizC (94% coverage, 100% confidence) (Figure 5C).

HPV-52 E7: The model is based on d2b9da1 (47% coverage, 99.8% confidence) (Figure 5D).

3.5.1 Key findings

E6 proteins are helix-rich and compact, with fewer β-turns, suggesting that most linear epitopes lie in random coil loops.

E7 proteins contain ≥ 50% random coil, indicating extensive surface exposure and many potential linear epitopes.

HPV-31 and HPV-52 E6/E7 structures are highly conserved overall, with only minor local deviations that may underlie subtype-specific immunogenic differences.

3.6 Homology and phylogenetic analysis (Clustal X 2.0 & MEGA 7.0)

3.6.1 Amino acid identity and conserved regions

Multiple sequence alignment of E6 proteins (HPV-16, 18, 31, 33, 35, 45, 52, 56, 58, 61) revealed conserved motifs at positions 8–15, 25–34, 41–77, 79–89, 96–112, 114–141 for HPV-31 E6, and 8–16, 25–31, 41–56, 59–69, 71–79, 81–89, 101–107, 109–119, 123–125, 130–136 for HPV-52 E6 (Figure 6A). E7 proteins exhibited conserved regions at 1–17, 20–28, 30–36, 38–45, 52–77, 82–87, 89–

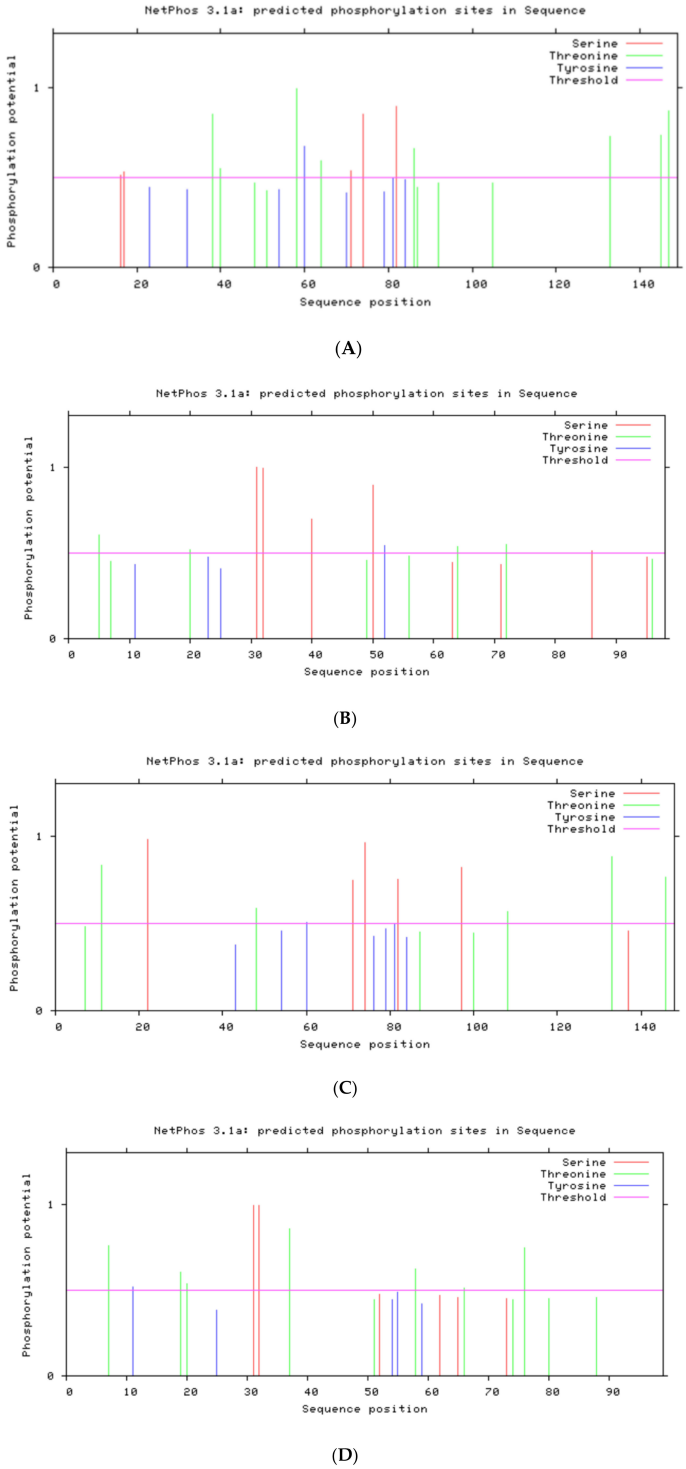


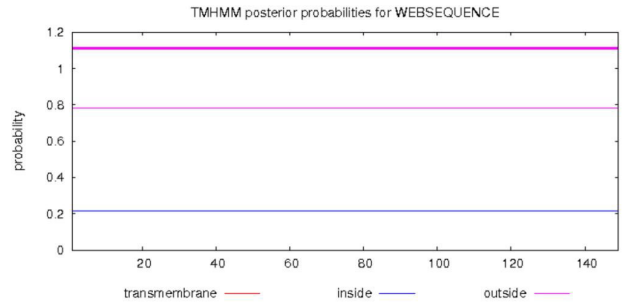
FIGURE 2
Phosphorylation sites: **(A)** HPV31 E6 **(B)** HPV31 E7 **(C)** HPV52 E6 **(D)** HPV52 E7.

94 (HPV-31) and 10–15, 24–28, 30–36, 39–46, 53–59, 62–70, 76–96 (HPV-52) (Figure 6C).

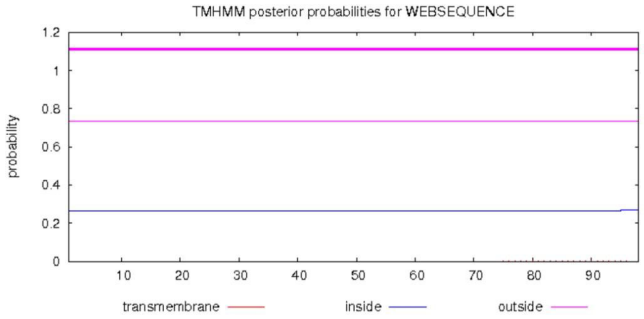
Conserved regions overlap predicted epitope regions, suggesting potential cross-reactivity among related types (48). The HPV−31 E6 45–53 region aligns with the HPV−16 E6 45–53 region, indicating possible shared immune responses.

3.6.2 Phylogenetic tree construction

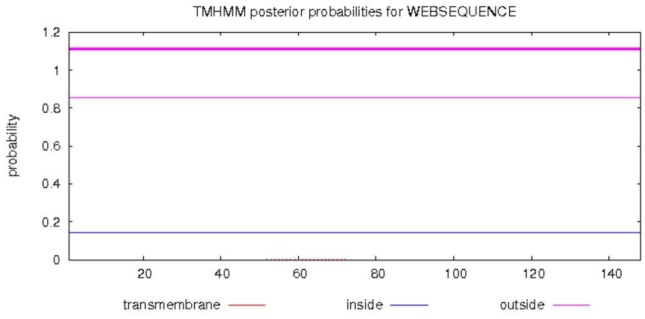
Neighbor-Joining trees (bootstrap = 1,000) placed HPV-31 E6 in a close clade with HPV-35 E6 (Figure 6B), and HPV-52 E6 in a close clade with HPV-33 E6. For E7, HPV-31 clustered with HPV-16, while HPV-52 clustered with HPV-33 (Figure 6D).



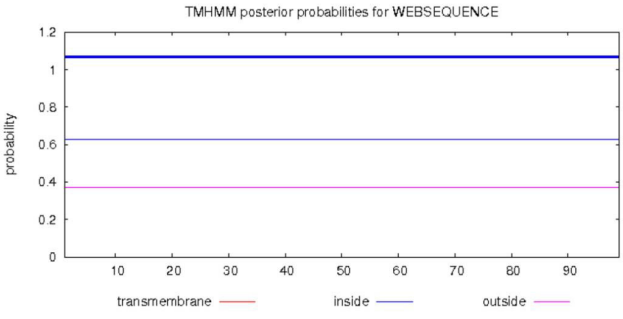
(A)



(B)



(C)



(D)

FIGURE 3
TMHMM analyzed the transmembrane domain of the proteins. **(A)** HPV31 E6 **(B)** HPV31 E7 **(C)** HPV52 E6 **(D)** HPV52 E7.

TABLE 4 Summarizes secondary structure content.

Protein	α-Helix (%)	β-Sheet (%)	β-Turn (%)	Random Coil (%)	Interpretation
HPV-31 E6	49.66	14.56	4.43	35.44	Relatively compact, fewer surface coils
HPV-31 E7	25.51	22.45	0.00	52.04	More random coils, implies greater surface exposure
HPV-52 E6	54.05	10.81	1.35	33.78	Compact with predominant α-helices
HPV-52 E7	27.27	21.21	0.00	51.52	Loose structure with significant coils

3.7 Linear epitopes of B cells

3.7.1 Methods brief

Tools: ABCpred v2.0 (peptide length = 16; threshold = 0.51), BepiPred 1.0 (threshold = 0.35), BCPREDS 1.0 (peptide length = 20; specificity = 75%), and SVMTrip v1.0 (peptide length = 20; threshold = 0.51).

Criterion: Retain only peptides predicted by ≥2 algorithms and restrict to loop/turn regions identified by SOPMA.

After excluding α-helix and β-sheet regions, the top five predicted epitopes per method were compared. Using the four B-cell prediction tools, overlapping epitopes (predicted by ≥2 servers) were identified as dominant (Supplementary Tables 1–16). After cross-referencing, the dominant B-cell epitopes were Table 5:

HPV-31 E6: 55–61 (RDDTPYG), 112–116 (PEEKQ), 125–131 (FHNIGGR)

HPV-31 E7: 8–17 (LQDYVLDLQPEATDLYC), 16–20 (QPEAT), 29–41 (PDSSDEEDVIDEP), 42–48 (AGQAKPDT)

HPV-52 E6: 110–119 (LCPEEKERHV), 129–141 (MGRWTGRCSECWR)

HPV-52 E7: 11–19 (YILDLQPET), 23–27 (HCYEQ), 29–38 (GDSSDEEDTD), 36–48 (DTDGVDRPDGQAE)

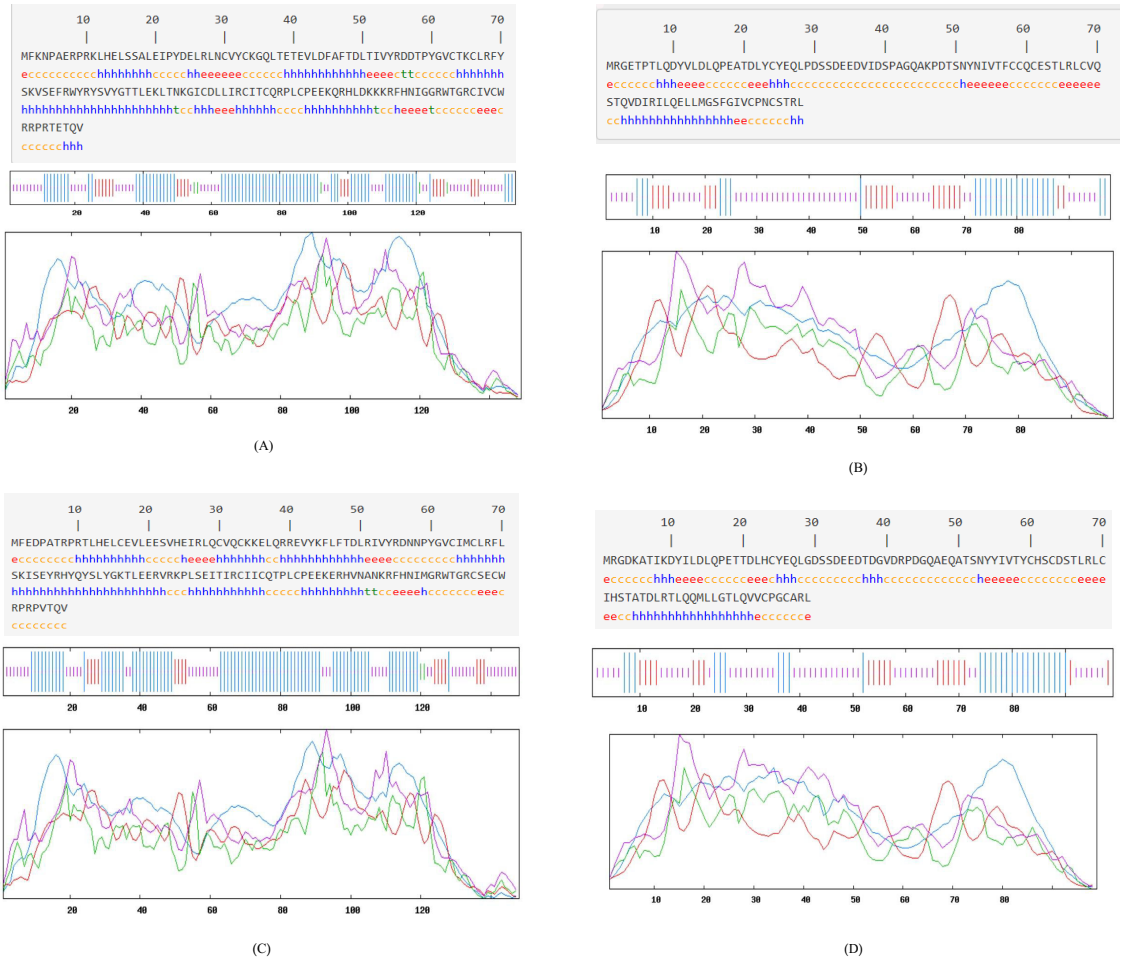


FIGURE 4
Secondary structure prediction: **(A)** HPV31 E6 oncoprotein; **(B)** HPV31 E7 oncoprotein; **(C)** HPV52 E6 oncoprotein; **(D)** HPV52 E7 oncoprotein.

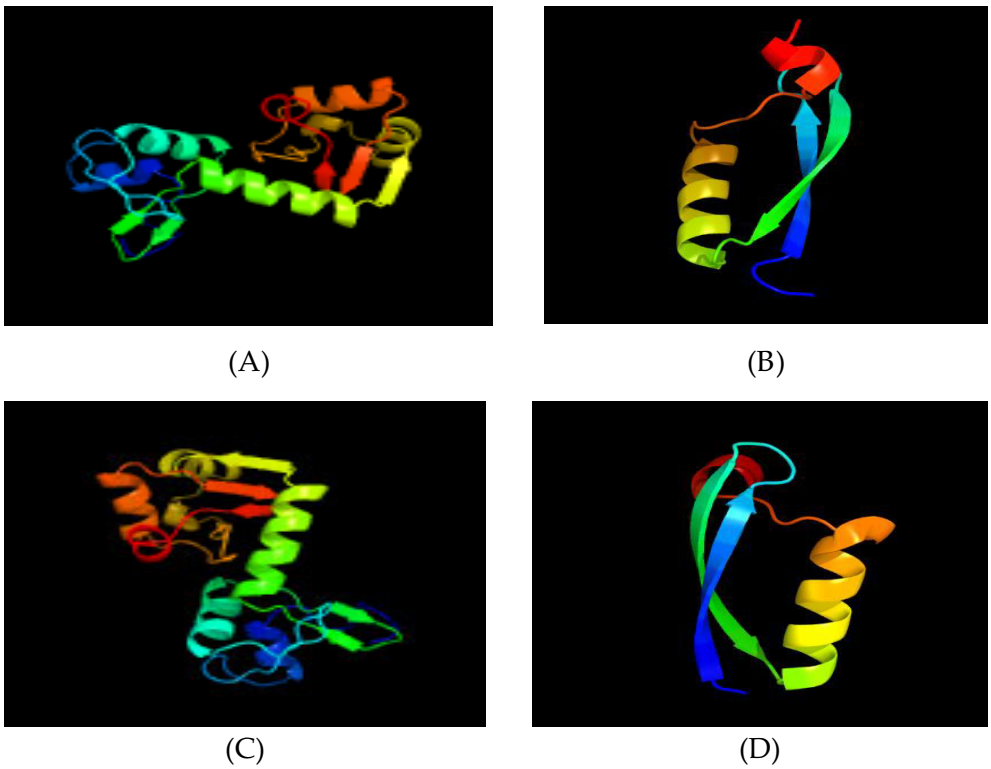


FIGURE 5
Tertiary structure prediction. **(A)** HPV31 E6 protein; **(B)** HPV31 E7 protein; **(C)** HPV52 E6 protein; **(D)** HPV52 E7 protein.

3.7.2 Key findings

HPV-31 E6 candidate epitopes (e.g., 55–61 RDDTPYG) are located in a random coil adjacent to LxxLL, suggesting potential for neutralizing antibodies.

The HPV-31 E7 region 29–41 (PDSSDEEDVIDEP) is consistently predicted by four methods and is located within a highly exposed coil loop.

The C-terminal loops of HPV-52 E6/E7 (e.g., 129–141 in E6, 36–48 in E7) are strong candidates for B-cell epitopes.

3.8 Linear epitopes of T cells

3.8.1 $CD4^+$ T cell epitope prediction (HLA-DRB1*1501)

The SYFPEITHI and IEDB MHC II tools (percentile rank ≤ 10; positive control) were used. Supplementary Tables 17–20 present the top five predictions. The final dominant $CD4^+$ epitopes (overlapping high-scoring predictions) are as follows:

- HPV-31 E6: 45–53 (FAFTDLTIV), 72–80 (KVSEFRWYR).
- HPV-31 E7: 7–15 (TLQDYVLDL), 11–19 (YVLDLQPEA), 82–90 (LLMGSFGIV).
- HPV-52 E6: 45–53 (FLFTDLRIV), 82–87 (SLYGKT).
- HPV-52 E7: 84–90 (MLLGTLQ), 53–59 (NYYIVTY), 11–19 (YILDLQPET).

3.8.2 $CD8^+$ T–cell epitope prediction (HLA-A1101, A0201)

IEDB MHC I binding (NetMHCpan 4.1; percentile rank ≤ 1) was used. Supplementary Tables 21–24 present the results. The final dominant $CD8^+$ epitopes are as follows (Table 6):

- HPV-31 E6: 82–90 (SVYGTTLEK; HLA-A1101 rank 0.01), 45–53 (FAFTDLTIV; HLA-A0201 rank 0.93)
- HPV-31 E7: 7–15 (TLQDYVLDL; HLA-A0201 rank 0.09), 37–46 (VIDSPAGQAK; HLA-A1101 rank 0.33)
- HPV-52 E6: 86–94 (KTLEERVRK; HLA-A1101 rank 0.01), 18–26 (VLEESVHEI; HLA-A0201 rank 0.03)
- HPV-52 E7: 84–92 (MLLGTLQVV; HLA-A0201 rank 0.08), 51–59 (TSNYYIVTY; HLA-A1101 rank 0.74)

Notably, the overlapping T-cell epitope 45–53 appears in both E6 proteins and is conserved between HPV-31 and HPV-52, suggesting a promiscuous HLA-binding region that could elicit cross-type T-cell responses.

4 Discussion

In this study, integrative bioinformatics approaches were employed to analyze the E6 and E7 proteins of HPV-31 and HPV-52, identifying key structural features and dominant

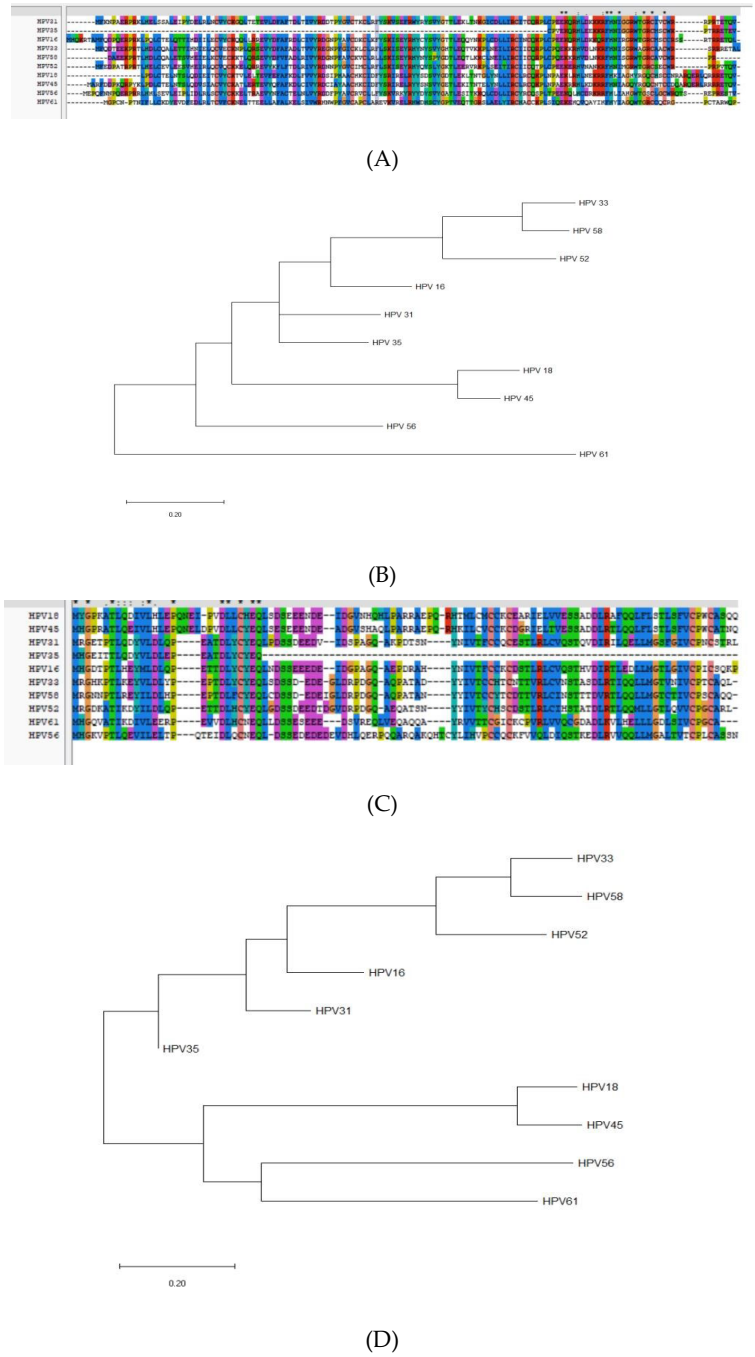


FIGURE 6
Homology and molecular evolution analysis. **(A)** Homology analysis of E6 proteins of HPV; **(B)** The molecular evolutionary tree of E6 proteins of HPV; **(C)** Homology analysis of E7 proteins of HPV; **(D)** The molecular evolutionary tree of E7 proteins of HPV.

antigenic epitopes. The key findings and their biological implications are addressed in the subsequent sections.

4.1 Physicochemical properties and implications for immunogenicity

Viral proteins with molecular weights exceeding 10 kDa typically exhibit sufficient immunogenicity for epitope recognition (46, 49). All four E6 and E7 proteins of HPV-31 and HPV-52 exceed this threshold (17.8–18.0 kDa) and are classified by ProtParam as "unstable" (instability index > 40), a feature associated with increased post-translational susceptibility and potential antigenicity (37, 50, 51). Negative GRAVY scores categorize these proteins as hydrophilic, thereby promoting solubility and enhancing epitope exposure (52). These properties correlate with an enhanced potential for antigen presentation, which is critical for vaccine design.

10.3389/fimmu.2025.1561572

TABLE 5 HPV-31/52 E6/E7 B-Cell epitope candidates (ABCpred; BepiPred; BCPREDS; SVMTrip).

Protein	Algorithm Combination	Residues	Sequence	Structural Context	Tool Version/Threshold
HPV-31 E6	ABCpred & BepiPred & BCPREDS & SVMTrip	55–61	RDDTPYG	Random coil adjacent to LxxLL binding pocket	ABCpred v2.0 (threshold = 0.51) BepiPred 1.0 (threshold = 0.35) BCPREDS 1.0 (length = 20; specificity = 75%) SVMTrip v1.0 (length = 20; threshold = 0.51)
	BepiPred & BCPREDS	112–116	PEEKQ	β-turn at surface	Same as above
	ABCpred & SVMTrip	125–131	FHNIGGR	C-terminal random coil near functional region	Same as above
HPV-52 E6	ABCpred & BepiPred	110–119	LCPEEKERHV	C-terminal β-turn near Zn-finger	Same as above
	BepiPred & SVMTrip	129–141	MGRWTGRCSECWR	Random coil loop; structurally exposed	Same as above
HPV-31 E7	ABCpred & BepiPred	8–17	QDYYLDLQP	N-terminal random coil, high hydrophilicity	Same as above
	BepiPred & BCPREDS	16–20	QPEAT	Small β-turn in central region	Same as above
	BCPREDS & SVMTrip	29–41	PDSSDEEDVIDEP	Long random coil loop with high immunogenicity	Same as above
	SVMTrip & ABCpred	42–48	AGQAKPDT	C-terminal loop region accessible to antibodies	Same as above
HPV-52 E7	ABCpred & BepiPred	11–19	YILDLQPET	N-terminal random coil loop	Same as above
	BepiPred & BCPREDS	23–27	HCYEQ	Small β-turn	Same as above
	BCPREDS & SVMTrip	29–38	GDSSDEEDTD	Central random coil loop	Same as above
	SVMTrip & ABCpred	36–48	DTDGVDRPDGQAE	C-terminal loop region	Same as above

TABLE 6 HPV-31/52 E6/E7 T-Cell Epitope Candidates (SYFPEITHI; IEDB).

Protein	Type	HLA Allele	Residues	Sequence	Affinity Metric (IEDB percentile)	SYFPEITHI Score	Structural Context	Tool Version/ Threshold
HPV-31 E6	$CD4^+$	DRB1*1501	45–53	FAFTDLTIV	3.10	25	Zn-finger region; likely helper epitope	SYFPEITHI v1.0 (score ≥ 20) IEDB MHC-II (percentile ≤ 10) IEDB MHC-I (NetMHCpan 4.1; percentile ≤ 1)
	$CD4^+$	DRB1*1501	72–80	KVSEFRWYR	4.50	22	β-turn at surface; T_H potentiation	Same as above
	$CD8^+$	A*1101	82–90	SVYGTTLEK	0.01	—	Conserved helix; cross-subtype CTL potential	Same as above
	$CD8^+$	A*0201	45–53	FAFTDLTIV	0.93	—	Overlaps with $CD4^+$ 45–53; candidate for poly-epitope design	Same as above

(Continued)

TABLE 6 Continued

Protein	Type	HLA Allele	Residues	Sequence	Affinity Metric (IEDB percentile)	SYFPEITHI Score	Structural Context	Tool Version/ Threshold
HPV-52 E6	$CD4^+$	DRB1*1501	45–53	FLFTDLRIV	3.70	24	Conserved block; cross-protection candidate	Same as above
	$CD4^+$	DRB1*1501	82–87	SLYGKT	4.00	23	Loop region; potential helper epitope	Same as above
	$CD8^+$	A*1101	86–94	KTLEERVRK	0.01	—	Zn-finger adjacency; strong CTL candidate	Same as above
	$CD8^+$	A*0201	18–26	VLEESVHEI	0.03	—	N-terminal helix; antigen-presenting potential	Same as above
HPV-31 E7	$CD4^+$	DRB1*1501	7–15	TLQDYVLDL	7.10	21	N-terminal random coil; T_H epitope candidate	Same as above
	$CD4^+$	DRB1*1501	82–90	LLMGSFGIV	8.00	20	C-terminal coil; possible cross-reactive	Same as above
	$CD8^+$	A*0201	7–15	TLQDYVLDL	0.09	—	Overlaps $CD4^+$ 7–15; poly-epitope design potential	Same as above
	$CD8^+$	A*1101	37–46	VIDSPAGQAK	0.33	—	Central coil loop; strong CTL candidate	Same as above
HPV-52 E7	$CD4^+$	DRB1*1501	11–19	YILDLQPET	9.00	19	Loop region; intermediate T_H immunogenicity	Same as above
	$CD4^+$	DRB1*1501	84–90	MLLGTLQ	13.00	18	C-terminal coil; modest helper response	Same as above
	$CD8^+$	A*0201	84–92	MLLGTLQVV	0.08	—	C-terminal coil; strong CTL candidate	Same as above
	$CD8^+$	A*1101	51–59	TSNYYIVTY	0.74	—	Central coil; potential CTL memory locater	Same as above

4.2 Post-translational modifications and functional context

Predicted phosphorylation sites were mapped to residues involved in the interactions of E6 and E7 with host regulators. For instance, conserved serine residues (S82 in both E6 proteins) reside within the LxxLL-binding pocket, which is crucial for E6AP-mediated p53 degradation (15, 19). CK2 phosphorylation motifs overlapping this region may modulate binding affinity and subsequent ubiquitination (10, 17). Similarly, E7 CK2 sites (e.g., residues 7–10 encompassing the LxCxE motif) likely regulate Rb binding, contributing to cell cycle dysregulation (11, 13). PKC sites adjacent to the C-terminal zinc-finger (E6 133–135) may influence nuclear localization and stability (15). These in silico insights align with experimental evidence showing that kinase-mediated phosphorylation directly alters oncoprotein function (10, 19).

4.3 Secondary/tertiary structures and template selection

SOPMA analysis reveals that the E6 proteins are predominantly composed of α-helices (49.66% in HPV-31; 54.05% in HPV-52), suggesting compact cores that may shield specific epitopes. In contrast, the E7 proteins exhibit a higher proportion of random coils (52.04% and 51.52%, respectively), indicating flexible surface regions conducive to antibody binding (37, 53). Previous studies have shown that random coils frequently coincide with B-cell

epitope hotspots (52, 53), supporting our predictions of dominant linear B-cell epitopes within coil-rich segments, such as residues 8–17 (HPV-31 E7) and 23–27 (HPV-52 E7).

Homology models generated by Phyre2 (confidence > 99.8%) confirm conserved structural motifs, including zinc-binding Cys motifs, consistent with experimental structures (19, 40). The 3D models generated by Phyre2, validated by high confidence scores, display conserved zinc-finger motifs and binding pockets. While AlphaFold3 (2025 release) could generate full-length models, Phyre2's template-based approach allowed for a direct comparison with known E6/E7 structures. We selected Phyre2 templates (c4gizC/d2ewla1/d2b9da1) due to their high sequence identity (>50%) and prior experimental validation (18, 19).

4.4 Homology and evolutionary insights

Multiple sequence alignment and phylogenetic analysis position HPV-31 E6 closely with HPV-35, and HPV-52 E6 with HPV-33, while E7 clusters similarly with HPV-16 and HPV-33 (18, 24). Conserved regions (e.g., E6 positions 41–77; E7 positions 52–77) overlap with predicted T-cell epitopes, suggesting potential cross-reactivity and cross-protection among high-risk HPV types (16, 17). This cross-immunity is essential for the design of multivalent vaccines targeting broad high-risk HPV coverage (6).

4.5 Antigenic epitope identification and validation potential

Dominant B-cell epitopes were identified (e.g., HPV-31 E6: 55–61, 112–116, 125–131; HPV-52 E7: 23–27, 29–38, 36–48) and T-cell epitopes (e.g., HPV-31 E6: 45–53; HPV-52 E6: 86–94), predicted by multiple algorithms (ABCpred, BepiPred 1.0, BCPREDS, SVMTriP) (37, 54, 55). $CD8^+$ epitopes, such as HPV-31 E7: 7–15 (TLQDYVLDL), exhibited a strong binding affinity to HLA-A0201 (IEDB rank 0.09), consistent with known CTL responses against HPV-16 E7 (11, 56). $CD4^+$ epitopes (e.g., HPV-52 E7: 11–19) exhibited favorable binding to HLA-DRB1*1501, which is crucial for helper T-cell activation (42). These in silico predictions align with experimental data linking epitope immunodominance to surface accessibility and structural features (54, 55). Subsequent empirical validation, such as peptide-MHC binding assays and T-cell activation studies, is necessary (42, 56).

4.6 Comparison with previous studies

Previous studies have characterized the sequence variability of HPV-31/52 (17, 18, 21) and resolved individual E6 crystal structures (19). Kogure et al. further demonstrated that HPV-31 genomes exhibit significant intra-patient heterogeneity (20), suggesting that E6 and E7 epitopes may evolve during disease progression. However, to date, no study has integrated physicochemical properties, post-translational modification site prediction, secondary and tertiary structure modeling, and multilayered immunoinformatic epitope mapping for both E6 and E7 of HPV-31 and HPV-52 into a single comprehensive analysis. Our work addresses this gap by correlating predicted phosphorylation sites with functional motifs (e.g., LxxLL, LxCxE) (27, 57) and mapping B- and T-cell epitopes to conserved, surface-exposed regions identified through structural modeling. Furthermore, Song et al. and Firdaus et al. have highlighted the immunogenic potential of HPV-52 (17, 22, 23), particularly in Asian populations, thus validating the public health relevance of our subtype-specific epitope predictions. Kesheh et al. proposed region-tailored multivalent vaccine designs based on L1 gene diversity (58), offering translational context for our E6 and E7-based epitope candidates.

4.7 Application to vaccine design

Although this study did not experimentally construct virus-like particles (VLPs) or multivalent peptide vaccines, the predicted epitopes provide a foundation for rational vaccine design:

4.7.1 Cross-subtype conserved $CD8^+$ epitopes

The E6 45–53 segment in HPV-31 (FAFTDLTIV) and HPV-52 (FLFTDLRIV) exhibits strong binding affinity for HLA-A0201 and HLA-A1101 (IEDB percentile ≤ 1) and is highly conserved across high-risk types, making it an ideal candidate for inclusion as a universal cytotoxic T-lymphocyte (CTL) epitope in a multi-epitope DNA or peptide vaccine.

4.7.2 Helper T-cell ($CD4^+$) epitopes

E6 72–80 (KVSEFRWYR) in HPV-31 and E6 82–87 (SLYGKT) in HPV-52 exhibit moderate binding affinity to HLA-DRB1*1501 (IEDB percentile ≤ 10) and could be fused with CTL epitopes into a single recombinant protein or synthetic long peptide construct to enhance helper T-cell responses, as suggested by He et al (57).

4.7.3 B-cell neutralizing epitopes on VLP platforms

The B-cell epitope HPV-31 E6 55–61 (RDDTPYG) and HPV-52 E6 110–119 (LCPEEKERHV) reside in exposed random coil regions. Firdaus et al. successfully inserted analogous linear epitopes into the L1 VLP platform to elicit neutralizing antibodies (22), supporting the strategy of grafting these peptides onto L1 VLPs to generate subtype-specific antibody responses.

4.7.4 Multivalent peptide/protein vaccine constructs

Building on Firdaus et al.'s reverse vaccinology design for HPV-52 L1 (23), one could concatenate top $CD4^+$ and $CD8^+$ epitopes (e.g., E6 45–53, 72–80; E7 7–15) with appropriate linkers and trafficking signals to create a chimeric protein capable of eliciting robust humoral and cell-mediated immunity in preclinical HLA-transgenic mouse models.

5 Limitations and future directions

Although the integrated in silico pipeline provides a comprehensive epitope landscape, experimental validation—such as peptide-MHC binding assays, ELISpot, and crystallographic studies—is crucial to confirm immunogenicity (54, 55). Additionally, molecular dynamics simulations could refine epitope conformations and assess stability within MHC binding grooves (32, 51). This study relies solely on in silico predictions and lacks direct *in vitro* or *in vivo* validation, representing a primary limitation. Pinheiro et al. confirmed that certain E6 and E7 regions correlate with cervical cancer aggressiveness at the genomic level (21), yet these findings require empirical confirmation through immunological assays. Kogure et al. observed intra-patient HPV-31 variants across different lesion stages (20), emphasizing the need to validate epitope immunogenicity across clinical time points. Future studies should involve:

5.1 Experimental binding assays

Use ELISPOT or flow cytometry with peptide-stimulated peripheral blood mononuclear cells (PBMCs) from HLA-typed donors to validate $CD4^+$ and $CD8^+$ T-cell responses against the predicted epitopes.

5.2 Antibody neutralization studies

Synthesize candidate B-cell epitopes (e.g., HPV-31 E6 55–61; HPV-52 E6 110–119) and assess their ability to induce neutralizing antibodies in ELISA or pseudovirus neutralization assays.

5.3 Animal model validation

Evaluate peptide-based or VLP-based vaccine constructs (e.g., insertion of linear epitopes into L1 VLPs, as demonstrated by Firdaus et al., 2023) in HLA-transgenic mouse models to measure protective efficacy against HPV-induced tumorigenesis.

In summary, the integrative bioinformatics analysis illuminates subtype-specific structural and immunogenic features of HPV-31 and HPV-52 E6 and E7 proteins, laying the groundwork for experimental validation and rational vaccine design aimed at reducing the HPV-associated cervical cancer burden.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/.

Author contributions

QC: Conceptualization, Data curation, Software, Writing – original draft, Writing – review & editing. YF: Data curation, Software, Writing – original draft. WD: Formal Analysis, Visualization, Writing – original draft. YM: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. 1) Mechanism study of specific histone deacetylase 6 inhibitor in treating asthma by regulating mast cell function, National Natural Science Foundation of China (82170038)2) Identification of clinical phenotypes of bronchial asthma and establishment of decision tree model China International Medical Foundation (Z-2017-24-2301)3) Mechanism study of immune imbalance in bronchial asthma and establishment of precise treatment system, Liaoning Provincial Science and Technology Program Joint Program (2023JH2/101700100).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1561572/full#supplementary-material

SUPPLEMENTARY TABLE 1
Prediction results of ABCpred B cell epitopes of HPV-31 E6 protein.

SUPPLEMENTARY TABLE 2
Prediction of ABCpred B cell epitopes of HPV-31 E7 protein.

SUPPLEMENTARY TABLE 3
Prediction results of ABCpred B cell epitopes of HPV-52 E6 protein.

SUPPLEMENTARY TABLE 4
Prediction of ABCpred B cell epitopes of HPV-52 E7 protein.

SUPPLEMENTARY TABLE 5
Prediction results of the Bepipred 1.0 Server B cell epitopes of HPV-31 E6 protein.

SUPPLEMENTARY TABLE 6
Prediction of the Bepipred 1.0 Server B cell epitopes of HPV-31 E7 protein.

SUPPLEMENTARY TABLE 7
Prediction of the Bepipred 1.0 Server B cell epitopes of HPV-52 E6 protein.

SUPPLEMENTARY TABLE 8
Prediction of the Bepipred 1.0 Server B cell epitopes of HPV-52 E7 protein.

SUPPLEMENTARY TABLE 9
Prediction results of BCPreds B cell epitopes for HPV-31 E6 protein.

SUPPLEMENTARY TABLE 10
Prediction of BCPreds B cell epitopes for HPV-31 E7 protein.

SUPPLEMENTARY TABLE 11
Prediction of BCPreds B cell epitopes for HPV-52 E6 protein.

SUPPLEMENTARY TABLE 12
Prediction of BCPreds B cell epitopes for HPV-52 E7 protein.

SUPPLEMENTARY TABLE 13
Prediction of SVMTRIP B cell epitopes of HPV-31 E6 protein.

SUPPLEMENTARY TABLE 14
Prediction of SVMTRIP B cell epitopes of HPV-31 E7 protein.

SUPPLEMENTARY TABLE 15
Prediction of SVMTRIP B cell epitopes of HPV-52 E6 protein.

SUPPLEMENTARY TABLE 16
Prediction of SVMTRIP B cell epitopes of HPV-52 E7 protein.

SUPPLEMENTARY TABLE 17
CD4+T cell epitope prediction of HPV-31 E6 protein with HLA-DRB1*1501 as allele parameter.

SUPPLEMENTARY TABLE 18
Prediction of CD4+T cell epitopes of HPV-31 E7 protein using HLA-DRB1*1501 as allele parameter.

SUPPLEMENTARY TABLE 19
Prediction of CD4+T cell epitopes of HPV-52 E6 protein with HLA-DRB1*1501 as allele parameter.

SUPPLEMENTARY TABLE 20
Prediction of CD4+T cell epitopes of HPV-52 E7 protein with HLA-DRB1*1501 as allele parameter.

SUPPLEMENTARY TABLE 21
Prediction of CE8+T cell epitopes of HPV-31 E6 oncogene using alleles HLA-A*1101 and HLA-A*0201 as parameters.

SUPPLEMENTARY TABLE 22
Prediction of CE8+T cell epitopes of HPV-31 E7 oncogene using alleles HLA-A*1101 and HLA-A*0201 as parameters.

SUPPLEMENTARY TABLE 23
Prediction of CE8+T cell epitopes of HPV-52 E6 oncogene using alleles HLA-A*1101 and HLA-A*0201 as parameters.

SUPPLEMENTARY TABLE 24
Prediction of CE8+T cell epitopes of HPV-52 E7 oncogene using alleles HLA-A*1101 and HLA-A*0201 as parameters.

SUPPLEMENTARY TABLE 25
Prediction of HPV-31 E6 antigen dominant epitopes in T cells and B cells.

SUPPLEMENTARY TABLE 26
Prediction of HPV-31 E7 antigen dominant epitopes in T cells and B cells.

SUPPLEMENTARY TABLE 27
Prediction of HPV-52 E6 antigen dominant epitopes in T cells and B cells.

SUPPLEMENTARY TABLE 28
Prediction of HPV-52 E7 antigen dominant epitopes in T cells and B cells.

References

1. Castle PE, Einstein MH, Sahasrabuddhe VV. Cervical cancer prevention and control in women living with human immunodeficiency virus. *CA: Cancer J Clin*. (2021) 71:505–26. doi: 10.3322/caac.21696

2. Hao L, Jiang Y, Zhang C, Han P. Genome composition-based deep learning predicts oncogenic potential of hpvs. *Front Cell infection Microbiol*. (2024) 14:1430424. doi: 10.3389/fcimb.2024.1430424

3. Rosendo-Chalma P, Antonio-Véjar V, Ortiz Tejedor JG, Ortiz Segarra J, Vega Crespo B, Bigoni-Ordóñez GD. The hallmarks of cervical cancer: molecular mechanisms induced by human papillomavirus. *Biology*. (2024) 13:77. doi: 10.3390/biology13020077

4. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *New Engl J Med*. (2003) 348:518–27. doi: 10.1056/NEJMoa021641

5. Lee J, Kim DJ, Lee HJ. Assessment of Malignant potential for hpv types 16, 52, and 58 in the uterine cervix within a korean cohort. *Sci Rep*. (2024) 14:14619. doi: 10.1038/s41598-024-65056-7

6. Cuzick RA, Wheeler CM. HPV genotype-specific risk for cervical cancer (2021). Available online at: www.HPVWorld.com (Accessed June 27, 2025).

7. Abate A, Munshea A, Nibret E, Alemayehu DH, Alemu A, Abdissa A, et al. Characterization of human papillomavirus genotypes and their coverage in vaccine delivered to Ethiopian women. *Sci Rep*. (2024) 14:7976. doi: 10.1038/s41598-024-57085-z

8. So KA, Lee IH, Lee KH, Hong SR, Kim YJ, Seo HH, et al. Human papillomavirus genotype-specific risk in cervical carcinogenesis. *J gynecologic Oncol*. (2019) 30:e52. doi: 10.3802/jgo.2019.30.e52

9. Bruyere D, Roncarati P, Lebeau A, Lerho T, Poulain F, Hendrick E, et al. Human papillomavirus E6/E7 oncoproteins promote radiotherapy-mediated tumor suppression by globally hijacking host DNA damage repair. *Theranostics*. (2023) 13:1130–49. doi: 10.7150/thno.78091

10. Bhattacharjee R, Das SS, Biswal SS, Nath A, Das D, Basu A, et al. Mechanistic role of hpv-associated early proteins in cervical cancer: molecular pathways and targeted therapeutic strategies. *Crit Rev oncology/hematology*. (2022) 174:103675. doi: 10.1016/j.critrevonc.2022.103675

11. Yim EK, Park JS. The role of hpv E6 and E7 oncoproteins in hpv-associated cervical carcinogenesis. *Cancer Res Treat*. (2005) 37:319–24. doi: 10.4143/crt.2005.37.6.319

12. Tewari KS, Monk BJ. New strategies in advanced cervical cancer: from angiogenesis blockade to immunotherapy. *Clin Cancer research: an Off J Am Assoc Cancer Res*. (2014) 20:5349–58. doi: 10.1158/1078-0432.Ccr-14-1099

13. Yeo-Teh NSL, Ito Y, Jha S. High-risk human papillomaviral oncogenes E6 and E7 target key cellular pathways to achieve oncogenesis. *Int J Mol Sci*. (2018) 19. doi: 10.3390/ijms19061706

14. Pal A, Kundu R. Human papillomavirus E6 and E7: the cervical cancer hallmarks and targets for therapy. *Front Microbiol*. (2019) 10:3116. doi: 10.3389/fmicb.2019.03116

15. Peng Q, Wang L, Zuo L, Gao S, Jiang X, Han Y, et al. Hpv E6/E7: insights into their regulatory role and mechanism in signaling pathways in hpv-associated tumor. *Cancer Gene Ther*. (2024) 31:9–17. doi: 10.1038/s41417-023-00682-3

16. Li S, Ye M, Chen Y, Gong Q, Mei B. Genetic variation of E6 and E7 genes of human papillomavirus 52 from central China. *J Med Virol*. (2021) 93:3849–56. doi: 10.1002/jmv.26690

17. Song Z, Cui Y, Li Q, Deng J, Ding X, He J, et al. The genetic variability, phylogeny and functional significance of E6, E7 and lcr in human papillomavirus type 52 isolates in sichuan, China. *Virol J*. (2021) 18:94. doi: 10.1186/s12985-021-01565-5

18. Ferenczi A, Gyöngyösi E, Szalmás A, László B, Kónya J, Veress G. Phylogenetic and functional analysis of sequence variation of human papillomavirus type 31 E6 and E7 oncoproteins. *Infection Genet evolution: J Mol Epidemiol evolutionary Genet Infect Dis*. (2016) 43:94–100. doi: 10.1016/j.meegid.2016.05.020

19. Conrady MC, Suarez I, Gogl G, Frecot DI, Bonhoure A, Kostmann C, et al. Structure of high-risk papillomavirus 31 E6 oncogenic protein and characterization of E6/E6ap/P53 complex formation. *J Virol*. (2020) 95:e00730-20. doi: 10.1128/jvi.00730-20

20. Kogure G, Tanaka K, Matsui T, Onuki M, Matsumoto K, Iwata T, et al. Intra-patient genomic variations of human papillomavirus type 31 in cervical cancer and precancer. *Viruses*. (2023) 15:2104. doi: 10.3390/v15102104

21. Pinheiro M, Harari A, Schiffman M, Clifford GM, Chen Z, Yeager M, et al. Phylogenomic analysis of human papillomavirus type 31 and cervical carcinogenesis: A study of 2093 viral genomes. *Viruses*. (2021) 13:1948. doi: 10.3390/v13101948

22. Firdaus MER, Mustopa AZ, Ekawati N, Chairunnisa S, Arifah RK, Hertati A, et al. Optimization, characterization, comparison of self-assembly vlp of capsid protein L1 in yeast and reverse vaccinology design against human papillomavirus type 52. *Journal Genet Eng Biotechnol*. (2023) 21:68. doi: 10.1186/s43141-023-00514-9

23. Firdaus MER, Mustopa AZ, Triratna L, Syahputra G, Nurfatwa M. Dissection of capsid protein hpv 52 to rationalize vaccine designs using computational approaches immunoinformatics and molecular docking. *Asian Pacific J Cancer prevention: APJCP*. (2022) 23:2243–53. doi: 10.31557/apjcp.2022.23.7.2243

24. Malla R, Kamal MA. E6 and E7 oncoproteins: potential targets of cervical cancer. *Curr medicinal Chem*. (2021) 28:8163–81. doi: 10.2174/0929867327666201111145546

25. Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the swiss bioinformatics resource portal, as designed by its users. *Nucleic Acids Res*. (2021) 49:W216–w27. doi: 10.1093/nar/gkab225

26. Chen Z, Zhu Y, Sha T, Li Z, Li Y, Zhang F, et al. Design of a new multi-epitope vaccine against brucella based on T and B cell epitopes using bioinformatics methods. *Epidemiol infection*. (2021) 149:e136. doi: 10.1017/s0950268821001229

27. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. (2004) 4:1633–49. doi: 10.1002/pmic.200300771

28. Hao X, Li J, Gao S, Tuerxun Z, Chang X, Hu W, et al. Sspsah, a H subunit of the photosystem I reaction center of suaeda salsa, confers the capacity of osmotic adjustment in tobacco. *Genes Genomics*. (2020) 42:1455–65. doi: 10.1007/s13258-020-00970-4

29. Rizal FA, Ho KL, Omar AR, Tan WS, Mariatulqabtiah AR, Iqbal M. Sequence analysis of the Malaysian low pathogenic avian influenza virus strain H5n2 from duck. *Genes*. (2023) 14:1973. doi: 10.3390/genes14101973

30. Nielsen H, Teufel F, Brunak S, von Heijne G. Signalp: the evolution of a web server. *Methods Mol Biol (Clifton NJ)*. (2024) 2836:331–67. doi: 10.1007/978-1-0716-4007-4_17

31. Dristy TT, Noor AR, Dey P, Saha A. Structural analysis and conformational dynamics of socs1 gene mutations involved in diffuse large B-cell lymphoma. *Gene*. (2023) 864:147293. doi: 10.1016/j.gene.2023.147293

32. Rouka E, Gourgoulianni N, Lüpold S, Hatzoglou C, Gourgoulianis K, Blanckenhorn WU, et al. The drosophila septate junctions beyond barrier function: review of the literature, prediction of human orthologs of the sj-related proteins and identification of protein domain families. *Acta physiologica (Oxford England)*. (2021) 231:e13527. doi: 10.1111/apha.13527

33. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. (2015) 10:845–58. doi: 10.1038/nprot.2015.053

34. Kumar NV, Rani ME, Gunaseeli R, Kannan ND, Sridhar J. Modeling and structural analysis of cellulases using clostridium thermocellum as template. *Bioinformation*. (2012) 8:1105–10. doi: 10.6026/97320630081105

35. Satitsuksanoa P, Kennedy M, Gilis D, Le Mignon M, Suratannon N, Soh WT, et al. The minor house dust mite allergen der P 13 is a fatty acid-binding protein and an activator of a tlr2-mediated innate immune response. *Allergy*. (2016) 71:1425–34. doi: 10.1111/all.12899

36. Zheng D, Liang S, Zhang C. B-cell epitope predictions using computational methods. *Methods Mol Biol (Clifton NJ)*. (2023) 2552:239–54. doi: 10.1007/978-1-0716-2609-2_12

37. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. (2006) 65:40–8. doi: 10.1002/prot.21078

38. Li XW, Zhang N, Li ZL, Dibo N, Ma ZR, Lu B, et al. Epitope vaccine design for toxoplasma gondii based on a genome-wide database of membrane proteins. *Parasites Vectors*. (2022) 15:364. doi: 10.1186/s13071-022-05497-z

39. Motamedi H, Ari MM, Shahlaei M, Moradi S, Farhadikia P, Alvandi A, et al. Designing multi-epitope vaccine against important colorectal cancer (Crc) associated pathogens based on immunoinformatics approach. *BMC Bioinf*. (2023) 24:65. doi: 10.1186/s12859-023-05197-0

40. Yao B, Zheng D, Liang S, Zhang C. Svmtrip: A method to predict B-cell linear antigenic epitopes. *Methods Mol Biol (Clifton NJ)*. (2020) 2131:299–307. doi: 10.1007/978-1-0716-0389-5_17

41. Heidarinia H, Tajbakhsh E, Rostamian M, Momtaz H. Two peptides derivate from acinetobacter baumannii outer membrane protein K as vaccine candidates: A comprehensive in silico study. *BMC Res Notes*. (2023) 16:128. doi: 10.1186/s13104-023-06409-9

42. Vita R, Blazeska N, Marrama D, Duesing S, Bennett J, Greenbaum J, et al. The immune epitope database (Iedb): 2024 update. *Nucleic Acids Res*. (2025) 53:D436–d43. doi: 10.1093/nar/gkae1092

43. Li Y, He J, Bao XJ, Qiu QC, Yuan XN, Xu C, et al. A study on allele frequencies and mismatching proportion of hla-a, B, cw, drb1 and dqb1 on high-resolution donor-recipient typing in chinese han population. *Zhonghua yi xue yi Chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chin J Med Genet*. (2011) 28:92–8. doi: 10.3760/cma.j.issn.1003-9406.2011.01.021

44. Beskow AH, Josefsson AM, Gyllensten UB. Hla class ii alleles associated with infection by hpv16 in cervical cancer in situ. *Int J Cancer*. (2001) 93:817–22. doi: 10.1002/ijc.1412

45. Ghaderi M, Nikitina L, Peacock CS, Hjelmström P, Hallmans G, Wiklund F, et al. Tumor necrosis factor a-11 and dr15-dq6 (B*0602) haplotype increase the risk for cervical intraepithelial neoplasia in human papillomavirus 16 seropositive women in northern Sweden. *Cancer epidemiology Biomarkers prevention: Publ Am Assoc Cancer Research cosponsored by Am Soc Prev Oncol*. (2000) 9:1067–70.

46. Butsashvili M, Kajaia M, Kochlamazashvili M, Zarandia M, Gagua T, Meskhishvili D, et al. Genotypic distribution of hpv among women of reproductive age in Georgia. *Georgian Med News*. (2016) 258):40–3.

47. Ropón-Palacios G, Chenet-Zuta ME, Otazu K, Olivos-Ramirez GE, Camps I. Novel multi-epitope protein containing conserved epitopes from different leishmania species as potential vaccine candidate: integrated immunoinformatics and molecular dynamics approach. *Comput Biol Chem*. (2019) 83:107157. doi: 10.1016/j.compbiolchem.2019.107157

48. van den Hende M, Redeker A, Kwappenberg KM, Franken KL, Drijfhout JW, Oostendorp J, et al. Evaluation of immunological cross-reactivity between clade A9 high-risk human papillomavirus types on the basis of E6-specific cd4+ Memory T cell responses. *J Infect Dis*. (2010) 202:1200–11. doi: 10.1086/656367

49. Condie D. Roitt's Essential Immunology – 10th Edition [Book Review]. *The Australian Journal of Medical Science*. (2003) 24:212.

50. Madeleine MM, Johnson LG, Smith AG, Hansen JA, Nisperos BB, Li S, et al. Comprehensive analysis of hla-a, hla-B, hla-C, hla-drb1, and hla-dqb1 loci and squamous cell cervical cancer risk. *Cancer Res*. (2008) 68:3532–9. doi: 10.1158/0008-5472.Can-07-6471

51. Barlow DJ, Edwards MS, Thornton JM. Continuous and discontinuous protein antigenic determinants. *Nature*. (1986) 322:747–8. doi: 10.1038/322747a0

52. Chenzhang Y, Wen Q, Ding X, Cao M, Chen Z, Mu X, et al. Identification of the impact on T- and B- cell epitopes of human papillomavirus type-16 E6 and E7 variant in southwest China. *Immunol Lett*. (2017) 181:26–30. doi: 10.1016/j.imlet.2016.09.013

53. Sela-Culang I, Ofran Y, Peters B. Antibody specific epitope prediction-emergence of a new paradigm. *Curr Opin Virol*. (2015) 11:98–102. doi: 10.1016/j.coviro.2015.03.012

54. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. (2006) 2:2. doi: 10.1186/1745-7580-2-2

55. Yao B, Zhang L, Liang S, Zhang C. Svmtrip: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PloS One*. (2012) 7:e45152. doi: 10.1371/journal.pone.0045152

56. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. Syfpeithi: database for mhc ligands and peptide motifs. *Immunogenetics*. (1999) 50:213–9. doi: 10.1007/s002510050595

57. He J, Li Q, Ma S, Li T, Chen Y, Liu Y, et al. The polymorphism analysis and epitope predicted of alphapapillomavirus 9 E6 in sichuan, China. *Virol J*. (2022) 19:14. doi: 10.1186/s12985-021-01728-4

58. Mobini Kesheh M, Shavandi S, Azami J, Esghaei M, Keyvani H. Genetic diversity and bioinformatic analysis in the L1 gene of hpv genotypes 31, 33, and 58 circulating in women with normal cervical cytology. *Infect Agents Cancer*. (2023) 18:19. doi: 10.1186/s13027-023-00499-7

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 01 September 2025
DOI 10.3389/fimmu.2025.1617251

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College and Hospital, India

REVIEWED BY
Salman Sadullah Usmani,
Albert Einstein College of Medicine, United States
Xiao-Yan Yang,
Zhuhai Campus of Zunyi Medical University, China
Mohit Chawla,
King Abdullah University of Science and Technology, Saudi Arabia

*CORRESPONDENCE
Sasikumar K
sasikumar.k@vit.ac.in

RECEIVED 24 April 2025
ACCEPTED 06 August 2025
PUBLISHED 01 September 2025

CITATION
Mudipalli Elavarasu S and K S (2025) Rational design of an epitope-centric vaccine against *Pseudomonas aeruginosa* using pangenomic insights and immunoinformatics approach. *Front. Immunol.* 16:1617251. doi: 10.3389/fimmu.2025.1617251

Rational design of an epitope-centric vaccine against *Pseudomonas aeruginosa* using pangenomic insights and immunoinformatics approach

Santhosh Mudipalli Elavarasu[1] and Sasikumar K[2*]

[1]Department of Integrative Biology, School of Biosciences and Technology, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India, [2]Department of Sensor and Biomedical Technology, School of Electronics Engineering, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India

Introduction: As a highly adaptable opportunistic pathogen, *Pseudomonas aeruginosa* presents a significant threat to people with weakened immune systems. This is because it naturally resists antibiotics and can form biofilms. These factors complicate treatment and underscore the urgent need for innovative therapeutic strategies, such as vaccines, to combat this pathogen.

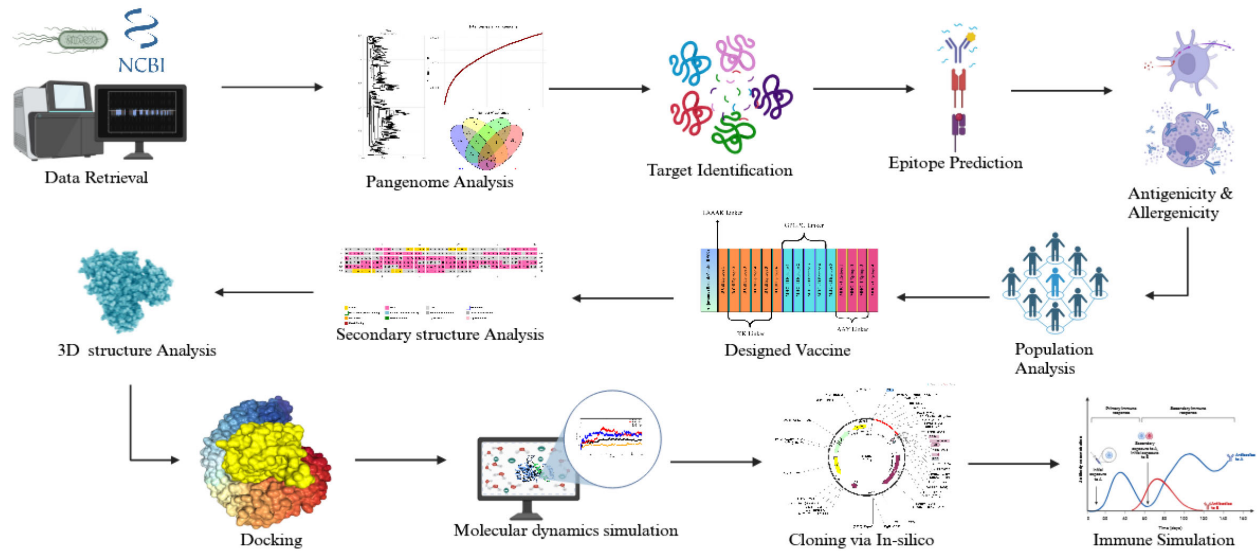
Methods: A pangenome analysis of *P. aeruginosa* genomes was performed to identify conserved core genes critical for bacterial survival and virulence. LptF, an outer membrane protein, was prioritized as a target for vaccine development. B-cell and T-cell epitopes were predicted from LptF using immunoinformatics tools, and a multi-epitope peptide vaccine was designed. The interaction between the vaccine candidate and Toll-like receptors (TLRs) was investigated through molecular docking and molecular dynamics simulations. Codon optimization and in-silico cloning were carried out to validate the vaccine's expression potential in *E. coli*. Immune response simulations evaluated the vaccine's immunogenicity.

Results: Our pangenome analysis identified highly conserved core genes, including LptF, which proved crucial for bacterial virulence. A multi-epitope peptide vaccine was designed using the most immunogenic B-cell and T-cell epitopes derived from LptF. Studies using molecular docking and dynamic simulation have shown stable interactions between the vaccine and TLRs, with the POA_V_RS09 construct exhibiting the highest stability. Codon optimization indicated high expression efficiency in *E. coli*. Immune simulations revealed robust adaptive immune responses, including sustained IgG production, the formation of memory B cells, and the activation of T-cell responses.

Discussion: The POA_V_RS09 vaccine candidate exhibited excellent stability, immunogenic potential, and expression efficiency, making it a promising candidate for combating *P. aeruginosa* infections. This study provides a strong foundation for developing effective therapeutic strategies to address the growing issue of antimicrobial resistance in *P. aeruginosa*. More experimental validation is needed to verify its effectiveness in preclinical and clinical environments.

KEYWORDS

***Pseudomonas aeruginosa*, pangenome analysis, immunoinformatics, epitope-based vaccine, molecular docking, immune simulation**



GRAPHICAL ABSTRACT

1 Introduction

Pseudomonas aeruginosa (*P. aeruginosa*), a highly adaptable opportunistic pathogen, is a significant cause of multidrug-resistant (MDR) infections, including diabetic foot infections, ventilator-associated pneumonia, wound infections, septicemia, and catheter-associated urinary tract infections (1). It poses a significant threat, particularly to immunocompromised individuals, due to its intrinsic resistance to antibiotics and its ability to thrive in diverse environments. Furthermore, *P. aeruginosa* can spread through medical equipment, increasing the risk of cross-contamination between patients and complicating infection control in healthcare settings (2). According to the World Health Organization (WHO), antimicrobial resistance (AMR) is expected to cause 10 million deaths annually by 2050, highlighting its severe impact as a global health threat (3). Hospital-acquired infections caused by ESKAPE pathogens, *Enterobacter* species, *P. aeruginosa*, *Staphylococcus aureus, Acinetobacter baumannii, Klebsiella pneumoniae*, and *Enterococcus faecium* are particularly concerning as they employ diverse mechanisms to resist antibiotics, making treatment increasingly challenging (4). Addressing *P. aeruginosa's* virulence and its role as a key contributor to AMR highlights the urgent need for new therapeutic strategies, such as vaccines, to mitigate its impact (5, 6).

According to the WHO's 2024 list of critical diseases, *P. aeruginosa* is a high-burden resistant bacterium resistant to last-resort antibiotics (7). Factors contributing to its pathogenicity include secretion systems, biofilm formation, and toxin production. Biofilms protect bacteria from host immune responses and

Abbreviations: WHO, World Health Organization; AMR, Antimicrobial Resistance; *P. aeruginosa*, *Pseudomonas aeruginosa;* NCBI, National Center for Biotechnology Information; VFDB, Virulence Factor Database; MD, Molecular dynamics; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Fluctuation; PCA, Principal Component Analysis; FEL, Free Energy Landscape; HBOND, Hydrogen bond analysis.

medications, promoting the formation of multidrug-resistant persister cells that cause recurrent infections, particularly in patients with cystic fibrosis (8). *P. aeruginosa* employs its Type III secretion system to inject effector proteins into the host cells, interfering with cellular processes and facilitating immune evasion (9). The bacterium exhibits three primary resistance mechanisms: intrinsic resistance (efflux pumps, antibiotic-inactivating enzymes, limited outer membrane permeability), acquired resistance (mutations or horizontal gene transfer leading to resistance to aminoglycosides, quinolones, and β-lactams), and adaptive resistance (driven by persister cells and biofilm formation) (10). Clinical outcomes of *P. aeruginosa* infections are generally worse than those caused by other bacteria (11–13). During the COVID-19 pandemic, despite a decrease in the overall number of isolates, the percentage of MDR *P. aeruginosa* isolates increased significantly, from 23.8% in 2019 to 38.8% in 2020 (14). This trend was influenced by longer hospital stays, increased ICU admissions, and a greater reliance on empirical antibiotics, primarily due to the severity of cases and the extensive use of mechanical ventilation. This highlights how AMR is exacerbated in healthcare settings during pandemics (15). With the overuse of antibiotics, slow development of new drugs, and increasing complexity of healthcare, AMR is expected to worsen, leading to higher mortality rates and a greater burden on healthcare systems globally. Traditional antibiotics are becoming ineffective against MDR and extensively drug-resistant (XDR) strains, which no longer respond to standard treatments (16). The limited efficacy of last-resort drugs, such as colistin, coupled with their toxicity risks, makes managing resistant infections even more challenging (17). The lack of specific, targeted therapies for resistant infections leaves healthcare providers with limited options, underscoring the need for novel treatments and more effective alternatives to combat AMR (18). Among vaccine development studies for *P. aeruginosa*, outer membrane proteins such as Porin F (OprF) and Lipoprotein I (OprI) have been extensively explored as potential antigen targets (19).

Vaccines are crucial for preventing infections and reducing antibiotic use in low- and middle-income countries, significantly contributing to the fight against AMR. By lowering the incidence of infectious diseases, vaccines help minimize antibiotic misuse and overuse, particularly in populations with limited access to healthcare (20). Vaccines hold significant promise in addressing AMR by preventing infections, reducing antibiotic dependency, and curbing the spread of resistant strains (21). However, designing a vaccine for *P. aeruginosa* has been challenging due to its complex genetic diversity, biofilm formation, and immune evasion capabilities (22). Recent advancements in genomics and immunoinformatics offer new opportunities to overcome these obstacles. Computational tools for identifying novel vaccine candidates pave the way for developing targeted vaccines that can address the diversity of *P. aeruginosa* strains and enhance immune protection (23). In this study, we employed a high-resolution pan-genomic analysis of complete *P. aeruginosa* genomes from the NCBI RefSeq database to identify core, virulence-associated proteins. Among the prioritized candidates, LptF, a component of the LPS transport system, has been classified as a lipotoxin (LPT) due to its ability to trigger strong pro-inflammatory responses via TLR2 activation, particularly in cystic fibrosis. LptF is a pro-inflammatory lipotoxin involved in the excessive induction of IL-8 in cystic fibrosis and remains underexplored as a vaccine target (24). Our pan-genome analysis has identified LptF as a key membrane-associated protein that interacts with virulence factors, such as OprI and LptE, which supports its potential as a new therapeutic candidate (25). Our pipeline integrates reverse vaccinology, structural modeling, and molecular dynamics simulations to design a multi-epitope subunit vaccine construct. Unlike previous studies that relied on reference strains, metabolic enzymes, or limited proteome screening, our approach emphasizes strain-wide conservation, immune accessibility, and functional relevance. This integrative, pathogen-focused design offers a rational and potentially effective strategy for developing a broad-coverage vaccine against MDR *P. aeruginosa*. Using linkers, these epitopes can be linked to effective adjuvants to develop vaccines.

2 Materials and methods

2.1 Genome data retrieval

A comprehensive dataset of *P. aeruginosa* genomes, all at the "complete" assembly level, was obtained from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) using the NCBI Genome Download Toolkit (26). To ensure comprehensive genomic representation, this dataset included a variety of strains, encompassing both clinical isolates and reference strains.

2.2 Pangenome construction and analysis

P. aeruginosa strains underwent pangenome analysis using the Roary tool (Version 3.13.0) (27). A diverse set of strains was initially selected to capture extensive genetic variability by collecting whole genomes from the NCBI RefSeq database. These genomes were annotated using Prokka (Version 1.14.6), which converted raw sequences into functional gene and protein data (28). Prokka is used to annotate essential genetic elements such as transfer RNA (tRNA), ribosomal RNA (rRNA), and coding sequences (CDS) for each genome, ensuring consistent annotation across all strains. Roary identifies the core and the accessory genes, revealing the conserved and variable genomic regions among *P. aeruginosa* strains. Core genes from all genomes were extracted from the Roary output for further detailed analysis, providing insights into essential genomic elements and potential targets for vaccine or therapeutic development. This pangenome analysis elucidated the genetic composition of the species and identified potential targets for further therapeutic advancements.

2.3 Prediction of subcellular localization

Following the identification of core genes, we employed the PSORTb tool (version 3.0.3) to predict their subcellular localization

(29). PSORTb, a robust tool for prokaryotic genome analysis, categorized the core genes based on their predicted cellular locations, including cytoplasmic, periplasmic, and outer membrane regions. This study primarily focused on outer membrane proteins due to their accessibility on the bacterial surface, making them ideal targets for vaccine development. To confirm that the selected outer membrane-associated genes did not show homology with the human proteins, we conducted a comparison against the human proteome using BLASTP analysis (E-value 0.0001) (30). This step was essential to avoid potential cross-reactivity and enhance the specificity of vaccine candidate selection.

2.4 Analysis using the virulence factor database

The identified outer membrane proteins were analyzed by comparing them to the Virulence Factor Database (VFDB) using BLASTP [E-value = 0.0001, protein sequences from the VFDB full dataset (set B)] (31). This comparative analysis aimed to determine whether the selected protein candidates possess virulence potential, thereby assessing their suitability as targets for therapeutic or vaccine development. By aligning these outer membrane proteins with known virulence factors, we identified candidates with established roles in pathogenicity, enhancing the selection of proteins with significant implications in host-pathogen interactions. The selected target underwent an additional BLASTP search against the *P. aeruginosa* database for further validation (32). This analysis provided insights into the protein's potential role, supporting its relevance for subsequent stages of the study.

2.5 Immunogenic potential and structural characterization of vaccine candidate

We evaluated the selected sequence's antigenic potential using the VaxiJen v2.0 (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) server (33) to determine its suitability as an antigenic candidate. The sequence was analyzed with Allertop v2.0 (34) to assess allergenic risk, ensuring it lacked properties that could trigger allergic reactions. We used the ProtParam tool (https://web.expasy.org/protparam/) to determine the physicochemical properties, including molecular weight, instability index, grand average of hydropathicity (GRAVY), and hydrophobicity (35). These analyses provided essential insights into the protein's suitability for vaccine development by assessing its immunogenicity, safety, and stability.

2.6 Signal peptide prediction analysis

SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) is a sophisticated bioinformatics tool designed to detect signal peptides in protein sequences and pinpoint their cleavage sites (36). Utilizing protein language models (LMs), it analyzes the N-terminal region of proteins. Based on the predicted pathway and cleavage mechanism, SignalP classifies signal peptides into various types, such as Sec/SPI and Tat/SPI. The tool provides crucial scores, including the S-score for signal peptide probability and the C-score for predicting cleavage sites. This is essential for developing vaccines, as it helps identify secreted or surface-exposed proteins that could serve as potential immunogenic targets.

2.7 Prediction of linear B-cell epitopes

For the prediction of linear B-cell epitopes, we utilized BepiPred 2.0, which relies on the Immune Epitope Database (IEDB) Analysis resource (https://www.iedb.org/) (37, 38). This tool analyses amino acid propensity scores and identifies patterns typical of B-cell epitopes, using propensity scales to identify regions likely to consist of these epitopes. Improved accuracy of predictions is achieved by training on known antigen-antibody complexes, with the Random Forest method refining the results. The antigenic potential of the predicted epitopes was further assessed using VaxiJen v2.0 to determine their ability to stimulate an immune response. In this study, it served as an additional screening tool to prioritize epitopes (B and T Cell epitopes) with higher intrinsic antigenic potential before subjecting them to downstream immunoinformatics and structural analyses. Allertop v2.0 assessed allergenicity, ensuring the epitopes would not trigger allergic reactions. Additionally, the toxicity profiles of the selected epitopes were evaluated using the ToxinPred server (39), making sure they had a low risk of allergic reactions was a key step in designing the vaccine.

2.8 Prediction of T-cell epitopes (MHC Class I and II)

Epitope prediction for helper (HTL) and cytotoxic (CTL) T lymphocytes was performed using the NetMHCpan 4.1 algorithm provided by the Immune Epitope Database (IEDB) Analysis Resource (40). The focus was on non-structural (NS) proteins, which are conserved across various strains of *P. aeruginosa* and serve as key targets for immune responses. A human-specific approach was employed for CTL epitopes, identifying 10-mer peptides (ten amino acids long) that included 27 common HLA alleles as a reference panel. These epitopes were chosen for their ability to bind to MHC class I molecules and activate cytotoxic T cells, which is essential for targeting and eliminating infected cells. We selected T-cell epitopes based on recommendations from the IEDB for binding predictions. Specifically, we selected epitopes with a percentile rank of ≤ 1% for MHC class I, and a median percentile rank of ≤20% for MHC class II. These thresholds represent high and moderate affinity binders, and we mapped them to our scoring scale (≥ 0.60 for class I and ≥ 0.75 for class II) to include biologically relevant epitopes (41). For HTL epitopes, 15-mer peptides likely to stimulate helper T cells were identified using the IEDB-

recommended method. These epitopes were designed to bind to MHC class II molecules, thereby activating B cells and initiating the humoral immune response. The input included antigenic, non-allergenic, and NS proteins from *P. aeruginosa* to ensure the predicted epitope's efficacy and safety for vaccine development.

2.9 Prediction of interferon-γ inducing MHC-II epitopes

In this study, the IFNepitope server was used to predict MHC-II epitopes capable of inducing Interferon-gamma (IFN-γ) responses. This web-based tool leverages a comprehensive dataset from the IEDB, comprising 6,728 non-inducing epitopes and 3,705 confirmed IFN-γ-inducing epitopes (42). Utilizing the Support Vector Machine (SVM) technique, the server combines sequence analysis with predictive algorithms to identify epitopes with a high potential to stimulate IFN-γ production. We also analyzed the IL-4 prediction web server (43), the IL-6 prediction web server (44), the IL-10 prediction web server (45), and the IL-13 prediction web server (46). Epitopes were selected for vaccine development based on their prior assessment for antigenicity and non-allergenicity. This tool also prioritizes safe and immunologically relevant epitopes, which boosts the chances of a successful immune response.

2.10 Analysis of population coverage

The finalized T-cell epitopes and their associated HLA binding data were evaluated using the IEDB's Population Coverage module to determine their global distribution across diverse populations (47). This analysis provided valuable insights into the epitope's coverage across different demographics and regions, enhancing our understanding of their potential immunological effectiveness. By examining the global distribution of these epitopes, the study highlighted their relevance to diverse demographic groups. This crucial step demonstrated the epitope's ability to address global healthcare needs, ensuring the vaccine candidate's potential to protect a wide range of populations, thereby increasing its worldwide applicability and efficacy.

2.11 Vaccine design and construction

This study enhanced the vaccine design by incorporating carefully selected adjuvants, linkers, and epitopes to amplify the immune response. Two adjuvants were selected for their immune-boosting properties: RS-09 (APPHALS), a short peptide mimicking bacterial lipopolysaccharide, and Beta-defensin, a potent antimicrobial peptide known for its strong immunological activation (48, 49). Four linkers were used to achieve the best positioning and presentation of the epitopes. The EAAAK linker connected the adjuvants to the epitopes. This rigid helical linker promotes spatial separation between the adjuvant and the epitope region, thereby minimizing potential structural interference and enhancing adjuvant-mediated immune activation. The Alanine-Alanine-Tyrosine (AAY) linker was employed between MHC-I epitopes to enhance processing and presentation by MHC class I molecules. The MHC-II epitopes were separated using the Glycine-Proline-Glycine-Proline-Glycine (GPGPG) linker, which is a flexible and hydrophilic linker that aids in preserving epitope integrity and enhances recognition by helper T cells. Finally, the KK (Lysine-Lysine) linker was used to connect B-cell epitopes, ensuring adequate exposure for B-cell activation while maintaining their conformational flexibility and immunogenicity (50). These strategic additions of adjuvants and linkers were designed to optimize the vaccine's ability to elicit strong and targeted immune responses, effectively combating the intended disease.

2.12 Analysis of the physicochemical properties of the formulated vaccines

The ProtParam server was utilized to conduct a physicochemical analysis of the developed vaccine candidates, assessing their stability and suitability for development (51). We analyzed the amino acid sequences to identify key structural and functional features. We calculated the molecular weight to estimate the proteins' size, solubility, and potential antigenicity. To assess their biochemical behavior under physiological conditions, we determined the theoretical isoelectric point (pI), which indicated their net charge and acid–base characteristics. We also computed the instability index to predict the likelihood of protein degradation. However, the aliphatic index was evaluated to determine temperature stability based on the contribution of aliphatic amino acids. The GRAVY index was also evaluated to determine the vaccine's overall hydrophobic or hydrophilic nature, aiding in understanding its solubility and stability.

2.13 Secondary structure analysis and prediction

The secondary structure of the developed vaccine was predicted using the PSIPRED tool (52), a widely used online resource for protein structure annotation and prediction. PSIPRED offers comprehensive protein analysis tools (53), with a focus on structural feature prediction. This analysis yielded valuable insights into how the vaccine might interact, its stability, and its functional properties. After entering the amino acid sequence of the final vaccine construct, the PSIPRED server analyzed the sequence and predicted the secondary structure, identifying coil, β-sheet, and α-helical regions. These predictions provide crucial insights about the overall structure and organization of the vaccine's protein backbone.

2.14 Prediction and computational refinement of tertiary structure

To predict the three-dimensional (3D) structure of the developed vaccine and facilitate docking analysis, the ROBETTA server and AlphaFold (54, 55), which employ deep-learning

techniques for accurate protein modeling, were utilized. The complete amino acid sequence of the vaccine was entered into both platforms, resulting in the prediction of multiple 3D structures in PDB format. These structures were enhanced in quality and accuracy using the GalaxyRefine tool (56). This tool refines the models by correcting structural errors, optimizing energy levels, and minimizing steric clashes. A comparative analysis of the refined models was conducted, and the best-performing structure, as determined by structural validation using a Ramachandran plot and other quality metrics, was selected for further docking studies.

2.15 Molecular docking and interaction studies

We employed molecular docking analysis to examine the interactions between the vaccine construct and Toll-like receptors TLR2 and TLR4, which are critical mediators of innate immune responses to infection. TLR2 was selected for its ability to detect various pathogen-associated molecular patterns and initiate immune responses (57). RS09 is a synthetic TLR4 agonist peptide that stimulates innate immunity. The TLR4 receptor recognizes a TLR4 agonist linked to the N-terminus of the vaccine construct. When TLR4 is activated, it triggers an intracellular signaling process via the NF-kB pathway, resulting in the production of inflammatory cytokines (58, 59). We retrieved the 3D structures of TLR2 and TLR4 from the RCSB PDB database, using IDs 2Z7X and 3FXI for TLR2 and TLR4, respectively, for further analysis (60, 61). Before docking, we thoroughly prepared the receptor structures by removing heteroatoms, water molecules, and bound ligands to ensure accurate analysis. This step was vital to prevent any interference during the docking process. Docking simulations were performed using the HDOCK web server (http://hdock.phys.hust.edu.cn/) (62). It employs a hybrid docking algorithm that combines template-based and free docking methods. In this study, we did blind docking to allow unbiased prediction of potential interaction sites between the vaccine construct and immune receptors. HDOCK, which is known for its intuitive interface and robust protein-protein docking capabilities, facilitated the simulation process by leveraging the refined 3D structure of the vaccine and the immune receptor models of TLR2 and TLR4. The docking affinity scores, indicating the strength of interaction between the vaccine and the receptors, were used to evaluate the results. Additionally, key residues involved in binding interactions were identified, providing insights into how these immune receptors recognize the vaccine. This study helps elucidate how the vaccine may effectively interact with TLR2 and TLR4, key components of the innate immune system, to trigger an immune response.

2.16 Molecular dynamics simulation analysis

To conduct molecular dynamics (MD) simulations for 1000ns, we utilized the CHARMM-GUI server's Solution Builder protocol, applying the CHARMM36 force field to generate the necessary input files (63). The TIP3P water model was used to solvate the protein-protein complexes, creating a realistic simulation environment by enclosing the system in a periodic cubic box extending 10 Å from the protein atoms in all directions (64). Counter ions were added to neutralize the system, ensuring overall charge balance. The Verlet cutoff method was employed with a 10 Å cutoff distance, striking a balance between computational efficiency and accuracy to calculate electrostatic and van der Waals interactions. Bond constraints were applied using the LINCS algorithm to stabilize the simulation by maintaining fixed bond lengths. The Particle Mesh Ewald (PME) method was used to precisely calculate long-range electrostatic interactions, enhancing simulation accuracy in systems with periodic boundary conditions (65). To remove undesirable interactions and stabilize the system, the solvated system was subjected to energy minimization using the steepest descent technique (66). Two equilibration phases followed: the first in the NVT ensemble (constant Number of particles, Volume, and Temperature) to stabilize temperature, and the second in the NPT ensemble (constant number of particles, Pressure, and Temperature) to stabilize pressure. Proper thermostat and barostat techniques maintained constant temperature and pressure levels. This dual equilibration ensured system stability before the production run. The simulation recorded coordinates every 1 ps with a time step of 2 fs, striking a balance between computational efficiency and accuracy. CHARMM-GUI provided Python scripts to convert topology (top) and parameter (itp) files into GROMACS-compatible formats, simplifying input file preparation (67). Following the post-production run, we performed thorough trajectory analyses, including calculating Root Mean Square Deviation (RMSD) for structural stability, Root Mean Square Fluctuation (RMSF) for flexibility, hydrogen bond analysis (HBOND) for molecular interactions, Principal Component Analysis (PCA) for dominant motion patterns, Buried Surface Area (BSA) for evaluating binding stability, and Free Energy Landscape (FEL) analysis for the conformational states of the protein-protein complexes. Free energy calculations were performed for the interaction between TLR complexes and the vaccine construct (POA_V_RS09, POA_V_BDEF) using the MM-PBSA method with a Poisson-Boltzmann approach (68, 69). These approaches account for various energy components, including bonded interactions, van der Waals forces, electrostatic effects, and both polar and non-polar solvation energies. Here in MM-PBSA, the polar solvation energy is derived from the Poisson–Boltzmann equation, utilizing the molecular dynamics (MD) trajectory to compute interaction energies throughout the simulation. These analyses provided valuable insights into structural stability, flexibility, interaction dynamics, and potential conformational changes, enhancing our understanding of protein-protein interactions over time (70–72).

2.17 *In silico* cloning and expression analysis

To ensure optimal expression in the desired host, the gene of interest was first subjected to codon optimization using the

GenScript program (www.genscript.com), aligning the gene sequence with the host's preferred codon usage (73). Using SnapGene software (https://www.snapgene.com/), we cloned the vaccine construct via in silico. The result showed that the gene of interest and the pET-28a(+) plasmid did not share any restriction sites. This was addressed by adding specific nucleotide sequences to the gene's N-terminal and C-terminal ends, which matched the restriction sites XhoI and NdeI, thereby aiding in cloning. These sequences provided suitable restriction sites for the accurate insertion of the gene into the plasmid. The appropriate recombinant plasmid construct was then produced by cloning the codon-optimized gene into the pET-28a(+) plasmid in silico using additional restriction sequences (74).

2.18 C-IMMSIM-based immune simulation

The C-IMMSIM server (https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php) (75), a widely used tool for simulating immune responses, was employed to evaluate the in-silico immunological response of the developed vaccine. This server utilizes a simulation-based framework to replicate the function of the human immune system and its organs, with a particular emphasis on primary lymphoid tissues. It uses a position-specific scoring matrix, enhanced by machine learning algorithms, to predict immune reactions. To achieve a total simulation period of 1050 steps, the vaccine and adjuvant were given in three doses: an initial dose, a second dose administered 84 days later, and a third dose administered 1050 days later, spaced eight hours apart. The adjuvant concentration was set to 100, and the injected antigen amount was 1000, following the server's default parameters. This setup enabled a comprehensive evaluation of the immune response triggered by the vaccine or the adjuvant.

3 Results

3.1 NCBI data retrieval

A diverse array of 864 complete *P. aeruginosa* genomes, encompassing strains such as PAO1, PA14, PAK, LESB58, and CF39S, was sourced from the NCBI Assembly database utilizing the NCBI-genome-download toolkit. Supplementary Table S1 contains detailed information on all included genomes, ensuring a comprehensive genomic representation for subsequent analyses.

3.2 Pangenome analysis

A thorough pangenome analysis was performed on 864 complete genomes of *P. aeruginosa* sourced from the NCBI Assembly database. Genome annotation was executed using Prokka, followed by pangenome analysis with Roary, which identified a total of 63,239 genes. Of these, 3,325 were classified as core genes. Within this core set, 296 genes were consistently present across all genomes, with 79 hypothetical genes excluded from further analysis. Additionally, 3,149 accessory genes were identified in 15–95% of the genomes. The significant genomic diversity revealed by this open pangenome analysis highlights the extensive variability within *P. aeruginosa* (Figure 1). The figures were generated using R. This variability provides crucial insights into strain-specific adaptations, pathogenicity, and antibiotic resistance. Furthermore, identifying universally conserved targets among the core genes points to promising candidates for vaccine development applicable across diverse *P. aeruginosa* strains. These findings are pivotal in guiding future research and therapeutic strategies.

3.3 Subcellular localization and virulence prediction

PSORTb analysis identified three outer membrane proteins, while the remaining proteins were classified as cytoplasmic or belonging to other categories (Table 1). Subsequent BLASTP analysis against the human proteome in NCBI showed no homologous hits for the outer membrane proteins, ensuring their specificity and minimizing the risk of cross-reactivity in vaccine development. BLASTP analysis against the VFDB revealed that only the PAL_1 protein matched known virulence factors, confirming its potential as a relevant target for further therapeutic or vaccine development. Further analysis of PAL_1 against the *P. aeruginosa* database identified the protein as LptF, with an e-value of 0. To validate the conservation of the selected vaccine target LptF across diverse *P. aeruginosa* strains, a multiple sequence alignment was performed using LptF sequences from 864 genomes using Python (76). The conservation analysis revealed that over 98% of the amino acid positions were fully conserved (with 100% identity), and a pairwise sequence identity of greater than 99% was observed among all strains. A corresponding heatmap of the pairwise identity matrix further confirmed the uniform conservation pattern (Supplementary Figure S1). These results underscore the evolutionary stability of LptF and support its candidacy as a universal target for vaccine or therapeutic development.

3.4 Analysis of immunogenic and physicochemical characteristics

The ProtParam tool was used to predict the physicochemical characteristics of the LptF protein. It has a molecular weight of 28.5 kDa and displays slight instability under standard laboratory conditions, with an instability index of 42.30. The GRAVY index of -0.574 indicates its hydrophilic nature. With an aliphatic index of 80.15, which reflects the protein's thermostability, LptF is considered a strong candidate for vaccine development due to its stability at physiological temperatures. Its potential as an immunogenic candidate is further supported by an antigenicity score of 0.6442 (classified as likely antigenic with a threshold of 0.4) and its classification as non-allergenic by AllerTOP.

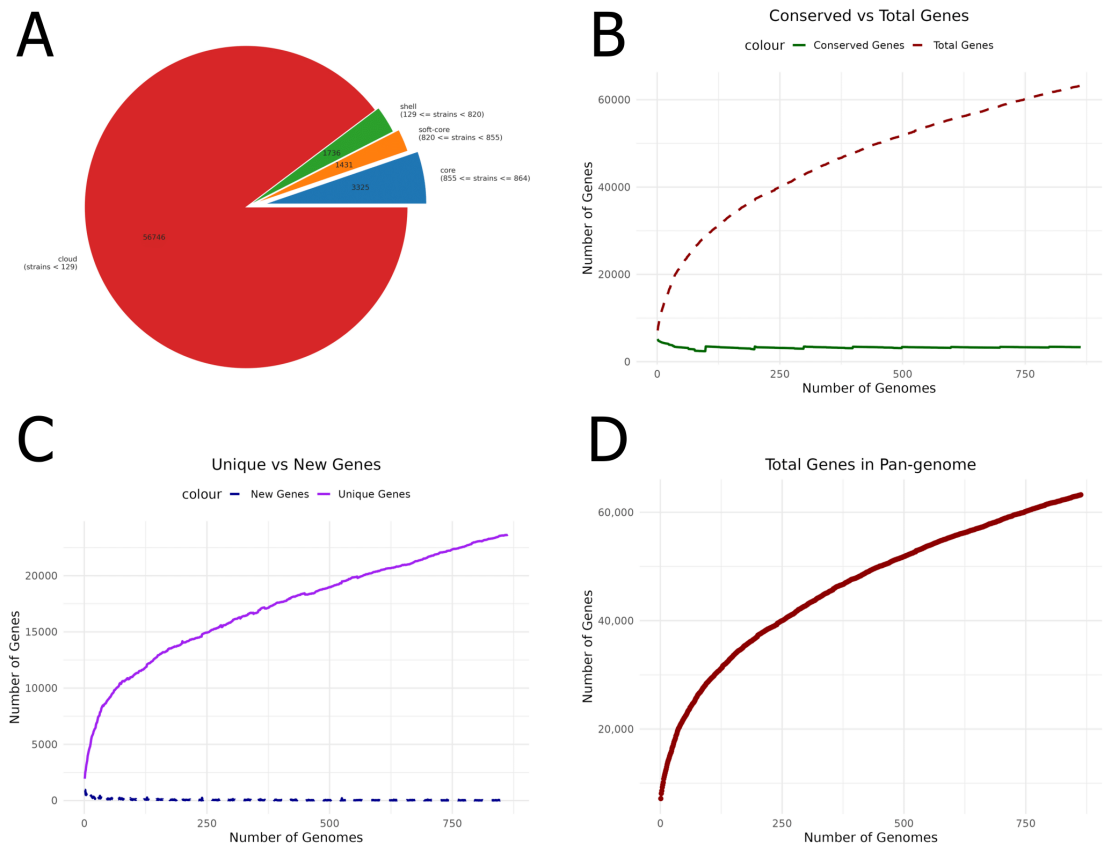


FIGURE 1
Pangenome analysis of *P. aeruginosa* genomes. **(A)** Pie chart showing the distribution of core, accessory, and unique genes. **(B)** Comparison of conserved genes with the total number of genes, highlighting genetic conservation across genomes. **(C)** Unique versus new gene ratio, emphasizing genome variability. **(D)** The number of genes identified within the pangenome provides insights into overall genomic diversity and its potential impact on vaccine development.

3.5 Signal peptide prediction

The analysis identified a Sec/SPII cleavage site at position 20 of the protein sequence, indicating the presence of a signal peptide that is likely cleaved during the maturation process via the Sec-dependent secretion pathway or the Sec/SPII system (Figure 2). With the signal peptide removed, the mature protein sequence begins at position 20. The signal peptide was excluded from further analysis, and the mature protein sequence was used in subsequent bioinformatics analyses. This sequence underwent secondary structure prediction, functional annotation, and potential epitope mapping, all of which are essential for understanding the protein's biological function and its potential use in vaccine design. This approach ensures that only the biologically relevant mature protein is considered for downstream analyses.

3.6 Prediction of B-cell epitope

The BepiPred Linear Epitope Prediction 2.0 tool was initially used to predict B-cell epitopes, identifying nine epitopes for the LptF protein. One of these epitopes, a 72-mer, was re-analyzed to ensure no potential epitopes were missed. This re-evaluation revealed eight additional epitopes (Figure 3), with figures generated in R (77). They were carefully selected based on several critical factors to confirm the suitability of the identified epitopes for vaccine development. VaxiJen v2.0 predicted high antigenicity scores for these epitopes, indicating their potential to trigger a robust immune response. Additionally, their non-toxic nature was confirmed using ToxinPred, ensuring they would not cause adverse effects. The non-allergenic properties of the epitopes were verified using AllerTOP v2.0, further ensuring their safety. As shown in Table 2, the selected epitopes were chosen for further research after careful consideration of these factors. Supplementary Table S2 provides a detailed analysis of the epitope's suitability for inclusion in potential vaccine formulations, including their toxicity, allergenicity, and antigenicity profiles.

3.7 Prediction of T-cell epitope (MHC-I and MHC-II)

The MHC-I and MHC-II epitopes were predicted for the LptF protein sequence using NetMHCpan 4.1 from IEDB. The finalized epitopes are presented in Tables 3, 4, with detailed T-cell epitope analyses in Supplementary Tables S3, S4. While VaxiJen v2.0 is primarily designed for complete proteins, it was utilized here as an additional tool to assess the antigenicity of both MHC class I and II T

TABLE 1 Subcellular localization predictions for selected proteins based on PsortB analysis.

Sno	Protein	PSORTb result
1	oprB	OuterMembrane – 10.00
2	bamB	OuterMembrane – 10.00
3	pal_1	OuterMembrane – 10.00

cell epitopes, supporting selection alongside MHC binding, immunogenicity, and toxicity criteria. Initially, 6,265 MHC-I epitopes were predicted and filtered based on a rank cutoff of 0.5 and a core score cutoff of 0.60. Similarly, 6,130 MHC-II epitopes were filtered using a rank cutoff of 2 and a score of 0.75. These thresholds were chosen because lower rank and score values indicate a higher binding affinity to MHC alleles, which is crucial for identifying effective immunogenic candidates. The finalized epitopes were further assessed for toxicity, antigenicity, and allergenicity to confirm their immunogenic potential while minimizing the risk of adverse reactions. All selected epitopes were predicted to be IL-10 inducers, suggesting their potential to regulate immune responses and prevent excessive inflammation. Notably, epitope 3 exhibited balanced induction of IL-4, IL-6, IL-10, and IL-13, making it a strong vaccine candidate. Epitopes 2 and 4 also induced IL-6 alongside IL-10, supporting a mixed pro-inflammatory and regulatory profile (Supplementary Table S5). Interferon-γ scores were computed for MHC-II epitopes to rank those that could elicit a strong immunological response. The chosen MHC-I and MHC-II epitopes, identified according to these criteria, are presented in Table 4.

3.8 Vaccine design and conservancy evaluation

Two vaccine constructs were developed, incorporating adjuvants such as RS-09 and Beta-defensin and with the predicted epitopes from the LptF protein. Each vaccine included the selected epitopes, comprising five MHC-II, four MHC-I, and five B-cell epitopes. Fifteen epitopes were incorporated into the final vaccine constructs (Figure 4). The sequences and corresponding lengths of both constructions are described in Table 5, and the proposed vaccines ranged from 248 to 283 amino acids. The combined term for these constructs was POA_V. The presence of the chosen epitopes in *P. aeruginosa* was verified using a BLASTP analysis. The results demonstrated 100% sequence similarity across *P. aeruginosa* strains, indicating that the chosen epitopes are conserved and present in all strains.

3.9 Analysis of population coverage

Based on estimated population coverage, the vaccine could potentially reach 87.35% of the global population. Tables 3, 4 comprehensively analyse the epitope distribution, demonstrating its adaptability across different regions and demographic groups. Additionally, Figure 5 visually represents the global coverage, underscoring the vaccine's potential for widespread impact (Supplementary Table S6, Supplementary Figure S2). Notably, regions such as the United States (98.33%), Kenya (98.58%), Germany (98.26%), Brazil (97.93%), France (98.04%), and Canada (95.58%) showed high predicted population coverage, emphasizing the vaccine's potential effectiveness across diverse geographic and genetic backgrounds. Moderate coverage was observed in countries such as India (85.56%), Japan (87.60%), and China (89.81%), further confirming the vaccine's adaptability in densely populated and genetically diverse regions. On the other hand, lower coverage was observed in regions such as the United Kingdom (56.38%), Hong Kong (56.64%), and American Samoa (56.40%), which may be attributed to regional HLA allele distribution patterns. Overall, the population coverage analysis strongly supports the broad usability and potential of the designed vaccine to fight the targeted pathogen worldwide.

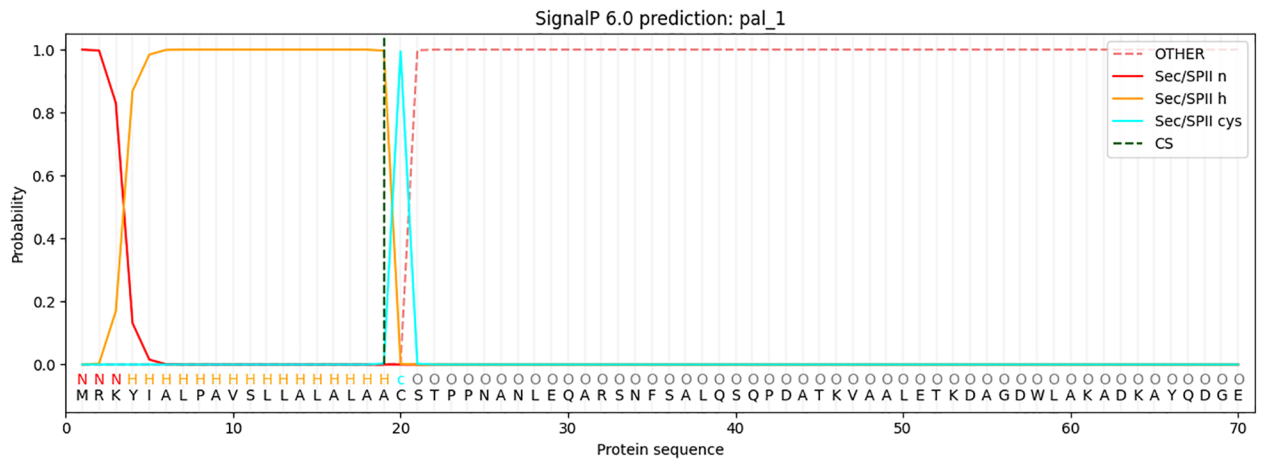


FIGURE 2
Signal peptide prediction for the LptF protein sequence was performed using SignalP.

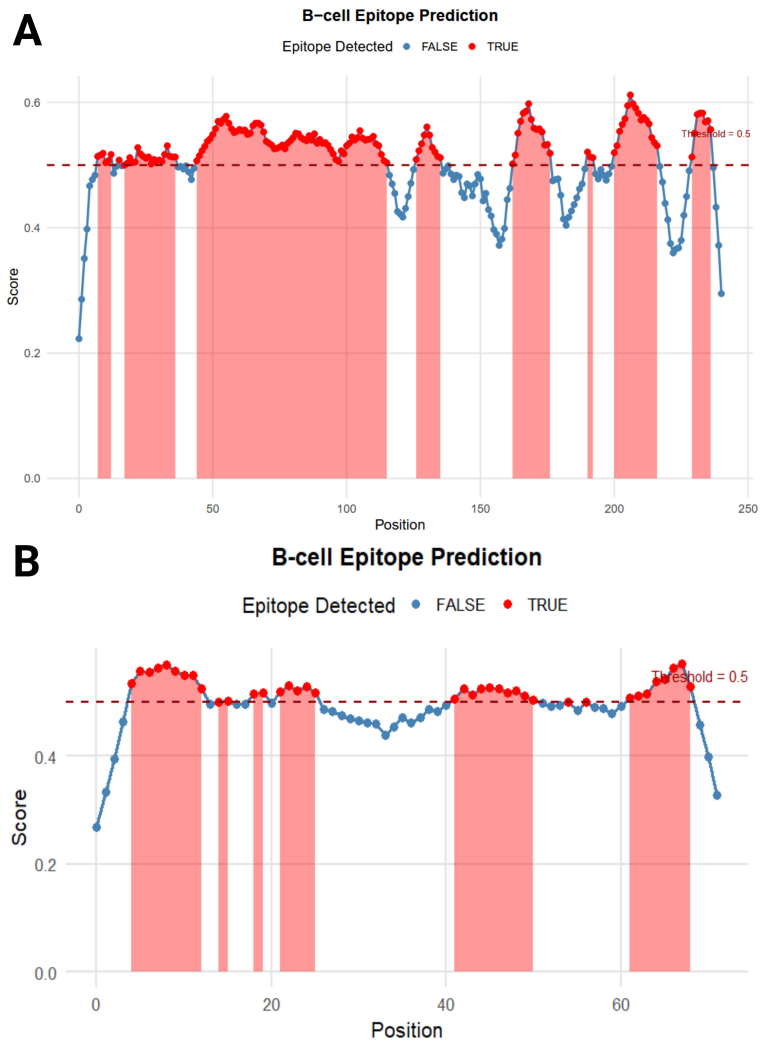


FIGURE 3
B-cell Epitope Prediction. **(A)** depicts the prediction of epitopes from the entire protein sequence, while **(B)** highlights the re-analyzed 72-mer epitope to ensure no potential epitope is overlooked.

3.10 Physicochemical property analysis

The physicochemical evaluation of the vaccine candidates POA_V_RS09 and POA_V_BDEF underscores their potential viability. POA_V_RS09, with a molecular weight of 25,734.39 Da comprising 248 amino acids, has an isoelectric point (pI) of 9.43. It exhibits hydrophilicity, as indicated by a GRAVY score of -0.856, and is considered stable with an instability index of 23.29. Similarly, POA_V_BDEF has a molecular weight of 29,763.15 Da, a pI of 9.53, and consists of 283 amino acids. Its instability index, 26.34, also suggests stability, and the GRAVY score of -0.761, which confirms its hydrophilic nature. These favorable stability and solubility

TABLE 2 Predicted B-cell epitopes for the LptF protein, identified as potential targets for vaccine development.

No	Start	End	Peptide	Length	Antigenicity Score	Probable Antigen	Allergenicity	Toxicity
8	201	217	YGKEYPVASNGTSSGRA	17	1.3767	Antigen	Non-allergen	Non-toxic
5	127	136	DLDKSDLKPG	10	1.1824	Antigen	Non-allergen	Non-toxic
3	18	37	LQSQPDATKVAALETKDAGD	20	0.7914	Antigen	Non-allergen	Non-toxic
1	5	13	GEDQRDVDQ	9	1.4255	Antigen	Non-allergen	Non-toxic
5	42	51	SAQRAQARLD	10	1.2283	Antigen	Non-allergen	Non-toxic
8	62	69	SQLNAKQT	8	1.4671	Antigen	Non-allergen	Non-toxic

TABLE 3 Finalized MHC-I epitopes identified for the LptF protein.

Allele	Length	Peptide	Score	Rank	Antigenicity Score	Antigen	Allergenicity	Toxicity
HLA-A*01:01	10	YTDSTGSANY	0.9955	0.01	1.3013	Antigen	Non-Allergen	Non-Toxin
HLA-B*57:01	10	QTSRGTMVTF	0.7892	0.22	0.5176	Antigen	Non-Allergen	Non-Toxin
HLA-B*40:01	10	GEDQRDVDQL	0.6866	0.16	1.0027	Antigen	Non-Allergen	Non-Toxin
HLA-A*31:01	10	KSDLKPGAMR	0.6452	0.19	0.9618	Antigen	Non-Allergen	Non-Toxin

properties render both candidates promising for further validation as vaccine prospects.

3.11 Secondary structure, tertiary structure, and refinement

The secondary structure of the vaccines was estimated using the PSIPRED approach, concentrating on the ratios of coils, β-sheets, and α-helices due to their immunogenic potential. PSIPRED's analysis of the vaccine candidates POA_V_RS09 and POA_V_BDEF revealed distinct structural features. POA_V_RS09 comprised 58.87% alpha helices, 4.03% beta strands, and 37.10% random coils. In contrast, POA_V_BDEF consisted of 50.53% alpha helices, 7.07% beta strands, and 42.40% random coils (Supplementary Figure S3). These findings indicate a predominance of alpha helices and a significant presence of coils in both candidates, with a relatively low content of beta strands. This structural profile suggests a balance between stability and flexibility, which is beneficial for antigenic presentation in vaccine design. Using ROBETTA and AlphaFold, we modelled the vaccine's 3D structures. Following structure generation, we refined all the models using GalaxyRefine to improve stereochemical accuracy. Among the generated models, Model 1 demonstrated superior performance for both vaccines, with RMSD values ranging from 0.9744 to 0.9889. Further validation was conducted using QMEAN4 scores and Ramachandran plot analysis to evaluate the structural quality at both global and local levels. For the POA_V_BDEF construct, the ROBETTA model yielded a QMEAN4 score of –0.72, while the AlphaFold model scored –2.67. Similarly, for the POA_V_RS09 construct, the ROBETTA model scored –0.19, compared to –2.24 for the AlphaFold prediction. QMEAN4 integrates four structural descriptors and is widely used to evaluate model quality in the absence of a native structure. These results indicate that the ROBETTA-generated models exhibit superior reliability and structural accuracy for both constructs. Further structural assessment using Ramachandran plot analysis (Table 6) revealed that ROBETTA models have over 96% of residues in favored regions, with only 0.71–0.81% falling in the disallowed areas. In contrast, AlphaFold models had a higher percentage of disallowed residues (up to 2.44%), particularly in functionally important loops and epitope-accessible regions. While AlphaFold has shown remarkable success in protein structure prediction and has been used in several recent vaccine design studies with promising results (78), we opted for ROBETTA-refined models in our research. This decision was based on comparative structural validation, where ROBETTA constructs exhibited fewer steric clashes and better Ramachandran statistics. Therefore, the ROBETTA-generated models were chosen for both POA_V_RS09 and POA_V_BDEF constructs and used in all downstream docking and immunological simulations to ensure structural reliability and predictive robustness. The Ramachandran plot of POA_V_RS09 and POA_V_BDEF shows that the structural value exceeds 90% of residues in favored regions, indicating a good overall geometry (Supplementary Figure S4). For the POA_V_BDEF construct, residues like Ser-26 and Gly-134 were located in disallowed areas, while for the POA_V_RS09 construct, residues Pro-100 and Gly-141 were also found in similar disallowed areas. The vaccine models are detailed in Supplementary Figure S5.

TABLE 4 Finalized MHC-II epitopes identified for the LptF protein, optimized for vaccine design.

Allele	Peptide	Score	Rank	Antigenicity Score	Antigen	Allergenicity	IFN-γ Score
HLA-DRB1*03:01	VEVTISNDAKPVAPR	0.9766	0.05	0.4458	Antigen	Non-Allergen	0.0856
HLA-DQA1*01:02/DQB1*06:02	VLRNAEAQLQNASAQ	0.8975	0.01	0.7313	Antigen	Non-Allergen	0.5321
HLA-DRB1*01:01	EAQLQNASAQRAQAR	0.8624	0.61	1.3141	Antigen	Non-Allergen	0.7388
HLA-DQA1*01:02/DQB1*06:02	IVLRNAEAQLQNASA	0.8601	0.03	0.7121	Antigen	Non-Allergen	0.2149
HLA-DQA1*05:01/DQB1*03:01	EAQLQNASAQRAQAR	0.8171	0.31	1.3141	Antigen	Non-Allergen	0.7388
HLA-DQA1*01:02/DQB1*06:02	TIVLRNAEAQLQNAS	0.7961	0.10	0.5276	Antigen	Non-Allergen	0.2391
HLA-DQA1*01:02/DQB1*06:02	EAQLQNASAQRAQAR	0.7768	0.13	1.3141	Antigen	Non-Allergen	0.7388

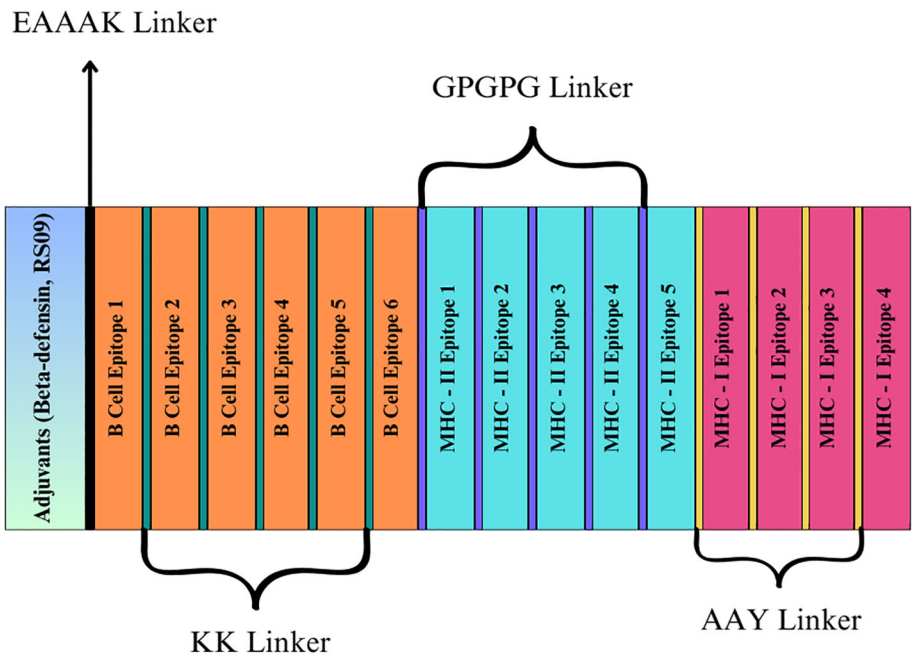


FIGURE 4
Graphical representation showcasing the formulation of the designed vaccine.

3.12 Molecular docking analysis

The HDOCK server performed docking tests to evaluate the interactions between the suggested vaccine candidates POA_V_RS09 and POA_V_BDEF and the immunological receptors TLR2 and TLR4, respectively. These receptors are vital in recognizing pathogen-associated molecular patterns (PAMPs) and triggering immune responses, such as cytokine production and the recruitment of immune cells. The results indicated strong binding affinities for all complexes, with POA_V_RS09 achieving the highest docking scores of -310.2 (kcal/mol) for TLR4 and -286.76 (kcal/mol) for TLR2 (Table 7, Figure 6). The MD simulations were conducted to further validate the interactions by examining the stability and conformational behavior of the docked complexes under physiological conditions. Both vaccine candidates exhibited stable interactions, with minimal fluctuations at the receptor-binding interface, suggesting their ability to engage immune receptors and potentially elicit robust immune responses effectively. These findings highlight the promising immunogenic potential of the designed vaccines.

3.13 Molecular dynamics simulation analysis

The MD simulations for the vaccine complexes (POA_V_RS09 and POA_V_BDEF) with TLR2 and TLR4 were conducted over 1000 ns and revealed notable differences in stability and interaction properties (Table 8). The RMSD (backbone) value indicated that the POA_V_RS09 vaccine complex was the most stable, with the TLR4_POA_V_RS09 complex showing the lowest RMSD (0.57 ± 0.06 nm), followed by TLR2_POA_V_RS09 (0.80 ± 0.19) nm. These complexes remained stable throughout the 1000-ns MD simulation. In contrast, the POA_V_BDEF-based vaccine

TABLE 5 Amino acid sequences and sequence lengths of the finalized vaccine constructs.

POA_V	Sequence	Length
POA_V_BDEF	FTQGISNPSSCRRNRGFCLAFWCPGSMRQIGTCFGFPVKCCREAAAKSQLNAKQTKKGEDQR DVDQKKYGKEYPVASNGTSSGRAKKSAQRAQARLDKKDLDKSDLKPGKKLQSQPDATKVAAL ETKDAGDGPGPGVEVTISNDAKPVAPRGPGPGVLRNAEAQLQNASAQGPGPGEAQLQNASAQ RAQARGPGPGIVLRNAEAQLQNASAGPGPGTIVLRNAEAQLQNASAAYYTDSTGSANYAAYQT SRGTMVTFAAYGEDQRDVDQLAAYKSDLKPGAMR	283
POA_V_RS09	APPHALSEAAAKSQLNAKQTKKGEDQRDVDQKKYGKEYPVASNGTSSGRAKKSAQRAQARLDK KDLDKSDLKPGKKLQSQPDATKVAALETKDAGDGPGPGVEVTISNDAKPVAPRGPGPGVLRNAE AQLQNASAQGPGPGEAQLQNASAQRAQARGPGPGIVLRNAEAQLQNASAGPGPGTIVLRNAEA QLQNASAAYYTDSTGSANYAAYQTSRGTMVTFAAYGEDQRDVDQLAAYKSDLKPGAMR	248

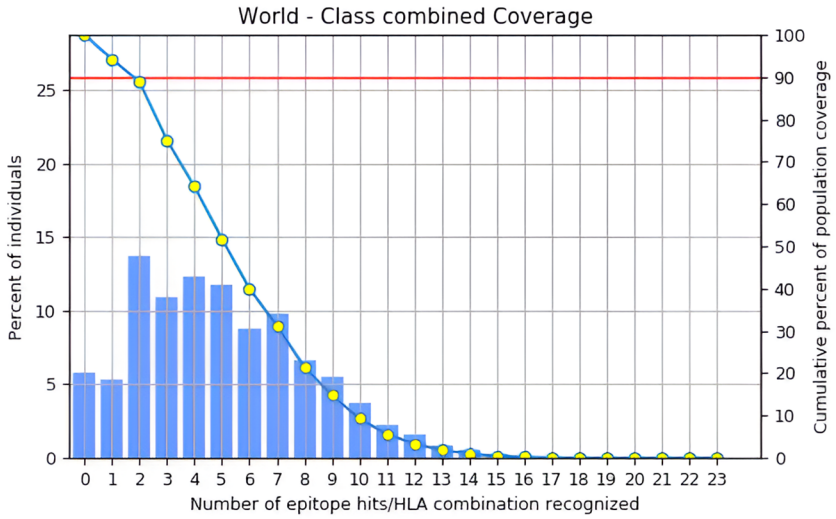


FIGURE 5
Global population coverage of the designed vaccine was analyzed using the IEDB tool, considering HLA allele frequencies across regions.

complexes had higher RMSD values, with TLR2_POA_V_BDEF (1.08 ± 0.14 nm) and TLR4_POA_V_BDEF (1.03 ± 0.10 nm), indicating more significant structural deviations and less stable interactions (Figure 7A). These complexes slightly fluctuated at the beginning of the MD simulation (0-200ns), and later they equilibrated at 1 nm. To investigate the observed fluctuations, we analyzed the backbone RMSD and Calpha RMSF of TLR2 and TLR4. Both receptors exhibited considerable structural stability, with average RMSD values of TLR2 in POA_V_BDEF at (0.33 ± 0.03 nm), TLR2 in POA_V_RS09 at (0.40 ± 0.07 nm), TLR4 in POA_V_BDEF at (0.26 ± 0.03 nm), and TLR4 in POA_V_RS09 at (0.23 ± 0.03 nm). The RMSF profiles also indicated stable conformations across all complexes, TLR2 in POA_V_BDEF at (0.15 ± 0.08 nm), TLR2 in POA_V_RS09 at (0.16 ± 0.14 nm), TLR4 in POA_V_BDEF at (0.15 ± 0.07 nm), and TLR4 in POA_V_RS09 at (0.14 ± 0.07 nm), as illustrated in Supplementary Figure S6. The predicted POA_V_BDEF complex displayed enhanced flexibility, primarily attributed to the presence of less structured epitope and linker regions Supplementary Figure S7. This inherent structural looseness likely accounts for the comparatively elevated average RMSD observed across its associated complexes. When coming to the vaccine stability in residue wise, RMSF (Calpha) analysis showed that the RS09 vaccine complexes were more rigid, with the TLR4_POA_V_RS09 complex showing the lowest RMSF (0.21 ± 0.07 nm) and the TLR2_POA_V_RS09 complex showing (0.32 ± 0.18 nm), indicating minimal flexibility at the interaction interface and that all the residues were around 0.5nm. Conversely, the POA_V_BDEF-based vaccine complexes had higher RMSF values, with TLR2_POA_V_BDEF (0.64 ± 0.23 nm) and TLR4_POA_V_BDEF (0.47 ± 0.25 nm), suggesting increased

TABLE 6 Structural validation of POA_V_RS09 and POA_V_BDEF-based vaccine models generated using Robetta and AlphaFold.

Metric	Robetta (RS09)	AlphaFold (RS09)	Robetta (BDEF)	AlphaFold (BDEF)
Total residues	248	248	283	283
Favored regions	96.34%	91.46%	96.09%	92.53%
Allowed regions	2.85%	6.10%	3.20%	5.69%
Disallowed regions	0.81%	2.44%	0.71%	1.78%
Disallowed residues	Pro-100, Gly-141	Asp-28, Asp-30, Pro-138, Glu-227, Asp-228, Gln-229	Ser-26, Gly-134	Val-75, Leu-112, Pro-175, Gly-261, Gly-280
Quality	Better geometry & fewer outliers	More outliers	Better geometry	More outliers
QMEANDisCo Global Score	0.41 ± 0.05	0.55 ± 0.05	0.52 ± 0.05	0.48 ± 0.05
QMEAN	−0.19	−2.24	−0.72	−2.67

TABLE 7 Molecular docking scores of POA_V_BDEF and POA_V_RS09 with TLR2 and TLR4, showing binding affinities.

Rank	Docking Score (kcal/mol)	Confidence Score	Interface residues	Complex
1	-299.98	0.9526	model_1	TLR2 - POA_V_BDEF
1	-286.76	0.9391	model_1	TLR2 - POA_V_RS09
1	-305.23	0.9571	model_1	TLR4 - POA_V_BDEF
1	-310.2	0.961	model_1	TLR2 - POA_V_RS09

flexibility and dynamic behaviour (Figure 7B). In contrast, the POA_V_BDEF vaccine construct exhibited pronounced fluctuations, particularly in regions interacting with TLR4 and TLR2. For the TLR4_POA_V_BDEF complex, notable flexibility was observed in the N-terminal linker region (residues 53–57), as well as in combined epitope and linker segments spanning residues 50–85, 165–180, and 207–216, in addition to the C-terminal end. Similarly, the TLR2_POA_V_BDEF complex showed continuous fluctuation across the linker (55–60), the epitope region (73–83), the extended linker–epitope stretch (109–150), and residues 173–180 and 195–220, along with the C-terminal region. In contrast to the POA_V_BDEF construct, the POA_V_RS09-based vaccine formulation demonstrated notably greater structural stability.

PCA was performed on the vaccine constructs extracted from their respective TLR2 and TLR4 complexes to evaluate their conformational dynamics. The POA_V_RS09 construct, when analyzed post-interaction with both TLR2 and TLR4, exhibited compact PCA clusters, indicating limited conformational fluctuations and stable structural behavior throughout the 1000 ns simulation. In contrast, the POA_V_BDEF construct displayed broader dispersions in PCA space, suggesting greater structural flexibility and reduced conformational stability. This trend remained consistent when the standalone vaccine models were analyzed, where POA_V_RS09 continued to show tight clustering and structural integrity, while POA_V_BDEF exhibited higher variability. These results align with earlier RMSD and RMSF analyses, collectively highlighting POA_V_RS09 as the more stable and potentially immunogenic vaccine candidate. (Supplementary Figure S8). Hydrogen bond (HBOND) analysis was conducted over the 1000 ns molecular dynamics simulation using GROMACS. The default criteria were used, which include a donor–acceptor distance cutoff of 0.35 nm and a hydrogen–donor–acceptor angle cutoff of ≥150° (i.e., ≤30°

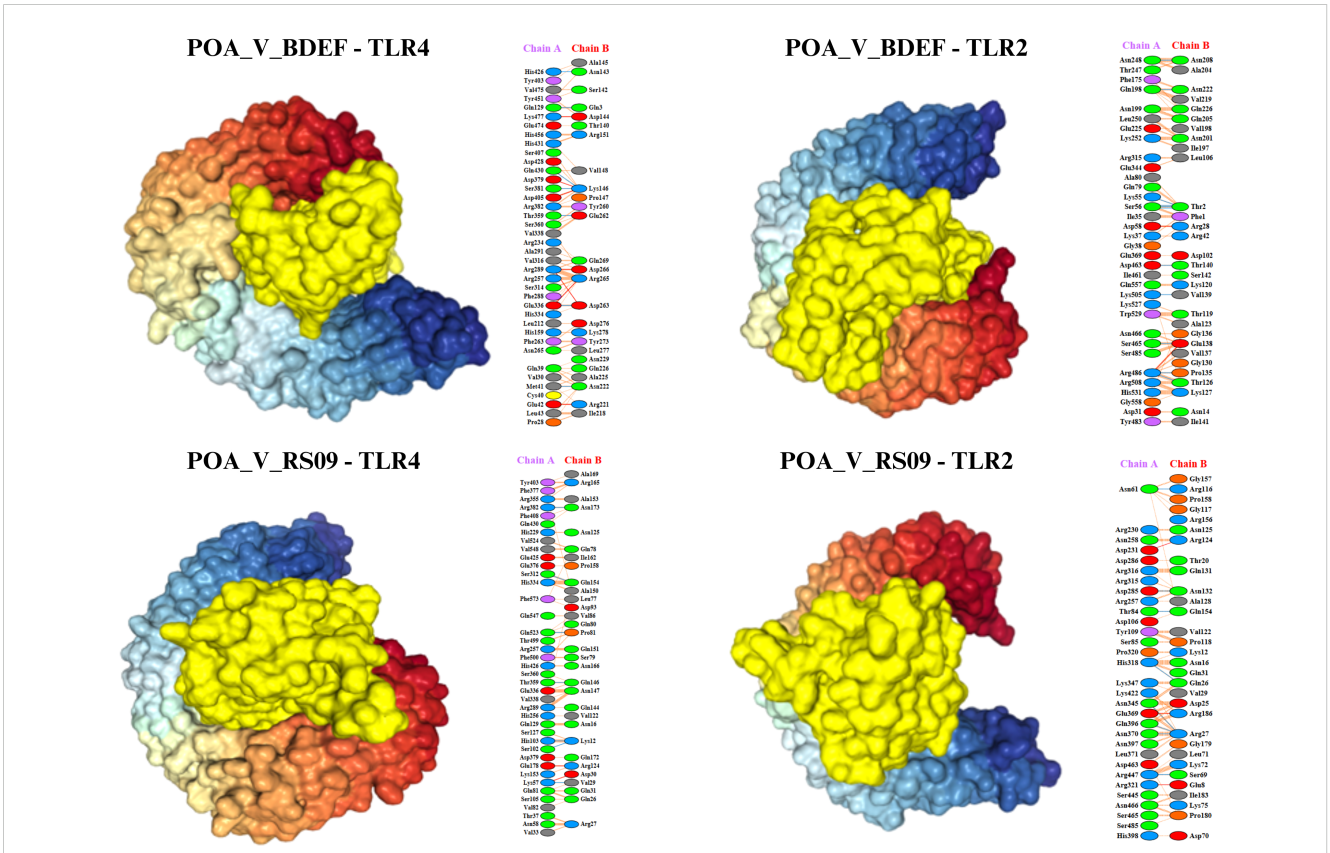


FIGURE 6
Illustrates the docking models of the vaccine constructs with the receptor, focusing on the lowest binding energy conformations. The identified interaction residues reveal strong binding affinities, highlighting critical contacts that contribute to the complex's stability and potential efficacy.

TABLE 8 Post-MD analysis averages for protein-protein complexes, including RMSD, RMSF, and H-bond values, reflecting structural stability and interactions.

Complexes	RMSD (nm)	RMSF (nm)	Avg. H-bond
POA_V_RS09_TLR2	0.80 ± 0.19	0.32 ± 0.18	11
POA_V_RS09_TLR4	0.57 ± 0.06	0.21 ± 0.07	10
POA_V_BDEF_TLR2	1.08 ± 0.14	0.64 ± 0.23	19
POA_V_BDEF_TLR4	1.03 ± 0.10	0.47 ± 0.25	14
POA_V_RS09_APO	0.83 ± 0.10	0.26 ± 0.13	–
POA_V_BDEF_APO	1.75 ± 0.25	1.07 ± 0.35	–

deviation from linearity), consistent with established definitions for biologically relevant hydrogen bonds. The analysis focused specifically on the intermolecular hydrogen bonds formed between the vaccine constructs and the TLR receptors. The POA_V_BDEF vaccine complexes exhibited a higher average number of hydrogen bonds (19 with TLR2 and 14 with TLR4) compared to the POA_V_RS09 complexes (11 with TLR2 and 10 with TLR4). However, the relatively higher RMSD and RMSF values observed in the POA_V_BDEF complexes suggest that these additional hydrogen bonds may be less stable or more transient (Figure 7C). The buried surface area (BSA) during the 1000 ns simulation at the interface of the TLR4_POA_V_RS09 complex was 42.09 nm² ± 3.98, indicating stable interactions and low variability. This was closely followed by TLR4_POA_V_BDEF, with a BSA of 41.12 nm² ± 5.83, showing a similar interaction pattern. In contrast, TLR2_POA_V_BDEF had a higher BSA of 52.37 nm² ± 6.60, while TLR2_POA_V_RS09 showed a BSA of 35.70 nm² ± 5.19, both with higher standard deviations, suggesting relatively fewer stable interactions (Figure 7D). Analysis of the apo forms revealed that POA_V_BDEF exhibited the highest RMSD (1.75 ± 0.25 nm) and RMSF (1.07 ± 0.35 nm), indicating significant conformational flexibility in the absence of receptor binding. In contrast, POA_V_RS09 exhibited lower deviation (0.83 ± 0.10 nm RMSD and 0.26 ± 0.13 nm RMSF), suggesting it remains relatively stable even when unbound. The FEL analysis effectively showed us the structural stability and flexibility of the vaccine-protein complexes. All complexes exhibited energy basins, indicating the presence of metastable states. However, notable differences were observed in the shape and depth of these energy wells. Complexes involving TLR4 exhibited more compact and deeper energy minima compared to those involving TLR2, suggesting a higher degree of structural stability. In particular, the TLR4_POA_V_RS09 complex exhibited a well-defined global minimum, indicating a stable and energetically favorable conformation throughout the simulation. Although the TLR4_POA_V_BDEF complex also reached stable conformations, it showed slightly more conformational variability. Conversely, the TLR2 complexes exhibited broader and more scattered low-energy regions, indicating increased conformational flexibility and less stable interaction patterns. Among them, the TLR2_POA_V_RS09 complex exhibited relatively smoother energy transitions compared to TLR2_POA_V_BDEF, which displayed more rugged features in its energy landscape (Supplementary Figure S9).

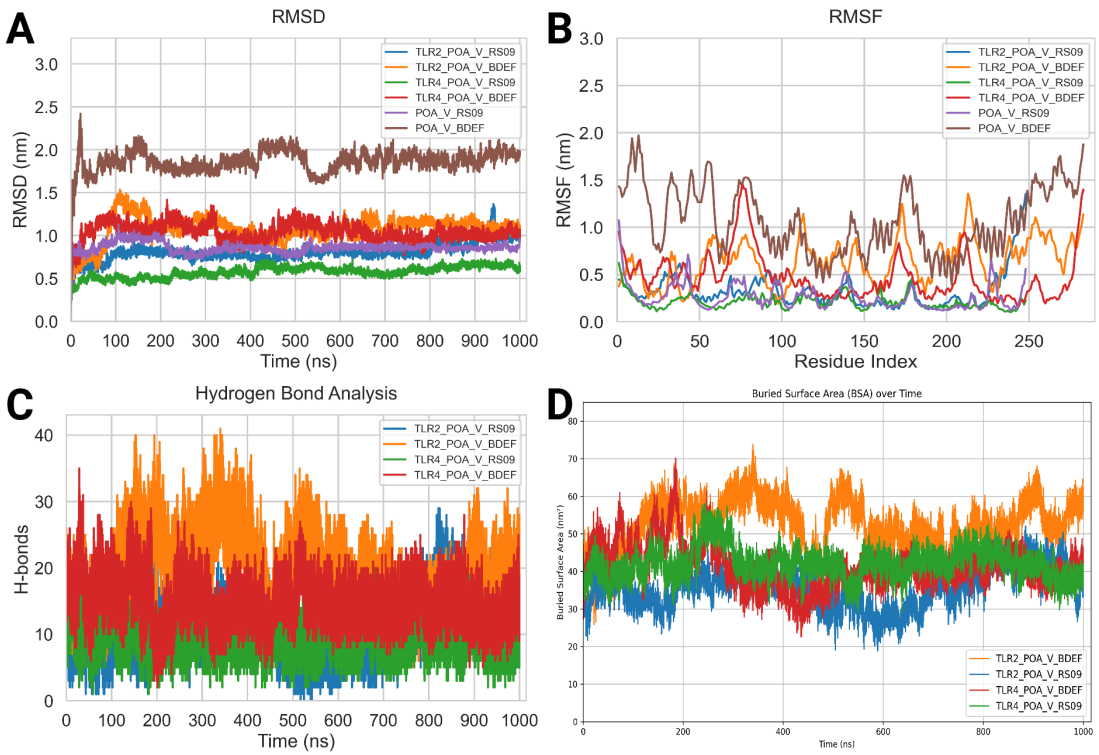


FIGURE 7
Molecular dynamics (MD) analysis results. **(A)** Root mean square deviation (RMSD) Backbone analysis. **(B)** Root mean square fluctuation (RMSF) Calpha Analysis. **(C)** Hydrogen bond (HBOND) analysis. **(D)** Buried surface analysis (BSA).

We calculated the binding free energy using MMPBSA, which revealed notable differences between the POA_V_RS09 and POA_V_BDEF-based vaccine constructs in complex with TLR2 and TLR4 receptors. The TLR2_POA_V_RS09 showed a more favorable binding energy (–1483.14 kJ/mol) than TLR2_POA_V_BDEF (–1335.16 kJ/mol), indicating that RS09 forms a more stable and energy-efficient complex with TLR2, compared to POA_V_BDEF. In contrast, TLR4_POA_V_BDEF exhibited a significantly stronger binding energy (–4600.83 kJ/mol) than TLR4_RS09 (–2682.66 kJ/mol), likely due to its extended area, which enables an increased contact surface. However, prior dynamic and structural analyses, such as RMSF and FEL plots, indicate that POA_V_BDEF is more flexible, particularly at the linker and epitope regions. This flexibility may contribute to reduced structural stability, especially in the TLR2 complex compared to POA_V_RS09.

3.14 *In silico* codon adaptation, cloning, and immune simulation

Codon optimization was performed using GenScript to enhance the expression of the POA_V_RS09 vaccine sequence in *E. coli* K-12. With a GC content of 60.22% and a total length of 744 base pairs, the optimized sequence falls within the ideal range (30–70%) for effective expression in *E. coli*. This balanced GC content ensures efficient transcription and translation, making the sequence suitable for high-level expression in the host. The optimized vaccine sequence was then used for in-silico cloning with SnapGene software, successfully inserting the gene into the pET-28a(+) expression plasmid (Figure 8). The immune response dynamics elicited by POA_V_RS09 are shown in Figure 9. Figure 9A illustrates the antigen (Ag) and antibody responses over a 350-day period, where an early antigen peak, followed by a sharp decline, indicates effective recognition and clearance by the host immune system. This is accompanied by a strong humoral response, characterized by an initial surge in IgM, typical of a primary response, followed by a sustained increase in IgG1 and IgG2, which shows class switching and maturation of the immune response. The dominance of IgG subclasses over time reflects the development of long-term protective immunity. Notably, IgG1 and IgG2 are associated with Th1-type immune responses, which are essential for combating pathogens such as viruses and certain bacteria. Figure 9B illustrates cytokine dynamics, where high levels of Interleukin-2 (IL-2) and IFN-γ early on indicate strong T-cell activation and a Th1-biased immune response, which is particularly important for combating pathogens. Their gradual decline over time suggests immune regulation and resolution of inflammation, highlighting the vaccine's safety profile. Figure 9C illustrates the dynamics of the B-cell population, exhibiting an increasing trend in memory B cells and a shift in isotype expression from IgM to IgG, which further validates class switching and the generation of long-lasting humoral memory. Figure 9D focuses on B-cell states, showing that active and antigen-internalizing B cells peak early, while anergic cells remain relatively constant, suggesting efficient antigen processing and presentation. The low level of anergic B cells suggests that immune tolerance is not induced, further supporting the vaccine's immunogenicity. Figure 9E presents helper T-cell (TH) populations, with an initial spike in total TH cells, followed by the emergence of memory TH cells, supporting long-term immunity. Finally, Figure 9F depicts cytotoxic T-cell (TC) states, showing early activity and duplication followed by a steady increase in resting memory TC cells, which are critical for sustained immune surveillance. This can contribute to the direct killing of bacteria or infected host cells in bacterial infections that evade extracellular immune mechanisms. These outcomes underscore the successful initiation of both humoral and cellular immune responses, supporting the potential for the development of long-lasting immunological memory. Moreover, the observed immune signatures align with the protective responses typically seen in effective bacterial vaccines, validating the rational design of POA_V_RS09, which incorporates TLR agonists, multi-epitope constructs, and immune-enhancing linkers to induce broad, durable immunity against bacterial pathogens.

4 Discussion

The opportunistic pathogen *P. aeruginosa* is a significant cause of hospital-acquired infections worldwide. It presents a serious threat to human health, especially in immunocompromised individuals, due to its innate antibiotic resistance and ability to develop biofilms (79, 80). The increasing prevalence of drug-resistant strains has complicated treatment options, highlighting the urgent need for alternative therapeutic strategies (81). Despite progress in antimicrobial therapies, no licensed vaccine exists for *P. aeruginosa*, revealing a critical gap in combating this pathogen (82). Immunoinformatics has emerged as a powerful tool in vaccine development, facilitating the rational design of in silico vaccines, as demonstrated in the development of vaccines against pathogens such as the Ebola virus, SARS-CoV-2, and Mycobacterium tuberculosis (83–85). These approaches hold significant promise for addressing the challenges posed by *P. aeruginosa*.

Previous immunoinformatics-based vaccine studies against *P. aeruginosa* often relied on single-strain datasets or focused on a narrow range of targets. Some selected cytoplasmic proteins have limited surface accessibility, while others used previously known antigens without assessing their conservation across diverse strains. Additionally, several studies selected targets which is completely based on literature without genome-wide screening (86–89). Other broader approaches involving multiple pathogens have also identified shared virulence or essential gene-derived epitopes while filtering for self-tolerance (90). However, such strategies typically lack species-specific optimization, structural validation, and comprehensive strain-level genomic coverage—critical aspects that our study addresses. In this study, we conducted a comprehensive pangenome analysis (91) of 864 *P. aeruginosa* genomes. This extensive dataset enabled robust pangenome analysis and the identification of conserved, surface-exposed, and virulence-associated targets, distinguishing our study from previous investigations. Here we identified 63,239 genes, including 3,325 core genes and 3,149 accessory genes. We focused on

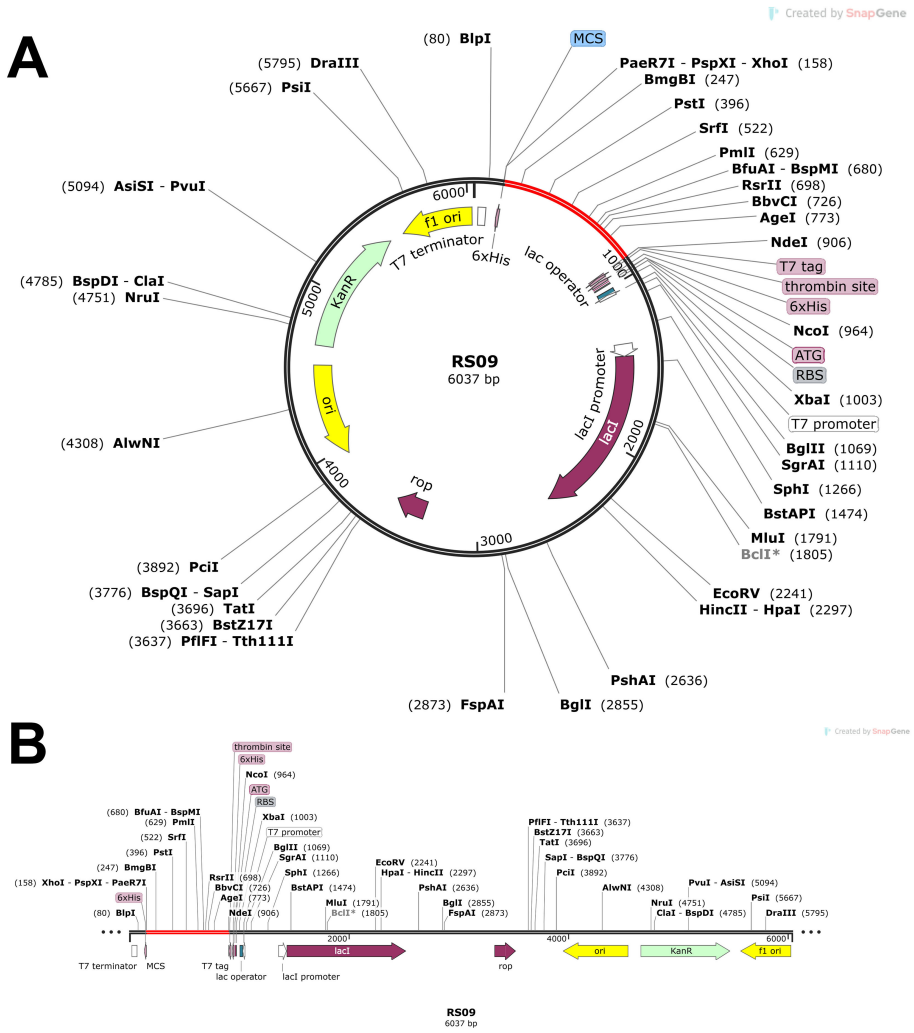


FIGURE 8
In silico cloning of POA_V_RS09. **(A)** Cloning of RS09 (which is POA_V_RS09) into the pET-28a(+) vector. **(B)** final vaccination design with additional restriction sites.

conserved core genes essential for bacterial survival and pathogenicity to ensure broad-spectrum coverage. The Pal_1 protein was identified as a potential vaccine candidate, which is classified as an outer membrane protein via PSORTb, and confirmed as a virulence-associated factor through BLASTP analysis against the VFDB. Its sequence was validated against the *P. aeruginosa* database, where it was identified with 100% confidence and an E-value of 0, and it is known as LptF (lipotoxin F). Sequence comparison with the human proteome confirmed the absence of homologous hits, minimizing the risk of adverse cross-reactivity. LptF, an OmpA-like outer membrane protein, plays a crucial role in *P. aeruginosa*'s survival, particularly in stressful environments such as lung colonization in cystic fibrosis, and may serve as an important target for therapeutic strategies (92). LptF remains an underexplored target. Its classification as a lipotoxin, along with evidence from structural proteomics revealing interactions with key membrane proteins like OprI and LptE, further highlights its relevance as a promising vaccine candidate against *P. aeruginosa*. Due to the increasing antibiotic resistance of *P. aeruginosa*, an effective vaccine is urgently needed, and lipotoxins have been identified as potential targets in studies (24). The LptF protein exhibited favorable physicochemical and immunogenic properties, with a molecular weight of 28.5 kDa, thermostability indicated by an aliphatic index of 80.15, and a hydrophilic nature reflected in a GRAVY index of -0.574. Immunogenic analysis revealed its suitability as a vaccine target, with an antigenicity score of 0.6442 and classification as a non-allergen. SignalP analysis showed that the protein has a signal peptide, which is likely removed during maturation. This was accounted for in later analyses that focused on the mature protein sequence. We identified B-cell epitopes and chose high-affinity T-cell epitopes based on their binding affinities to MHC-I and MHC-II molecules, with additional refinement based on their potential to stimulate interferon-γ production. Ultimately, 15 epitopes were incorporated into the vaccine design, comprising four MHC-I epitopes, five MHC-II epitopes, and six B-cell epitopes. Additionally, the cytokine prediction analysis revealed that all selected epitopes possess IL-10-inducing potential, a cytokine shown to be critical in controlling inflammation and enhancing bacterial

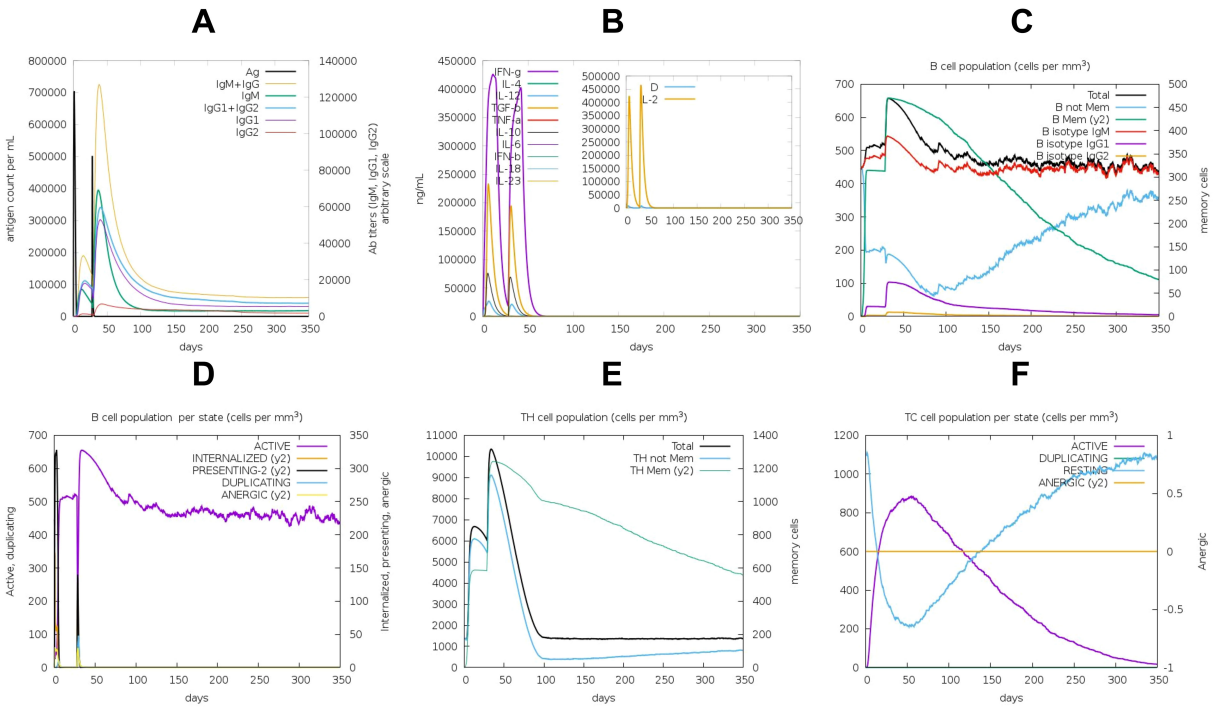


FIGURE 9
Immune Response Induced by the POA_V_RS09 Vaccine: **(A)** antibody response. **(B)** Cytokine response. **(C)** B-Cell population (cells/mm^3). **(D)** B-Cell population by state. **(E)** TH-cell population by state. **(F)** TC-cell population by state.

clearance during *P. aeruginosa* infection (93). The vaccine constructs POA_V_RS09 and POA_V_BDEF were designed, incorporating RS-09 and Beta-defensin as adjuvants. We selected RS09 as it functions as a TLR4 agonist, effectively stimulating the innate immune response. β-defensin was chosen for its dual role in activating both innate and adaptive immunity. Incorporating these adjuvants aimed to enhance the immunogenic potential of the constructs. This strategic design allowed us to evaluate and compare their impact on vaccine performance. These constructs exhibited broad global population coverage (87.35%) and strong immunogenic potential. Secondary structure analysis revealed that both vaccines predominantly consisted of α-helices and random coils, enhancing antigenic presentation. Structural validation through Ramachandran plot analysis ensured the reliability of the tertiary structures. We selected the Robetta server over AlphaFold for tertiary structure prediction because Robetta offers more reliable modeling for synthetic, chimeric constructs involving multiple domains, such as adjuvants, epitopes, and linkers. Unlike AlphaFold, which is optimized for natural protein sequences, Robetta's *de novo* prediction approach is better suited for handling novel vaccine constructs. For the BDEF-based construct, serine-26 (Ser-26) and glycine-134 (Gly-234) residues were found in disallowed regions, while the RS09-based construct showed proline-100 (Pro-100) and glycine-141 (Gly-141) residues in disallowed regions. These residues were primarily located in loop and linker regions and were not associated with key epitope or adjuvant domains, suggesting that they are unlikely to compromise the overall structural integrity of the protein. Therefore, the refined and validated vaccine structures were subsequently used for molecular docking studies to assess receptor interactions. Molecular docking analyses revealed high-affinity interactions between the vaccine constructs and Toll-like receptors TLR2 and TLR4, which play pivotal roles in initiating innate immune responses. POA_V_RS09 showed superior docking scores, with -310.2 (kcal/mol) for TLR4 and -286.76 (kcal/mol) for TLR2, while MD simulations confirmed the stability of these interactions under physiological conditions. We conducted extensive 1000-ns molecular dynamics simulations to analyze the long-term structural stability and interaction dynamics of the vaccine-receptor complexes. This extended simulation duration exceeds the standard practice in similar studies, providing deeper insights into conformational behavior, particularly in flexible regions such as linkers and epitopes. It enhances the structural validation of our vaccine constructs and reinforces the reliability of our results. The TLR4_POA_V_RS09 complex demonstrated minimal structural fluctuations, with the lowest RMSD and RMSF values, indicating stable interactions compared to POA_V_BDEF. Although POA_V_BDEF demonstrated good structural quality based on validation metrics, MD simulations revealed considerable flexibility, even in its apo form (without receptor binding). This inherent flexibility, especially in the epitope-linker regions, might weaken stable receptor binding and influence immune activation. While some mobility facilitates epitope presentation, too much fluctuation can reduce vaccine effectiveness. These findings underscore the importance of dynamic assessment in conjunction with static validation when

evaluating multi-epitope vaccine designs. Further analysis confirmed POA_V_RS09's stable conformational states, with tighter cluster dispersion and distinct energy minima. MMPBSA analysis showed that although POA_V_BDEF has a strong binding affinity with TLR4, POA_V_RS09 exhibits a more balanced and consistent interaction profile with both TLR2 and TLR4, along with fewer structural fluctuations. These qualities make POA_V_RS09 a robust and dependable adjuvant candidate for the design of multi-epitope vaccines. Codon optimization for POA_V_RS09 allowed efficient expression in *E. coli* (K-12), and in silico cloning into pET-28a(+) validated its expression potential. Immune simulations demonstrated robust adaptive immune responses, characterized by sustained IgG production, memory B-cell formation, and effective cytokine engagement, rendering POA_V_RS09 a promising vaccine candidate for long-term immunity. The POA_V_RS09 vaccine demonstrated strong stability, optimal expression potential, and robust immune activation, positioning it as an ideal candidate for further development. By targeting *P. aeruginosa*, a highly resistant pathogen, the POA_V_RS09 vaccine could offer a valuable strategy for preventing infections and addressing the global threat of antimicrobial resistance, ultimately improving patient outcomes. In this study, we designed two separate vaccine constructs using RS09 and β-defensin adjuvants to independently evaluate their immunostimulatory potential. This separation allows for comparative assessment of construct stability, population coverage, and immunogenicity. However, future studies could explore the integration of both adjuvants into a single construct, as combinatorial adjuvants have been shown to enhance immune responses more effectively than individual components (78). One limitation of this study is that it lacks experimental validation. While our computational approach provides a cost-effective and time-efficient method for epitope screening, future *in vitro* and *in vivo* studies (e.g., ELISA, ELISPOT) are essential to confirm immunogenicity and support vaccine development of the POA_V_RS09 vaccine candidate against *P. aeruginosa*. However, we have thoroughly examined the structural and immunological characteristics of the vaccine candidate through in silico methods, including 1000 ns molecular dynamics simulations, epitope mapping, TLR docking, population coverage analysis, and immunogenicity prediction. Long-timescale MD simulations allow for the capture of biologically relevant conformational changes, showing that microsecond to millisecond scale simulations can uncover protein folding pathways and slow structural transitions. This supports the use of 1000 ns MD to study dynamic molecular interactions (94). Several previous studies have demonstrated that immunoinformatics-based vaccine designs can reliably predict antigenic determinants and immune interactions, often correlating well with experimental outcomes (95–97). These findings support the translational relevance of computational predictions in the early stages of vaccine design. Further validation using comprehensive *in vitro* assays is also necessary to evaluate the safety profile and immunogenic potential of the POA_V_RS09-based vaccine, including its ability to induce pro-inflammatory cytokines, activate T cells, and generate specific antibody responses. Such investigations will provide valuable insights into the clinical feasibility of POA_V_RS09 as a vaccine candidate for *P. aeruginosa* infections.

5 Conclusion

This study utilized an integrated pangenome and immunoinformatics approach to develop an epitope-based peptide vaccine targeting *P. aeruginosa*. Through pangenome analysis, we identified LptF as a promising and underexplored vaccine target, specifically. From LptF, we predicted potential epitopes. The resulting vaccine candidate, POA_V_RS09, demonstrated promising immune response outcomes and strong binding affinity to immunological receptors (TLRs). Notably, the 1000-ns molecular dynamics simulation provided valuable insights into the structural stability of the vaccine–receptor complexes over an extended timescale, reinforcing the robustness of the construct under physiological conditions. This computational strategy holds significant potential for addressing the escalating issue of antimicrobial resistance, particularly in resource-limited settings and low-income countries. This strategy provides a comprehensive and practical approach to combating infections by targeting conserved NS proteins, identifying high-affinity B-cell and T-cell epitopes, and utilizing suitable adjuvants. Future studies should assess the vaccine's safety, effectiveness, and scalability through *in vitro* investigations, animal model testing, and ensuing clinical trials. To transform this computational framework into a valuable tool for combating *P. aeruginosa* resistance to multiple drugs, these steps are crucial.

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

SE: Validation, Writing – original draft, Formal analysis, Investigation, Visualization, Data curation, Methodology, Conceptualization, Writing – review & editing. SK: Visualization, Writing – review & editing, Conceptualization, Validation, Supervision, Methodology.

Funding

The author(s) declare that no financial support was received for the research, and/or publication of this article.

Acknowledgments

The authors express deep gratitude to the management of VIT University for all the support, assistance, and constant encouragement to carry out this work.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1617251/full#supplementary-material

References

1. Qin S, Xiao W, Zhou C, Pu Q, Deng X, Lan L, et al. Pseudomonas aeruginosa: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. *Signal Transduct Target Ther*. (2022) 7:199. doi: 10.1038/s41392-022-01056-1

2. Menon ND, Somanath P, Jossart J, Vijayakumar G, Shetty K, Baswe M, et al. Comparative molecular profiling of multidrug-resistant Pseudomonas aeruginosa identifies novel mutations in regional clinical isolates from South India. *JAC Antimicrob Resist*. (2023) 6. doi: 10.1093/jacamr/dlae001

3. De Oliveira DMP, Forde BM, Kidd TJ, Harris PNA, Schembri MA, Beatson SA, et al. Antimicrobial resistance in ESKAPE pathogens. *Clinical Microbiology Reviews* (2020) 33. doi: 10.1128/cmr.00181-19

4. Kalpana S, Lin W-Y, Wang Y-C, Fu Y, Lakshmi A, Wang H-Y. Antibiotic resistance diagnosis in ESKAPE pathogens—A review on proteomic perspective. *Diagnostics*. (2023) 13:1014. doi: 10.3390/diagnostics13061014

5. Teney C, Poupelin J-C, Briot T, Le Bouar M, Fevre C, Brosset S, et al. Phage therapy in a burn patient colonized with extensively drug-resistant pseudomonas aeruginosa responsible for relapsing ventilator-associated pneumonia and bacteriemia. *Viruses*. (2024) 16:1080. doi: 10.3390/v16071080

6. Pham TM, Büchler AC, Voor in 't holt AF, Severin JA, Bootsma MCJ, Gommers D, et al. Routes of transmission of VIM-positive Pseudomonas aeruginosa in the adult intensive care unit-analysis of 9 years of surveillance at a university hospital using a mathematical model. *Antimicrob Resist Infect Control*. (2022) 11:55. doi: 10.1186/s13756-022-01095-x

7. Organization WH. WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. (2024). Available online at: https://www.who.int/publications/i/item/9789240093461 (Accessed August 19, 2025).

8. Liao C, Huang X, Wang Q, Yao D, Lu W. Virulence factors of pseudomonas aeruginosa and antivirulence strategies to combat its drug resistance. *Front Cell Infect Microbiol*. (2022) 12:926758. doi: 10.3389/fcimb.2022.926758

9. Horna G, Ruiz J. Type 3 secretion system of Pseudomonas aeruginosa. *Microbiol Res*. (2021) 246:126719. doi: 10.1016/j.micres.2021.126719

10. Giovagnorio F, De Vito A, Madeddu G, Parisi SG, Geremia N. Resistance in pseudomonas aeruginosa: A narrative review of antibiogram interpretation and emerging treatments. *Antibiotics*. (2023) 12:1621. doi: 10.3390/antibiotics12111621

11. Kang C-I, Kim S-H, Park WB, Lee K-D, Kim H-B, Kim E-C, et al. Bloodstream infections caused by antibiotic-resistant gram-negative bacilli: risk factors for mortality and impact of inappropriate initial antimicrobial therapy on outcome. *Antimicrob Agents Chemother*. (2005) 49:760–6. doi: 10.1128/AAC.49.2.760-766.2005

12. Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis*. (2004) 39:309–17. doi: 10.1086/421946

13. Li J, Wang Y, Liu P, Zhang Y, Yang Y, Zhao S, et al. Trends and implications of antimicrobial resistance in Pseudomonas aeruginosa: Insights from a 19-year study in Zhejiang Province. *Medicine*. (2024) 103:e40606. doi: 10.1097/MD.0000000000040606

14. Yang AF, Huang V, Samaroo-Campbell J, Augenbraun M. Multi-drug resistant Pseudomonas aeruginosa: a 2019–2020 single center retrospective case control study. *Infect Prev Pract*. (2023) 5:100296. doi: 10.1016/j.infpip.2023.100296

15. Almutairy B. Extensively and multidrug-resistant bacterial strains: case studies of antibiotics resistance. *Front Microbiol*. (2024) 15:1381511. doi: 10.3389/fmicb.2024.1381511

16. Muteeb G, Rehman MT, Shahwan M, Aatif M. Origin of antibiotics and antibiotic resistance, and their impacts on drug development: A narrative review. *Pharmaceuticals*. (2023) 16:1615. doi: 10.3390/ph16111615

17. Mondal AH, Khare K, Saxena P, Debnath P, Mukhopadhyay K, Yadav D. A review on colistin resistance: an antibiotic of last resort. *Microorganisms*. (2024) 12:772. doi: 10.3390/microorganisms12040772

18. Singha B, Singh V, Soni V. Alternative therapeutics to control antimicrobial resistance: a general perspective. *Front Drug Discov*. (2024) 4:1385460. doi: 10.3389/fddsv.2024.1385460

19. Hassan R, El-Naggar W, Abd El-Aziz AM, Shaaban M, Kenawy HI, Ali YM. Immunization with outer membrane proteins (OprF and OprI) and flagellin B protects mice from pulmonary infection with mucoid and nonmucoid Pseudomonas aeruginosa. *J Microbiol Immunol Infect*. (2018) 51:312–20. doi: 10.1016/j.jmii.2016.08.014

20. Hasso-Agopsowicz M, Sparrow E, Cameron AM, Sati H, Srikantiah P, Gottlieb S, et al. The role of vaccines in reducing antimicrobial resistance: A review of potential impact of vaccines on AMR and insights across 16 vaccines and pathogens. *Vaccine*. (2024) 42:S1–8. doi: 10.1016/j.vaccine.2024.06.017

21. Costanzo V, Roviello GN. The potential role of vaccines in preventing antimicrobial resistance (AMR): an update and future perspectives. *Vaccines (Basel)*. (2023) 11:333. doi: 10.3390/vaccines11020333

22. Aroca Molina KJ, Gutiérrez SJ, Benítez-Campo N, Correa A. Genomic differences associated with resistance and virulence in pseudomonas aeruginosa isolates from clinical and environmental sites. *Microorganisms*. (2024) 12:1116. doi: 10.3390/microorganisms12061116

23. Sethi G, Varghese RP, Lakra AK, Nayak SS, Krishna R, Hwang JH. Immunoinformatics and structural aided approach to develop multi-epitope based subunit vaccine against Mycobacterium tuberculosis. *Sci Rep*. (2024) 14:15923. doi: 10.1038/s41598-024-66858-5

24. Firoved AM, Ornatowski W, Deretic V. Microarray analysis reveals induction of lipoprotein genes in mucoidPseudomonas aeruginosa: implications for inflammation in cystic fibrosis. *Infect Immun*. (2004) 72:5012–8. doi: 10.1128/iai.72.9.5012-5018.2004

25. Adila Nazli, Qiu J, Tang Z, He Y. Recent advances and techniques for identifying novel antibacterial targets. *Curr Medicinal Chem*. (2023) 31:464–501. doi: 10.2174/0929867330666230123143458

26. Blin K. ncbi-genome-download. *Zenodo*. (2023). doi: 10.5281/zenodo.8192486

27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. (2015) 31:3691–3. doi: 10.1093/bioinformatics/btv421

28. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. (2014) 30:2068–9. doi: 10.1093/bioinformatics/btu153

29. 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

30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinf*. (2009) 10:421. doi: 10.1186/1471-2105-10-421

31. Chen L. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. (2004) 33:D325–8. doi: 10.1093/nar/gki008

32. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res*. (2016) 44:D646–53. doi: 10.1093/nar/gkv1227

33. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf*. (2007) 8:4. doi: 10.1186/1471-2105-8-4

34. 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

35. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*. (1982) 157:105–32. doi: 10.1016/0022-2836(82)90515-0

36. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. (2022) 40:1023–5. doi: 10.1038/s41587-021-01156-3

37. 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

38. 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

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:e73957. doi: 10.1371/journal.pone.0073957

40. 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

41. 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

42. 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

43. Dhanda SK, Gupta S, Vir P, Raghava GPS. Prediction of IL4 inducing peptides. *Clin Dev Immunol*. (2013) 2013:1–9. doi: 10.1155/2013/263952

44. Dhall A, Patiyal S, Sharma N, Usmani SS, Raghava GPS. Computer-aided prediction and design of IL-6 inducing peptides: IL-6 plays a crucial role in COVID-19. *Briefings Bioinf*. (2020) 22. doi: 10.1093/bib/bbaa259

45. Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci Rep*. (2017) 7. doi: 10.1038/srep42851

46. Jain S, Dhall A, Patiyal S, Raghava GPS. IL13Pred: A method for predicting immunoregulatory cytokine IL-13 inducing peptides. *Comput Biol Med*. (2022) 143:105297–7. doi: 10.1016/j.compbiomed.2022.105297

47. 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 (Basel)*. (2021) 9:669. doi: 10.3390/vaccines9060669

48. Li X, Wang X, Du J, Bu X, Peng C, Duan X, et al. Applications of β-defensins against infectious pathogenic microorganisms. *Expert Rev Anti Infect Ther*. (2024) 22:501–10. doi: 10.1080/14787210.2024.2377677

49. Ruaro-Moreno M, Monterrubio-López GP, Reyes-Gastellou A, Castelán-Vega JA, Jiménez-Alberto A, Aparicio-Ozores G, et al. Design of a Multi-Epitope Vaccine against Tuberculosis from Mycobacterium tuberculosis PE_PGRS49 and PE_PGRS56 Proteins by Reverse Vaccinology. *Microorganisms*. (2023) 11:1647. doi: 10.3390/microorganisms11071647

50. Khan S, Rizwan M, Zeb A, Eldeen MA, Hassan S, Ur Rehman A, et al. Identification of a Potential Vaccine against Treponema pallidum Using Subtractive Proteomics and Reverse-Vaccinology Approaches. *Vaccines*. (2023) 11:72. doi: 10.3390/vaccines11010072

51. 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*. Humana Press, Totowa, NJ (2005). p. 571–607. doi: 10.1385/1-59259-890-0:571

52. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. (2000) 16:404–5. doi: 10.1093/bioinformatics/16.4.404

53. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*. (2019) 47:W402–7. doi: 10.1093/nar/gkz297

54. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. (2004). doi: 10.1093/nar/gkh468

55. 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

56. 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

57. Mogensen TH. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev*. (2009) 22:240–73. doi: 10.1128/CMR.00046-08

58. Shanmugam A, Rajoria S, George AL, Mittelman A, Suriano R, Tiwari RK. Synthetic toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PloS One*. (2012) 7:e30839. doi: 10.1371/journal.pone.0030839

59. Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H, Ogawa T, et al. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity*. (1999) 11:443–51. doi: 10.1016/S1074-7613(00)80119-3

60. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature*. (2009) 458:1191–5. doi: 10.1038/nature07830

61. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, 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

62. Yan Y, Tao H, He J, Huang S-Y. The HDOCK server for integrated protein–protein docking. *Nat Protoc*. (2020) 15:1829–52. doi: 10.1038/s41596-020-0312-x

63. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem*. (2008) 29:1859–65. doi: 10.1002/jcc.20945

64. Ong EES, Liow J-L. The temperature-dependent structure, hydrogen bonding and other related dynamic properties of the standard TIP3P and CHARMM-modified TIP3P water models. *Fluid Phase Equilib*. (2019) 481:55–65. doi: 10.1016/j.fluid.2018.10.016

65. Petersen HG. Accuracy and efficiency of the particle mesh Ewald method. *J Chem Phys*. (1995) 103:3668 3679. doi: 10.1063/1.470043

66. Haug EJ, Arora JS, Matsui K. A steepest-descent method for optimization of mechanical systems. *J Optim Theory Appl*. (1976) 19:401 424. doi: 10.1007/BF00941484

67. Nesabi A, Kalayan J, Al-Rawashdeh S, Ghattas MA, Bryce RA. Molecular dynamics simulations as a guide for modulating small molecule aggregation. *J Comput Aided Mol Des*. (2024) 38:11. doi: 10.1007/s10822-024-00557-1

68. Kumari R, Kumar R, Lynn A. g_mmpbsa—A GROMACS tool for highThroughput MMPBSA calculations. *J Chem Inf Model*. (2014) 54:1951–62. doi: 10.1021/ci500020m

69. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci United States America*. (2001) 98:10037–41. doi: 10.1073/pnas.181342398

70. Sharma AD, Samuel J, Singh H, Cuspoca F, Cavallo L, Chawla M. Immunoinformatics-driven design of a multi-epitope vaccine targeting neonatal rotavirus with focus on outer capsid proteins VP4 and VP7 and non structural proteins NSP2 and NSP5. *Sci Rep*. (2025) 15. doi: 10.1038/s41598-025-95256-8

71. Kaushik V, G SK, Gupta LR, Kalra U, Shaikh AR, Cavallo L, et al. Immunoinformatics aided design and *in-vivo* validation of a cross-reactive peptide based multi-epitope vaccine targeting multiple serotypes of dengue virus. *Front Immunol*. (2022) 13:3389/fimmu.2022.865180. doi: 10.3389/fimmu.2022.865180

72. Kaushik V, Jain P, Akhtar N, Joshi A, Gupta LR, Grewal RK, et al. ImmunoinformaticsAided design and *in vivo* validation of a peptideBased multi-epitope vaccine targeting canine circovirus. *ACS Pharmacol Transl Sci*. (2022) 5:679–91. doi: 10.1021/acsptsci.2c00130

73. GenScript. Make Research Easy - The leader in molecular cloning and gene synthesis, peptide synthesis, protein and antibody engineering. Available online at: https://www.genscript.com/.

74. Kamens J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res*. (2015) 43:D1152–7. doi: 10.1093/nar/gku893

75. 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

76. Sanner MF. Python: a programming language for software integration and development. *J Mol Graph Model*. (1999) 17:57–61.

77. Tuimala J, Kallio A. R, programming language. In: *Encyclopedia of Systems Biology*. Springer New York, New York, NY (2013). p. 1809–11. doi: 10.1007/978-1-4419-9863-7_619

78. Chawla M, Cuspoca AF, Akthar N, Magdaleno JSL, Rattanabunyong S, Suwattanasophon C, et al. Immunoinformatics-aided rational design of a multi-epitope vaccine targeting feline infectious peritonitis virus. *Front Vet Sci*. (2023) 10:1280273. doi: 10.3389/fvets.2023.1280273

79. Elfadadny A, Ragab RF, AlHarbi M, Badshah F, Ibáñez-Arancibia E, Farag A, et al. Antimicrobial resistance of Pseudomonas aeruginosa: navigating clinical impacts, current resistance trends, and innovations in breaking therapies. *Front Microbiol*. (2024) 15:1374466. doi: 10.3389/fmicb.2024.1374466

80. Assefa M, Amare A. Biofilm-associated multi-drug resistance in hospital-acquired infections: A review. *Infect Drug Resist*. (2022) 15:5061–8. doi: 10.2147/IDR.S379502

81. Sharma D, Misba L, Khan AU. Antibiotics versus biofilm: an emerging battleground in microbial communities. *Antimicrob Resist Infect Control*. (2019) 8:76. doi: 10.1186/s13756-019-0533-3

82. Lopes JA, Rghei AD, Thompson B, Susta L, Khursigara CM, Wootton SK. Overcoming barriers to preventing and treating P. aeruginosa infections using AAV vectored immunoprophylaxis. *Biomedicines*. (2022) 10:3162. doi: 10.3390/biomedicines10123162

83. Sharma R, Rajput VS, Jamal S, Grover A, Grover S. An immunoinformatics approach to design a multi-epitope vaccine against Mycobacterium tuberculosis exploiting secreted exosome proteins. *Sci Rep*. (2021) 11:13836. doi: 10.1038/s41598-021-93266-w

84. Alizadeh M, Amini-Khoei H, Tahmasebian S, Ghatrehsamani M, Ghatreh Samani K, Edalatpanah Y, et al. Designing a novel multi epitope vaccine against Ebola virus using reverse vaccinology approach. *Sci Rep*. (2022) 12:7757. doi: 10.1038/s41598-022-11851-z

85. Sarkar B, Ullah M, 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

86. Elhag M, Ruaa Mohamed Alaagib, Ahmed NM, Abubaker M, Haroun EM, Sahar, et al. Design of Epitope-Based Peptide Vaccine against Pseudomonas aeruginosa Fructose Bisphosphate Aldolase Protein Using Immunoinformatics. *J Immunol Res*. (2020) 2020:1–11. doi: 10.1155/2020/9475058

87. Zhu F, Qin R, Ma S, Zhou Z, Tan C, Yang H, et al. Designing a multi-epitope vaccine against Pseudomonas aeruginosa via integrating reverse vaccinology with immunoinformatics approaches. *Sci Rep*. (2025) 15. doi: 10.1038/s41598-025-90226-6

88. 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. doi: 10.1038/s41598-019-41496-4

89. Roy SK, Biswas MS, Raman MF, Hasan R, Zahidur Rahmann, Moyen M. A computational approach to developing a multi-epitope vaccine for combating Pseudomonas aeruginosa–induced pneumonia and sepsis. *Briefings Bioinf*. (2024) 25. doi: 10.1093/bib/bbae401

90. Nagpal G, Usmani SS, Raghava GPS. A web resource for designing subunit vaccine against major pathogenic species of bacteria. *Front Immunol*. (2018) 9:3389/fimmu.2018.02280. doi: 10.3389/fimmu.2018.02280

91. 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) 20:123. doi: 10.1186/s12859-019-2713-9

92. Damron FH, Napper J, Teter MA, Yu HD. Lipotoxin F of Pseudomonas aeruginosa is an AlgU-dependent and alginate-independent outer membrane protein involved in resistance to oxidative stress and adhesion to A549 human lung epithelia. *Microbiol (N Y)*. (2009) 155:1028–38. doi: 10.1099/mic.0.025833-0

93. Belo VA, Pereira JA, Souza SFD, Tana FL, Pereira BP, Lopes D de O, et al. The role of IL-10 in immune responses against Pseudomonas aeruginosa during acute lung infection. *Cell Tissue Res*. (2020) 383:1123–33. doi: 10.1007/s00441-020-03308-4

94. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*. (2010) 330:341–6. doi: 10.1126/science.1187409

95. Cheng X, Chen Z, Gao C, Zhang Y, Yang L, Wan J, et al. Structural and biological insights into outer membrane protein lipotoxin F of Pseudomonas aeruginosa: Implications for vaccine application. *Int J Biol Macromol*. (2023) 253:127634. doi: 10.1016/j.ijbiomac.2023.127634

96. Sharma S, Yadav PD, Cherian S. Comprehensive immunoinformatics and bioinformatics strategies for designing a multi-epitope based vaccine targeting structural proteins of Nipah virus. *Front Immunol*. (2025) 16:1535322. doi: 10.3389/fimmu.2025.1535322

97. Qiu J, Wei Y, Shu J, Zheng W, Zhang Y, Xie J, et al. Integrated in-silico design and *in vivo* validation of multi-epitope vaccines for norovirus. *Virol J*. (2025) 22:166. doi: 10.1186/s12985-025-02796-6

frontiers | Frontiers in Immunology

TYPE Methods
PUBLISHED 05 December 2025
DOI 10.3389/fimmu.2025.1628583

OPEN ACCESS

EDITED BY
Gurudeeban Selvaraj,
Aarupadai Veedu Medical College and Hospital, India

REVIEWED BY
Mi Liu,
Soochow University, China
Matthew Halma,
Frontline COVID-19 Critical Care Alliance, United States

*CORRESPONDENCE
Valentina Di Salvatore
valentinadisalvatore@unict.it

RECEIVED 14 May 2025
REVISED 11 November 2025
ACCEPTED 18 November 2025
PUBLISHED 05 December 2025

CITATION
Di Salvatore V, Cernuto F, Russo G and Pappalardo F (2025) A computational framework for optimizing mRNA vaccine delivery via AI-guided nanoparticle design and *in silico* gene expression profiling. *Front. Immunol.* 16:1628583. doi: 10.3389/fimmu.2025.1628583

A computational framework for optimizing mRNA vaccine delivery via AI-guided nanoparticle design and *in silico* gene expression profiling

Valentina Di Salvatore[1]*, Federica Cernuto[2], Giulia Russo[1] and Francesco Pappalardo[1]

[1]Department of Health and Drug Sciences, University of Catania, Catania, Italy, [2]Department of Mathematics and Computer Science, University of Catania, Catania, Italy

Recent concerns about off-target immune activation following non-targeted mRNA vaccine delivery have prompted the need for rational design strategies that optimize nanoparticle formulations. Building upon our previous in silico work using the Universal Immune System Simulator to characterize immune responses to mRNA vaccines, we present a computational framework that integrates synthetic transcriptomics with artificial intelligence-driven optimization to guide the development of safer and more targeted lipid nanoparticles. We generated biologically informed, synthetic RNA-seq datasets to emulate gene expression profiles in immune-related tissues post-vaccination. Differential gene expression analysis identified compartment-specific transcriptional responses, which were then used to construct a risk index based on predicted immune activation and the number of upregulated immune markers. Parallelly, we trained a Random Forest regression model on simulated lipid nanoparticles formulations to predict immune activation values and embedded this model into a genetic algorithm to identify optimal lipid nanoparticles design parameters (size, charge, polyethylene glycol content, and targeting). The proposed framework enables early-stage, fully in silico screening of mRNA vaccine delivery strategies. Our results highlight the potential of combining mechanistic immune modeling, synthetic transcriptomic validation, and Artificial Intelligence-based design to accelerate the development of safer and more effective mRNA-based therapies. By enabling rapid, data-driven optimization of delivery systems prior to experimental validation, this approach can significantly shorten vaccine development timelines, reduce costs, and support the creation of more personalized and adaptable immunization strategies. In the long term, this paradigm shift toward computationally guided vaccine development could redefine the future of immunization, paving the way for next-generation vaccines that are safer, more targeted, and rapidly adaptable to emerging infectious threats and individual patient needs.

KEYWORDS

mRNA vaccines, lipid nanoparticles, synthetic transcriptomics, AI-driven optimization, immune modeling, synthetic omics data, optimization algorithms, vaccine delivery

1 Introduction

Messenger RNA (mRNA) vaccines have revolutionized the field of immunization, offering rapid development timelines, high efficacy, and adaptability to various pathogens. The success of mRNA-based vaccines against COVID-19 has underscored their potential in combating infectious diseases and beyond. Central to the efficacy of these vaccines is the delivery system, with lipid nanoparticles (LNPs) emerging as the leading non-viral vectors for mRNA delivery. LNPs protect mRNA from degradation, facilitate cellular uptake, and promote endosomal escape, ensuring efficient translation of the antigenic protein (1).

Despite these advantages, significant challenges remain in optimizing LNP formulations to achieve an optimal balance between efficacy and safety. Variations in key physicochemical properties, such as particle size, surface charge, PEGylation density and lipid composition, can substantially affect biodistribution, cellular uptake, endosomal escape, and ultimately, the magnitude and specificity of the immune response. For example, LNPs with highly cationic surfaces may enhance cellular internalization but also activate Toll-like receptors (TLRs) or inflammasome pathways, potentially inducing undesired innate immune responses, systemic inflammation, or even reactogenicity. Conversely, overly neutral or PEG-shielded formulations may escape immune surveillance altogether, limiting antigen presentation and immunogenicity (2).

Moreover, the biodistribution of LNPs is highly context-dependent, influenced by physiological barriers, tissue tropism, and inter-patient variability, making empirical optimization challenging (3). Conventional methodologies for LNP design rely on iterative, trial-and-error testing of individual components, a process that is both time-consuming and resource-intensive, often requiring extensive *in vitro* and *in vivo* validation to assess delivery efficiency and immune activation profiles.

Traditional Design of Experiments (DOE) approaches have been widely employed to systematically explore the impact of formulation variables on nanoparticle characteristics and performance. By using structured experimental matrices, DOE enables the simultaneous evaluation of multiple parameters and their interactions, significantly improving the efficiency and robustness of formulation optimization compared to traditional one-variable-at-a-time methods (4, 5). For example, factorial and response surface methodologies have proven effective in optimizing lipid nanoparticle properties such as size, charge, and encapsulation efficiency for mRNA delivery (4). However, while DOE provides a powerful framework for structured experimentation, it still requires substantial experimental resources and may be limited in capturing the full complexity of biological responses. This highlights the need for complementary in silico approaches that can simulate biological systems, reduce experimental burden, and guide rational design more efficiently.

In this context, to overcome these limitations and fully capture the complexity of nanoparticle-biology interactions, computational modeling and artificial intelligence (AI)-driven optimization offer a powerful alternative for systematically exploring the vast design space of LNPs. By simulating biological outcomes and predicting key response metrics such as immunogenicity or off-target activation, these tools enable a more rational and cost-effective approach to LNP development, potentially accelerating the pipeline from formulation design to preclinical validation.

Recent advancements in computational biology and AI offer promising avenues to streamline LNP design. Machine learning models can predict the physicochemical properties of LNPs and their biological interactions, enabling the rational design of nanoparticles with desired characteristics (6). Additionally, synthetic transcriptomics allows for the simulation of gene expression profiles post-vaccination, providing insights into potential immune responses without the need for extensive *in vivo* studies.

Building upon our previous work utilizing the Universal Immune System Simulator (UISS) to model immune responses to mRNA vaccines (7), we propose an integrated in silico framework that combines synthetic transcriptomics with AI-driven optimization strategies. While recent advances in computational biology have introduced simulation-based approaches and machine learning for drug delivery design, comprehensive platforms that integrate immune modeling, synthetic omics data, and optimization algorithms for vaccine delivery remain scarce. Our framework addresses this gap by offering a modular, reproducible pipeline capable of generating biologically informed synthetic RNA-seq datasets, performing differential expression analysis, computing immune activation risk scores, and identifying optimal lipid nanoparticle (LNP) formulations via machine learning and evolutionary computation.

The pipeline was developed entirely in R and Python, leveraging robust and widely used packages. This integrated approach enables both hypothesis generation and rational design in the early stages of mRNA vaccine development, with the goal of minimizing off-target immune activation and maximizing targeted delivery efficiency. By simulating transcriptional responses and incorporating interpretable machine learning models into an optimization framework, our methodology aims to accelerate the design of safer and more effective mRNA-based therapeutics.

2 Methods

The workflow, shown in Figure 1, includes the following key steps:

a) Synthetic RNA-seq Data Generation

A synthetic RNA-seq dataset was constructed to mimic gene expression profiles post-vaccination. It included immune-related marker genes for key compartments (e.g., CD19 for B cells, CD3D for T cells, IGHG1 for plasma cells), with differential expression patterns reflecting simulated immune activation.

b) Transcriptomic Analysis and Immune Risk Indexing

The synthetic RNA-seq dataset was analyzed for differential gene expression. The number of significantly upregulated immune marker genes per compartment was used to compute a risk index by

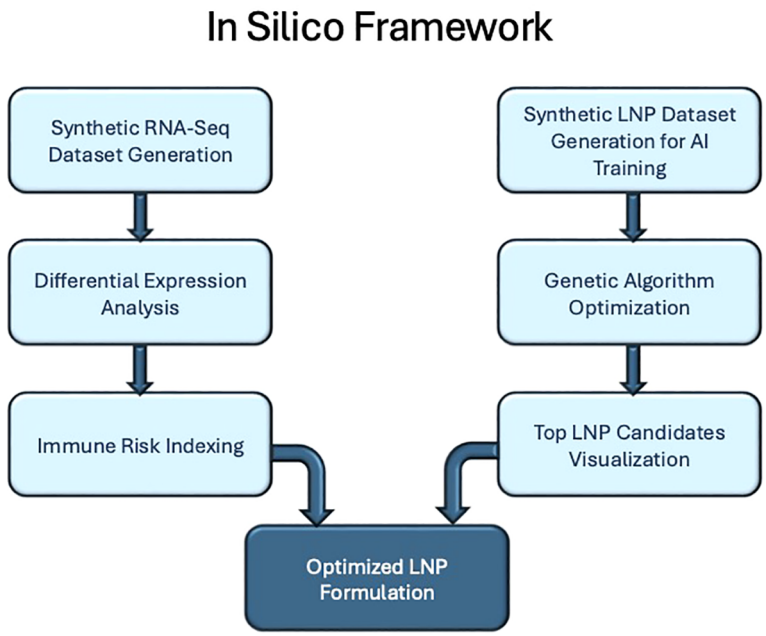


FIGURE 1
Graphical representation of the in silico framework for optimizing mRNA vaccine delivery.

multiplying with corresponding Delta_AUC values. This yielded a semi-quantitative estimate of off-target immune activation risk.

c) Synthetic LNP Dataset for AI Training

A synthetic dataset of LNP formulations was generated by varying four key physicochemical parameters: particle size (50–150 nm), surface charge (−10 to +10 mV), PEGylation percentage (0.1–0.5 mol%), and targeting ligand presence (binary). Delta_AUC values were assigned to each formulation using a custom nonlinear scoring function designed to reflect optimal biodistribution and immunogenicity.

d) Machine Learning Model Development

A Random Forest regression model was trained to predict Delta_AUC values based on LNP parameters. The model was validated internally using performance metrics such as RMSE and R^2.

e) Genetic Algorithm Optimization

The trained model was embedded within a genetic algorithm to identify LNP configurations predicted to maximize immune delivery efficiency while minimizing off-target activation. The top 10 candidates were selected for further analysis.

f) Data Visualization and Interpretation

A heatmap and ranked plots were used to summarize the properties of optimized LNP formulations and their predicted immune activation scores. These visualizations highlighted common design features among the best-performing candidates.

This approach enables rational design of mRNA vaccine formulations with improved targeting and reduced off-target immune activation, and it will be discussed in detail in next paragraphs.

All simulations, data generation, and analyses were performed using a custom R and Python-based workflow developed for this study. Core statistical procedures and expression modeling were conducted in R (v4.4.1) within RStudio 2024.04.2 + 764, leveraging established packages including DESeq2 (version '1.48.1') for differential gene expression analysis (8) *randomForest* (version '4.7.1.2') for predictive modeling (9), *GA* (version '3.2.4') for genetic algorithm optimization (10), and *ggplot2* (version '4.0.0') and *pheatmap* ('1.0.13') (https://github.com/raivokolde/pheatmap) for initial data visualization (11). To enhance figure aesthetics and consistency, key visualizations (e.g., ΔAUC comparisons, immune risk index, LNP ranking) were refined using Python (v3.13.2) in a virtual environment with the matplotlib (12) and seaborn (13) libraries.

All analyses were performed on an iMac with Apple M3 chip (8-core CPU, 10-core GPU) equipped with 24 GB unified memory, running macOS Sequoia 15.6.1.

2.1 Synthetic RNA-seq generation

To model transcriptional responses to mRNA vaccination, we generated a synthetic RNA-seq dataset based on biologically informed assumptions and guided by immunological response profiles simulated using the UISS platform in our previous work. The dataset comprised 300 genes measured across 10 samples (5 Control and 5 Post-Vaccination). A subset of genes was designed to simulate vaccine-induced immune activation: 30 genes were upregulated and 30 downregulated in the post-vaccination group relative to controls.

Additionally, well-established immune marker genes were included to represent specific compartments, B cells (*CD19*, *MS4A1*) (14), T cells (*CD3D*, *CD8A*, *CD4*) (15), plasma cells

(*IGHG1*, *IGHM*, *PRDM1*) (16), and others, artificially upregulated to reflect canonical immune activation following antigen exposure.

Gene expression values were sampled from normal distributions, with mean shifts used to simulate differential regulation. To preserve biological plausibility, negative values, resulting from the statistical properties of normal distributions, were truncated to zero. This step ensures that all simulated expression values remain non-negative, reflecting the reality that gene expression levels, being measures of transcript abundance, cannot be less than zero. This approach enables the simulation of genes with no detectable expression while avoiding artifacts that could compromise downstream analysis.

This synthetic dataset serves a dual purpose. On one hand, it allows controlled benchmarking of the transcriptomic analysis pipeline, particularly in assessing its ability to recover known patterns of immune activation. On the other hand, it acts as a bridge to validate predictions generated by the Universal Immune System Simulator (UISS), a mechanistic, agent-based platform capable of modeling immune responses at multiple scales, from molecular signaling to cellular interactions and tissue-level dynamics (17, 18).

Specifically, UISS has been used to simulate host responses to mRNA vaccines, including the biodistribution of lipid nanoparticles (LNPs), antigen presentation, and subsequent activation of adaptive immunity (7). Based on its simulations, UISS produces immunological outputs, such as the expansion of specific immune cell subsets or the secretion of key cytokines, that can be mapped to gene expression patterns. While UISS does not generate RNA-seq data directly, these outcomes can be qualitatively and semi-quantitatively translated into gene expression profiles, enabling the construction of biologically plausible synthetic datasets.

By constructing a synthetic RNA-seq dataset that reflects these expected transcriptional signatures, we can assess whether downstream analysis methods (e.g., differential expression, immune risk indexing) can faithfully recapitulate the immune activation patterns originally predicted by UISS. This integration provides a robust framework for evaluating the predictive alignment between mechanistic modeling and transcriptomic data analytics in the context of rational vaccine design.

2.2 Transcriptomic analysis and immune risk indexing

Differential gene expression analysis was performed using the *DESeq2* package in R, employing negative binomial distribution modeling and Wald tests to identify significantly differentially expressed genes between the post-vaccination and control groups within the synthetic RNA-seq dataset (8). Gene-wise fold changes and adjusted p-values (Benjamini-Hochberg correction) were computed to isolate significantly upregulated immune-related genes (FDR < 0.05).

To infer the immunological profiles of each condition, marker genes characteristics of major immune compartments were selected based on established immunological literature. Specifically, we considered CD19 and MS4A1 for B cells (19), CD3D, CD8A, and CD4 for T cells (20) and IGHG1, IGHM, and PRDM1 (BLIMP-1) for plasma cells (21).

Based on prior immune simulation results, we introduced a compartment-specific risk index designed to quantitatively evaluate the potential for unintended immune activation (off-target effects). The immune risk index for each compartment was calculated by multiplying the simulated Delta_AUC (area under the curve representing cumulative immune activation over time, as established in previous immunological modeling studies (17)) by the count of significantly upregulated marker genes identified in the differential expression analysis for that immune compartment. This integrated approach combines functional simulation data with empirical transcriptomic profiles, providing a robust, interpretable, and semi-quantitative metric for assessing immune activation risks associated with vaccination or other therapeutic interventions.

2.3 Synthetic LNP dataset for AI training

To support the development and evaluation of an AI-driven optimization pipeline for lipid nanoparticle (LNP) formulations, we generated a synthetic dataset consisting of diverse LNPs characterized by defined physicochemical parameters and corresponding immune activation scores (Delta_AUC). Each LNP formulation was parameterized based on four key physicochemical attributes known to significantly impact biodistribution, cellular uptake, and immunogenicity: particle size (ranging from 50 to 150 nm), which influences circulation time and tissue penetration (22); surface charge (−10 to +10 mV), affecting cellular interaction (23) and colloidal stability (24); PEGylation percentage (0.1 to 0.5 mol %), referring to the covalent attachment of polyethylene glycol (PEG) chains to the nanoparticle surface, a modification that confers a steric barrier against opsonization, reduces recognition and clearance by the mononuclear phagocyte system, prolongs systemic circulation time, and imparts a "stealth" property that enhances *in vivo* stability (24); and the presence or absence of active targeting ligands (binary encoded, where 0 represents untargeted and 1 represents targeted nanoparticles), enabling selective binding to specific cellular receptors (24, 25). A total of 100 distinct LNP formulations were systematically sampled across this four-dimensional parameter space, ensuring uniform representation and adequate coverage for robust AI model training. A summary of the main effects of these physicochemical parameters on biodistribution, cellular uptake, and immunogenicity, are summarized in Table 1:

Each formulation was assigned a Delta_AUC value, calculated using a biologically informed, non-linear scoring function explicitly designed to simulate realistic biodistribution and immunological response patterns observed experimentally:

$$\text{Delta\_AUC} = -0.01(\text{Size} - 90)^2 - 0.02(\text{Charge})^2 + 0.5\text{PEG} + 1.5\,\text{Targeting} + \varepsilon$$

TABLE 1 Physicochemical attributes of LNPs and their predicted biological effects.

LNP attribute	Main biological effects
Particle size (50-150 nm)	Affects biodistribution and tissue penetration: smaller LNPs circulate longer and diffuse more effectively, whereas larger LNPs tend to accumulate in the liver and spleen.
Surface Charge (-10 to +10 mV)	Modulates cellular uptake and stability; neutral charge improves circulation; positive charge increases uptake but may raise immunogenicity.
PEGylation (0.1-0.5 mol%)	Reduces opsonization and clearance, prolonging circulation, and provides stealth properties; excessive PEG reduces cellular uptake.
Targeting Ligands	Determines targeting specificity: without ligands, LNPs accumulate passively in the liver; with ligands, delivery is more specific, efficacy improves, and toxicity is reduced.

Each attribute, such as particle size, surface charge, PEGylation, and targeting ligands, affects biodistribution, cellular uptake, circulation time, and delivery specificity.

In this formulation, ε represents Gaussian-distributed noise (mean = 0, standard deviation = 0.1), reflecting biological variability and measurement uncertainty typically encountered in experimental settings (26).

The scoring function for Delta_AUC was constructed to reflect biologically plausible relationships between key nanoparticle properties and delivery performance, based on known experimental trends. Specifically, the quadratic penalty terms for size and surface charge were introduced to model the existence of optimal values: nanoparticle diameters around 90 nm and near-neutral charges are experimentally associated with enhanced circulation times and improved biodistribution profiles. Therefore, the terms $-0.01(\text{Size} - 90)^2$ and $-0.02(\text{Charge})^2$ penalize deviations from these optimal values, with the choice of coefficients scaling the relative importance of size and charge in the delivery performance.

Conversely, PEGylation and active targeting were modeled as linear contributors to performance. The positive coefficients (+0.5 for PEGylation and +1.5 for targeting) reflect the experimental evidence that moderate PEGylation improves nanoparticle stealth properties, and the presence of active targeting ligands substantially enhances cellular uptake by promoting receptor-mediated endocytosis.

Finally, Gaussian-distributed noise (ε, mean = 0, standard deviation = 0.1) was added to each Delta_AUC value to simulate biological variability and measurement uncertainty typically observed *in vivo* and *in vitro* assays. This biologically informed functional form allowed us to create a synthetic dataset, through an in-house R script, where optimal nanoparticle configurations (around 90 nm in size, with near-neutral surface charge, moderate PEGylation, and active targeting) systematically achieve higher Delta_AUC values, while suboptimal configurations are penalized. This design ensures that machine learning models trained on the dataset are exposed to realistic, non-linear, and multi-parametric optimization challenges, mimicking the complexity of real-world nanoparticle formulation tasks (27, 28).

This synthetic dataset was subsequently used to train and evaluate a supervised machine learning model, as described in the following section.

2.4 Machine learning model development

A supervised machine learning approach was employed to predict immune activation potential (ΔAUC) of lipid nanoparticle (LNP) formulations based on key physicochemical descriptors. A Random Forest regression model (29) was implemented using the randomForest package in R. Input features included particle size (nm), surface charge (mV), PEGylation percentage (mol%), and presence of targeting ligands (binary encoding).

The synthetic dataset described above, comprising 100 simulated LNP formulations generated by systematically varying key physicochemical parameters across biologically relevant ranges, was randomly partitioned into training (80%) and validation (20%) subsets. Model performance was assessed using root mean square error (RMSE) and the coefficient of determination (R^2) on the validation set, providing quantitative estimates of predictive accuracy and generalizability (30). RMSE measures the average magnitude of the prediction errors, providing an estimate of how close the predicted values are to the actual ones: lower values indicate better accuracy. R^2 quantifies the proportion of variance in the observed data that is explained by the model, with values closer to 1 indicating higher predictive power and generalizability.

Feature importance was assessed using the mean decrease in node impurity, a standard metric in Random Forest models that quantifies how much each variable contributes to improving decision tree splits. This analysis revealed that surface charge and PEGylation were the most influential predictors of ΔAUC, in line with their well-established roles in modulating nanoparticle biodistribution and immunogenicity. The trained model was subsequently used to predict ΔAUC values for all LNP candidates in the dataset. The top 10 formulations were selected based on their predicted scores and visualized accordingly, forming the basis for downstream optimization via genetic algorithms.

Full model and optimization settings, cross-validation protocol, and sensitivity analysis are reported in Supplementary Table S2 and Supplementary Methods S1.

2.5 Genetic algorithm optimization

Building on the predictive model trained on the synthetic LNP dataset, we used the optimized Random Forest as a surrogate fitness function within a genetic algorithm (GA) to search for new LNP formulations predicted to yield high Delta_AUC values, i.e., strong immune activation profiles. The GA was implemented using the GA library in R, which simulates an evolutionary process to solve optimization problems. We began with an initial population of 50 LNP formulations, randomly generated within biologically plausible parameter ranges (for size, charge, PEGylation, and targeting). Each formulation in the population was evaluated using the trained

Random Forest model, which predicted its Delta_AUC score: this prediction served as the fitness value for the GA. The selection of individuals for reproduction was performed using a tournament strategy, where multiple candidates compete and the best is chosen for mating. To simulate genetic diversity and exploration of the solution space, we applied crossover (with a probability of 0.8) to exchange parameter values between formulations, and mutation (with a probability of 0.2) to introduce small random changes. This process was repeated over 100 generations. As the algorithm progressed, it increasingly favored formulations with higher predicted Delta_AUC, gradually converging towards optimal solutions. At the end of the run, we selected the top 10 formulations, those with the highest predicted Delta_AUC scores, for further analysis.

2.6 Statistical confidence and clustering analyses

To quantify the robustness of the model predictions and the associated uncertainty, we performed statistical confidence and clustering analyses on both the immune risk index and the ΔAUC predictions.

For the immune risk index (Figure 2), 95% confidence intervals were estimated using a nonparametric bootstrap procedure (B = 1000 resamplings) applied to compartment-specific immune markers, weighted by their respective ΔAUC coefficients.

For the ΔAUC predictions (Figure 3), a bootstrap approach was applied to the random forest model, which was re-trained on 500 bootstrap samples of the synthetic LNP dataset to estimate prediction variability.

In addition, hierarchical clustering was incorporated into the heatmaps (Figure 4) to highlight parameter co-variation, and a correlation heatmap (Spearman's ρ) was generated to visualize relationships among LNP physicochemical parameters and ΔAUC values.

2.7 Data visualization and software

All data preprocessing, statistical analyses, and initial visualizations were performed using R (v4.4.1) within RStudio 2024.04.2 + 764. Differential expression analysis was carried out with DESeq2 (v1.48.1), while predictive modeling and optimization were implemented using randomForest (v4.7.1.2) and GA (v3.2.4), respectively. Exploratory plots were generated with ggplot2 (v4.0.0) and pheatmap (v1.0.13).

To refine figure design and ensure visual consistency, selected key plots—such as ΔAUC comparisons, immune risk index distributions, and LNP ranking—were reproduced using Python (v3.13.2) in a dedicated virtual environment with matplotlib (v3.9.2) and seaborn (v0.13.2).

All analyses were executed on an iMac with Apple M3 chip (8-core CPU, 10-core GPU) equipped with 24 GB unified memory, running macOS Sequoia 15.6.1. This hybrid R/Python workflow ensured both graphical uniformity and full reproducibility across the study.

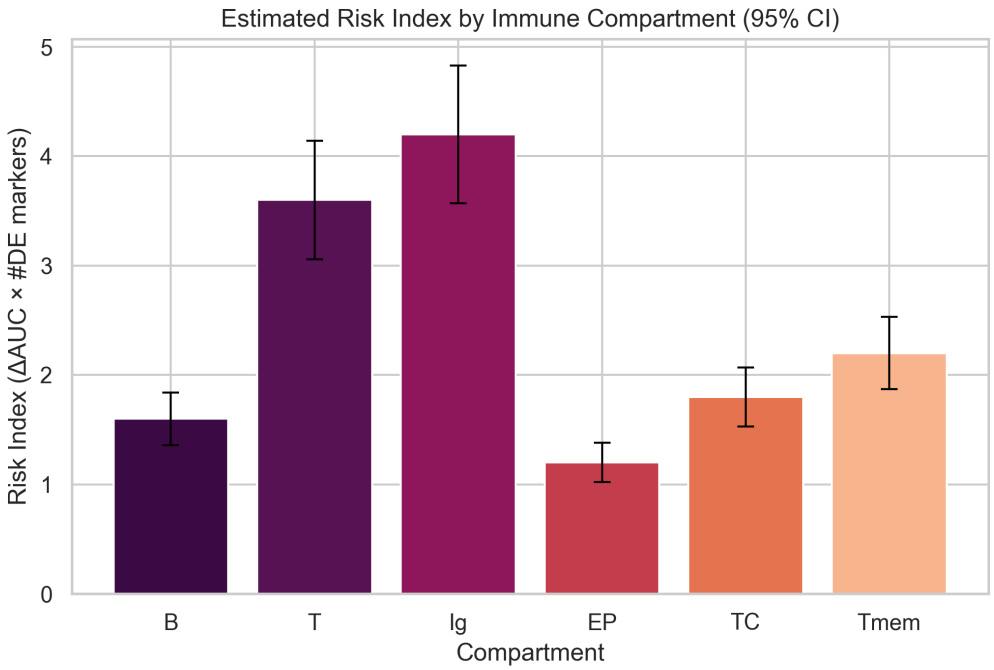


FIGURE 2
Estimated immune risk index by compartment, computed as the product of Delta_AUC and the number of upregulated immune marker genes. This index reflects potential off-target immune activation.

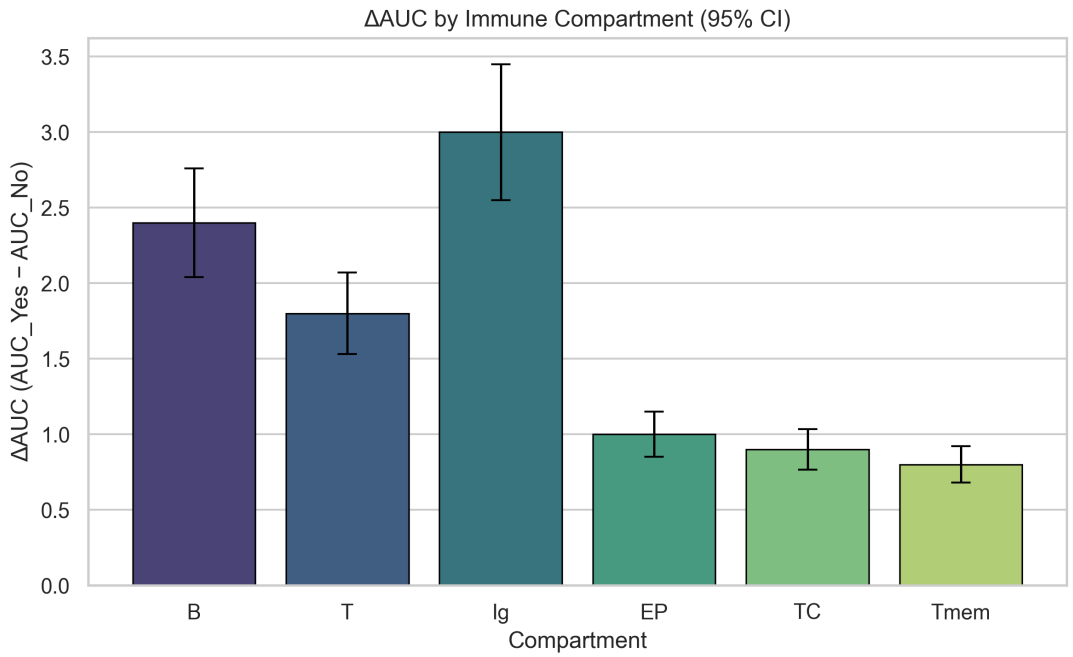


FIGURE 3
Delta_AUC values for immune compartments, calculated as the difference in activation between targeted and non-targeted formulations. Higher values indicate stronger compartment-specific immune responses to targeted delivery.

3 Results

3.1 Synthetic RNA-seq differential expression analysis

Differential gene expression analysis of the synthetic RNA-seq dataset accurately identified the simulated transcriptional changes. Among the 300 analyzed genes, all 30 genes designed to be upregulated, and the 30 genes designated as downregulated post-vaccination were correctly identified as significantly differentially expressed (FDR < 0.05), demonstrating the reliability and validity of the synthetic data generation methodology. Additionally, key immune marker genes representing distinct immune compartments, such as B cells (CD19, MS4A1), T cells (CD3D, CD8A, CD4), and plasma cells (IGHG1, IGHM, PRDM1), were significantly upregulated, consistent with expected immune activation patterns.

3.2 Immune risk indexing

To assess potential off-target immune activation, we computed a compartment-specific immune risk index by multiplying the predicted ΔAUC values by the number of differentially expressed (DE) immune marker genes within each compartment, as shown in Figure 2:

The Ig compartment, representing antibody-producing plasma cells, displayed the highest risk index (~4.2), suggesting a strong activation of humoral responses, consistent with mRNA vaccine effects (31). The T cell compartment followed (~3.6), indicating robust T cell engagement. Memory T cells (Tmem) and cytotoxic T cells (TC) showed moderate risk levels (~2.2 and ~1.8, respectively), while B cells had a slightly lower activation (~1.6). Notably, the EP compartment, likely representing epithelial or non-immune cells, had the lowest index (~1.2), suggesting minimal off-target transcriptional activation. These results support the capacity of the simulated nanoparticle formulation to preferentially activate relevant immune pathways while sparing non-target tissues, aligning with the immune response patterns previously predicted by UISS.

The calculated immune risk index effectively quantified compartment-specific immune activation, clearly distinguishing between post-vaccination and control conditions. Specifically, the highest immune risk index values were observed in the T cell compartment, driven by strong upregulation of CD3D, CD8A, and CD4 genes, in alignment with simulated Delta_AUC scores derived from the UISS model. B cell and plasma cell compartments exhibited moderate immune risk scores, correlating with fewer significantly upregulated marker genes. Overall, the immune risk indexing method demonstrated strong correlation with simulated immune activation, offering a robust and interpretable approach for evaluating potential off-target immune responses.

3.3 Simulated immune compartment activation

Based on prior UISS simulations, immune compartments showed distinct activation patterns when comparing targeted and non-targeted mRNA vaccine delivery. Delta_AUC values were calculated to quantify the difference in immune activation

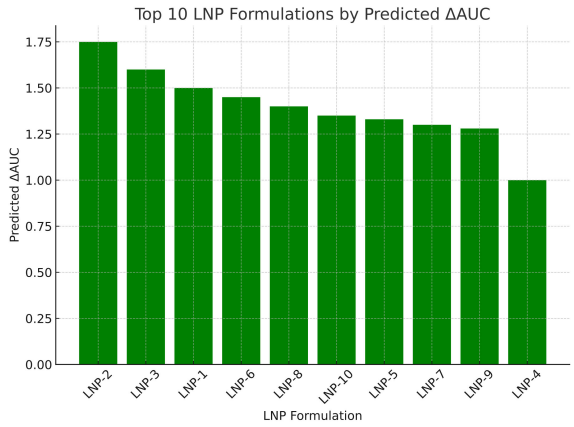


FIGURE 4
Top 10 LNP formulations ranked by predicted ΔAUC. Barplot showing the predicted immune activation scores (ΔAUC) for the top LNP candidates identified by the genetic algorithm. LNP-2 achieved the highest predicted score (ΔAUC = 1.73), with others following in descending order.

between conditions. Compartments such as B cells and plasma cells (Ig) showed the highest differential activation, indicating preferential targeting and stronger immune engagement when delivery was optimized.

The difference in immune activation between targeted and non-targeted formulations (ΔAUC) was computed for each immune compartment.

As illustrated in Figure 3, the Ig compartment exhibited the highest increase in ΔAUC, followed by B and T cells, indicating a stronger activation under targeted delivery.

In contrast, epithelial (EP), cytotoxic (TC), and memory T (Tmem) compartments showed smaller ΔAUC values, suggesting that their activation is less affected by the delivery modality within the current simulation setup.

3.4 AI-based prediction and ranking of LNP formulations

Using the synthetic dataset previously described, which uniformly sampled a four-dimensional physicochemical parameter space, we trained a Random Forest regression model to predict Delta_AUC values based on LNP characteristics. The model achieved strong predictive performance, with R^2 values exceeding 0.9 and low RMSE on the validation set, confirming its ability to capture non-linear relationships between input features and immune activation potential.

The model was then embedded as a surrogate fitness function within a genetic algorithm to identify LNP formulations predicted to maximize Delta_AUC. After 100 generations, the GA consistently converged toward optimal configurations, that is, nanoparticles around 90 nm in size, with near-neutral surface charge, moderate PEGylation, and active targeting, closely matching profiles known to enhance biodistribution and immunogenicity. Following model training and validation, ΔAUC values were predicted for the entire synthetic LNP dataset. After convergence, the genetic algorithm identified a set of top 10 LNP formulations that consistently exhibited superior predicted performance as shown in Table 1:

All selected candidates included active targeting ligands and exhibited particle sizes ranging from 88.8 to 93.9 nm, with a central tendency around 90–92 nm, aligning with theoretical optima for biodistribution. This outcome reflects the influence of the scoring function used during model training, which included a positive weighting for the presence of targeting ligands, thereby favoring formulations predicted to enhance receptor-mediated uptake and compartment-specific immune activation. Surface charges were consistently near-neutral, varying between −1.0 and −4.4 mV, and PEGylation percentages ranged from 0.26 to 0.34 mol%, centering around the biologically favorable 0.3 mol%. This near-neutral charge is known to minimize non-specific interactions with serum proteins and immune cells, thereby improving circulation time and reducing innate immune activation (32). Similarly, an optimal PEGylation density has been shown to balance nanoparticle stealth and cellular uptake, preventing rapid clearance while maintaining delivery efficiency.

The predicted ΔAUC values, calculated using the biologically informed non-linear scoring function described in the Methods section, ranged from 0.99 to 1.73. The highest score (1.73) was achieved by the top-performing formulation (Rank 2), while the lowest among the top 10 (Rank 4) was 0.99. Although the ΔAUC range was narrower than initially anticipated, the results highlight the genetic algorithm's ability to finely discriminate between LNP designs with subtle yet functionally meaningful differences.

Notably, all top-ranked formulations exhibited overlapping physicochemical features: particle sizes around 90–92 nm, near-neutral surface charges, and PEGylation levels close to 0.3 mol%, indicating strong convergence toward a shared optimal profile. These findings not only validate the effectiveness of the GA in identifying high-performing candidates but also reinforce design patterns observed in earlier model-driven rankings. In particular, the convergence toward moderate PEGylation and near-neutral charge mirrors experimental literature that associates such profiles

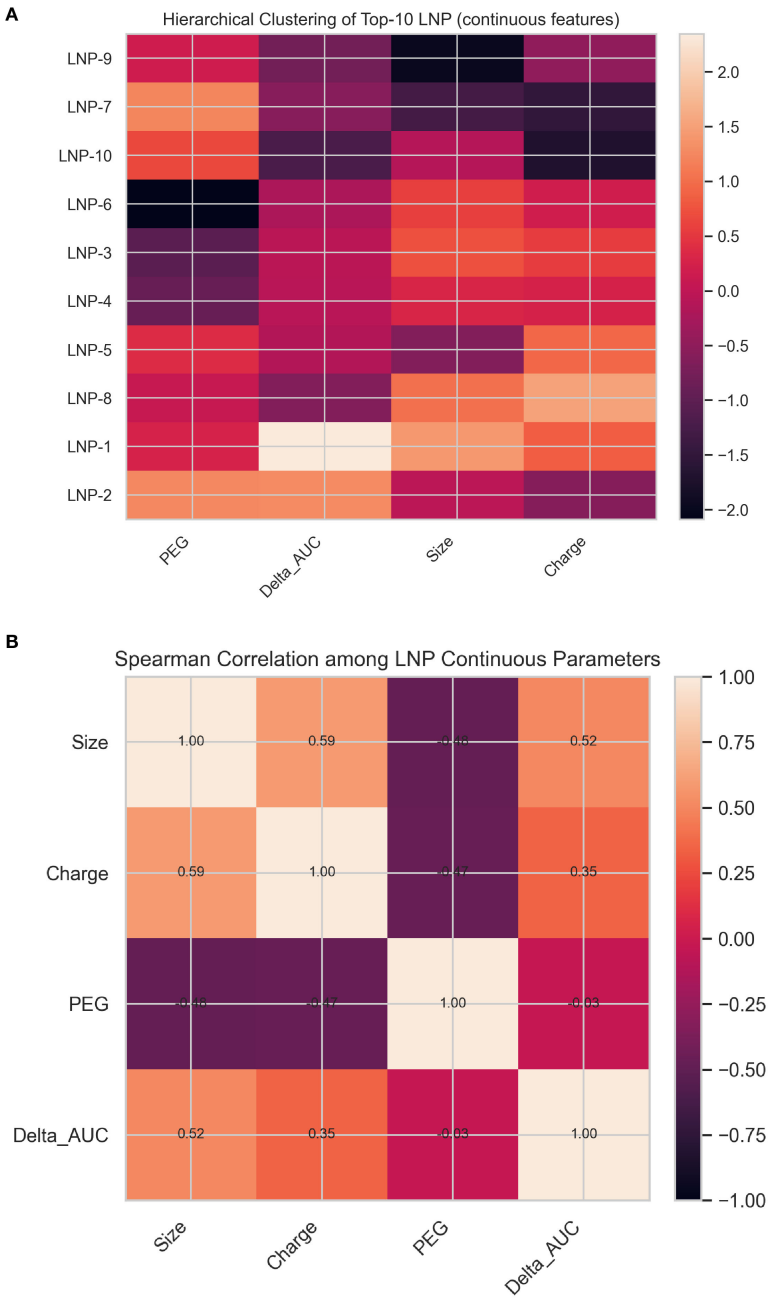


FIGURE 5
(A) Hierarchical clustering of the top 10 GA-optimized LNP formulations. Each column represents a normalized (z-scored) physicochemical parameter, and each row corresponds to an optimized LNP ranked by predicted ΔAUC. The color gradient indicates relative deviation from the mean, highlighting co-variation patterns among size, charge, PEG content, and predicted performance. **(B)** Spearman correlation matrix illustrating relationships among key continuous parameters. Positive correlations between size and charge, and negative associations with PEG content, reflect the balance between stability and delivery efficiency captured by the optimization framework.

with optimal biodistribution and reduced innate immune activation.

The distribution of predicted ΔAUC scores for the top 10 LNP candidates is shown in Figure 4. Notably, LNP-2 achieved the highest predicted score, followed by a gradual decline among the subsequent formulations.

To complement the tabulated summary of physicochemical features (Table 2), we generated a heatmap (Figure 5A) to provide a visual overview of parameter distributions among the top 10 GA-optimized LNP candidates. As previously noted, the selected formulations exhibited broadly consistent trends across size, surface charge, PEGylation, and targeting, reflecting convergence toward a shared optimal physicochemical profile. The heatmap reinforces these findings, offering an intuitive depiction of the design space occupied by the top-performing nanoparticles.

TABLE 2 Physicochemical characteristics and predicted ΔAUC values of the top 10 LNP formulations identified through genetic algorithm optimization.

Rank	Size (nm)	Charge (mV)	PEG (%)	Targeting	Predicted ΔAUC
1	91.2	-2.2	0.34	Yes	1.5
2	89.7	-2.2	0.29	Yes	1.73
3	91.6	-1.1	0.3	Yes	1.6
4	93.8	-4.4	0.26	Yes	0.99
5	89.4	-4.1	0.28	Yes	1.38
6	89.4	-2.3	0.3	Yes	1.42
7	93.9	-3	0.27	Yes	1.32
8	91.9	-1	0.31	Yes	1.4
9	88.8	-2.9	0.28	Yes	1.32
10	91.4	-3.6	0.29	Yes	1.39

Each formulation is characterized by its particle size, surface charge, PEGylation percentage, and presence of targeting ligands.

The heatmap presents the z-score–normalized physicochemical parameters—PEG content, predicted ΔAUC, particle size, and surface charge—for the ten GA-optimized LNP formulations.

Two main patterns emerge:

1. Consistency in design parameters:

Most top-performing LNPs occupy a narrow region of the design space, showing moderate PEG percentages (~0.27–0.31%), near-neutral to slightly negative charges (−4 to −1 mV), and diameters close to 90–94 nm. This convergence indicates that the optimization process favored formulations with balanced stability and cellular uptake potential.

2. ΔAUC-driven clustering:

The ΔAUC column highlights subtle differences in predicted immunogenic performance across formulations. LNP-1 and LNP-8 exhibit the highest relative ΔAUC (lighter shades), while others form a compact cluster with lower but comparable predicted responses, reflecting minor variations around the optimal region.

TABLE 3 To assess the plausibility of the simulated immune response, the synthetic transcriptomic signatures have been validated against public RNA-seq data (GSE171110).

Metric	Value	Description
Universe (shared genes)	11,342	Common genes between simulated and GSE171110 datasets
Simulated DEGs	43	DEGs identified in the synthetic dataset
Validation DEGs (GSE171110)	3,625	DEGs identified in the public RNA-seq dataset
Overlap	8 genes	Shared DEGs between simulated and real datasets
Fisher's exact test	p = 0.0707	Significance of overlap
Concordant direction	62.5%	DEGs with matching up/down-regulation
Pearson correlation (log2FC)	r = 0.22	Correlation of fold-change magnitudes

Overall, the figure visually reinforces the model-driven convergence toward an optimal physicochemical profile, characterized by ~90 nm size, low PEG content, and slightly negative charge, consistent with literature-reported parameters for clinically validated mRNA-LNP systems.

To further explore interdependencies among physicochemical variables, a Spearman correlation matrix (Figure 5B) was computed using the top 10 GA-optimized LNP formulations.

Size and surface charge showed moderate positive correlation (ρ = 0.59), while PEG content was inversely correlated with both size and charge, indicating that formulations with lower PEG fractions tend to have slightly larger and less negatively charged particles.

Collectively, these findings illustrate the effectiveness of the machine learning–driven design strategy in prioritizing LNP formulations for further refinement and experimental validation. This approach provides a rational and scalable pathway for accelerating the development of safe and effective mRNA delivery systems.

Finally, to evaluate whether the simulated transcriptomic patterns and model-driven predictions align with experimentally observed vaccine responses, we performed an external validation using public RNA-seq data from COVID-19–vaccinated individuals (GSE171110).

The results of this comparative analysis are presented in the following section.

3.5 Biological validation of simulated transcriptomics

To assess the biological plausibility of the simulated immune response, we validated the synthetic transcriptomic signatures against a public RNA-seq dataset (GSE171110) profiling peripheral blood samples from COVID-19–vaccinated and healthy individuals.

This dataset was selected because it captures *in vivo* immune activation after SARS-CoV-2 vaccination, closely reflecting the biological processes represented in our simulation (e.g., B-cell, T-cell, and immunoglobulin upregulation).

Differential gene expression analysis was performed using DESeq2 on both datasets with identical thresholds (|log2FC| > 0.5, FDR < 0.1). Comparative validation metrics were then computed between the two sets of differentially expressed genes (DEGs), including overlap significance (Fisher's exact test), directionality concordance, and Pearson correlation of log2 fold-changes.

These metrics are derived from the comparative analysis between simulated and real datasets and do not represent raw biological measurements.

As shown in Table 3, although the overlap between simulated and experimental DEGs was modest (8 shared genes, Fisher's p = 0.0707), 62.5 % of them displayed concordant regulation direction, and the overall fold-change correlation (r = 0.22) indicated a positive trend in expression magnitude, supporting the biological plausibility of the simulated immune response.

These results confirm that the simulated immune activation patterns, particularly those involving B-cell and plasma-cell markers, exhibit partial but consistent agreement with experimental vaccine transcriptomics. The positive correlation and directional concordance demonstrate that the synthetic simulation preserves biologically plausible immune activation trends without overfitting to specific datasets.

This validation step provides an important bridge between in silico predictions and experimental evidence, reinforcing the translational relevance of the proposed computational framework.

3.6 Comparison with existing COVID-19 mRNA–LNP formulations and experimental response variables

To contextualize the optimized LNPs generated by the in silico framework, their physicochemical characteristics were compared with those reported for clinically validated mRNA–LNP formulations, such as those used in the authorized COVID-19 mRNA vaccines. The parameter space explored in this study (particle size 50–150 nm, surface charge −10 to +10 mV, PEGylation 0.1–0.5 mol % and targeting presence/absence) was designed to represent generic LNPs carriers before mRNA encapsulation.

Publicly available data indicate vaccine LNPs to be small (80–100 nm), slightly negative (~ −5 mV), to contain PEG-lipids around 1.5–2 mol%, and to lack active targeting. Our optimized LNPs converge to the same size window (~90–92 nm) and to a similarly neutral/slightly negative charge, but to a lower PEGylation (~0.30 mol%) and to the presence of targeting ligands (33, 34).

These parameters are summarized in Table 4, together with the corresponding optimized values obtained from the top 10 genetic-algorithm candidates. The ideal LNP identified in this study falls within the experimentally observed range of vaccine-like LNPs, while exhibiting slightly more neutral surface charge, lower PEG-lipid content, and active targeting features predicted to enhance biodistribution and reduce off-target immune activation.

Despite being generated from a pre-encapsulation design space, the optimized LNPs fall within the clinically observed ranges for size and surface charge. Two systematic differences emerge: (i) the optimized candidates feature a lower PEG fraction (~0.30 mol%) than marketed vaccines (1.5–2 mol%), and (ii) they all include targeting ligands, while current products do not. The first difference reflects that our simulations considered PEGylation as an adjustable parameter within a simplified lipid mixture; extending the PEG dimension to 0–2 mol% in future simulations would be straightforward and would not require changing the optimization logic. The second difference reflects the objective function used here, which rewarded predicted improvements in delivery specificity and reduced off-target immune activation; this is consistent with next-generation LNPs but not yet with first-generation COVID-19 products.

The incorporation of mRNA is known to slightly alter these physicochemical properties, generally increasing particle size by 5–15 nm and shifting the surface charge neutrality, while maintaining values within the same overall range (35).

To provide an experimental reference for the biological effects associated with these physicochemical parameters, Table 5 summarizes how the key response variables, such as biodistribution, cellular uptake, and immunogenicity, are typically evaluated in mRNA–LNPs vaccines.

These variables are quantified through established experimental methods, such as *in vivo* imaging or qPCR for biodistribution, flow cytometry or confocal microscopy for cellular uptake, and immunoassays (ELISA, ELISpot, cytokine profiling) for immunogenicity (34, 36).

The reported experimental ranges highlight consistent biological behaviors across LNP-based vaccine systems, supporting the predictive validity and translational relevance of the optimized *in silico* framework.

A more detailed comparison between the optimized in silico parameters and experimental data from recent literature is provided in Supplementary Table S1.

TABLE 4 The table summarizes typical measurement methods and value ranges for biodistribution, cellular uptake, and immunogenicity reported in experimental studies of mRNA–LNP vaccines.

Response variable	Measurement method	Representative experimental values
Biodistribution	*In vivo* imaging of labeled LNPs, qPCR of mRNA per organ. (%ID/g)	Liver 40–60%; spleen 10–20%ID/g at 6-24h post-dose
Cellular uptake	Flow cytometry or confocal microscopy of LNP-positive APCs in draining lymph node.	20–50% positive cells depending on surface charge and PEGylation density
Immunogenicity	ELISA, ELISpot, cytokine profiling	Neutralizing Ab ≥ 1:1000; IFN-γ 100–500 pg/mL (Th1-biased)

These data outline the expected biological performance range of clinically validated formulations and support the relevance of the optimized in silico LNP profiles proposed in this work.

TABLE 5 Comparison of physicochemical parameters for vaccine-like and ideal LNPs.

Formulation	Particle size	Surface charge	PEGylation	Targeting
COVID-19 mRNA–LNP (Pfizer-like)	~90 nm (midpoint of 80–100)	~ −5 mV	~1.5 mol% (50:10:38.5:1.5)	No
COVID-19 mRNA–LNP (Moderna-like)	~90 nm	~ −5 mV	~1.5–2 mol%	No
Ideal LNP	~91 nm (midpoint of 88.8–93.9 nm)	~ −2.7 mV (range −1 − 4.4 mV)	~ 0.30 mol% (range 0.26–0.34 mol%)	Yes

The ideal LNP remains within the experimentally observed range of mRNA–LNP formulations but shows a more neutral charge, lower PEG content, and active targeting, features predicted to enhance biodistribution and reduce off-target immune activation.

This comparison indicates that the in silico search was conducted within clinically realistic physicochemical boundaries, while deliberately extending the design space toward targeted and lower-PEGylation to explore safer delivery profiles.

4 Discussion

The unprecedented success of mRNA vaccines against COVID-19 has propelled messenger RNA technology to the forefront of vaccinology, showcasing its potential for rapid development and high efficacy. Central to this success is the role of lipid nanoparticles (LNPs), which have emerged as the most clinically advanced non-viral platforms for mRNA delivery. LNPs protect the fragile mRNA strands and facilitate their efficient delivery into cells, overcoming previous challenges associated with mRNA therapeutics.

Building upon this foundation, our study presents an in silico framework that bridges mechanistic immune simulations with AI-driven optimization strategies to guide the rational design of safer and more effective mRNA vaccine delivery systems. By leveraging synthetic RNA-seq data aligned with immune activation patterns, predicted by multiscale simulations, and integrating these insights into a machine learning–guided formulation pipeline, we demonstrate a systematic approach to optimizing LNP parameters under biologically informed constraints.

Traditional Design of Experiments (DOE) methodologies have historically played a central role in formulation development by enabling structured exploration of formulation variables and their interactions. However, while DOE remains a cornerstone of experimental design, its reliance on extensive empirical data collection can limit its scalability, particularly in complex biological systems where multidimensional interactions are critical. Our in silico framework complements and extends the DOE philosophy by virtually exploring the formulation space, thereby significantly reducing experimental burden while maintaining a systematic and interpretable optimization process.

The application of a genetic algorithm, coupled with a predictive model trained on physicochemical attributes, enabled the identification of top-performing formulations that consistently shared favorable traits such as near-neutral charge, moderate PEGylation, and optimal size. These features are well-established in the literature as critical for efficient biodistribution and reduced immunogenicity of nanoparticle systems. Beyond enhancing delivery precision, this pipeline offers a powerful tool for hypothesis generation, dramatically reducing the need for costly and time-consuming *in vivo* screening in early-stage vaccine development.

Interestingly, the optimized LNP parameters predicted by our AI-guided workflow (~90 nm diameter, near-neutral charge, and ~0.3% PEG) are consistent with experimental findings reported in previous studies (37, 38).

The optimized formulations identified by our algorithm—ranging between 88.8 and 93.9 nm in diameter, with surface charges between −1.0 and −4.4 mV and PEGylation levels of 0.26–0.34 mol%—thus fall squarely within the experimental range associated with efficient lymphatic transport and reduced innate immune activation. This strong convergence between simulated and experimentally validated parameters reinforces the reliability and practical significance of our in silico design framework.

5 Limitations

This study introduces and tests a computational framework for in silico vaccine design by integrating artificially generated RNA-seq data and simulated immune activation profiles derived from a previously validated UISS-COVID19 model. While simulated datasets cannot fully capture the complexity and heterogeneity of biological systems, they provide a valuable platform for prototyping analytical pipelines, exploring mechanistic hypotheses, and informing experimental design in data-scarce contexts.

The synthetic RNA-seq data were generated under biologically grounded assumptions, including expected transcriptional shifts following mRNA vaccination and compartment-specific immune activation. Simulated immune activation scores (Delta_AUC) were assigned to virtual lipid nanoparticle (LNP) formulations using a custom scoring function to reflect known principles of biodistribution and immunogenicity. These components were combined with AI-based optimization strategies, such as random forest regression and genetic algorithms, to identify LNP configurations predicted to minimize off-target activation and maximize delivery efficiency.

All transcriptomic data were simulated and must ultimately be validated using experimental datasets. Similarly, the predictive model was trained on artificially generated Delta_AUC values, which, although biologically plausible, do not replace empirical measurements. The framework is modular and scalable, but its predictive accuracy remains sensitive to the assumptions embedded in the simulation and data generation processes. Therefore, all findings derived from synthetic data should be interpreted as proof-of-concept rather than biological evidence.

Nonetheless, this in silico foundation offers a powerful tool for early-phase vaccine development, enabling efficient hypothesis generation, risk estimation, and preclinical prioritization of candidate formulations prior to experimental validation.

6 Conclusion

This study presents a novel in silico pipeline that integrates multiscale immune simulation outcomes with synthetic RNA-seq data and machine learning algorithms to systematically identify optimized mRNA-LNP formulations. By simulating post-vaccination gene expression profiles and using these to guide the selection of physiochemically favorable LNP candidates, our framework provides a rational and scalable approach for early-stage vaccine design. The integration of a predictive model with a genetic algorithm allowed us to converge on nanoparticle configurations exhibiting key features, such as near-neutral surface charge, appropriate particle size, and moderate PEGylation, associated with enhanced delivery efficiency and minimal off-target effects.

Our findings underscore the feasibility of computational vaccine design, complementing and accelerating empirical approaches that are often time-consuming, costly, and ethically challenging due to the need for extensive *in vivo* testing. The pipeline supports more sustainable and reproducible development processes by minimizing experimental burden and enabling rapid, data-driven iteration.

Moreover, the framework is modular and adaptable: it can be extended to incorporate patient-derived transcriptomic data, support personalized vaccine strategies, or be applied to other therapeutic delivery systems beyond mRNA, such as siRNA, CRISPR components, or protein-based biologics. Its compatibility with existing data standards and modeling infrastructures also makes it suitable for integration into industrial development pipelines and regulatory decision-making workflows. As computational tools continue to evolve, this integrative strategy holds promise for accelerating the development of safe, targeted, and cost-effective immunotherapies and vaccines with wide-ranging applications in infectious disease, oncology, and beyond.

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. All R/Python scripts used for data generation, analysis, and figure rendering have been made publicly available to ensure full reproducibility, in line with Frontiers' Open Science policy, at the following GitHub repository: https://github.com/ValeDS/A-Computational-Framework-for-Optimizing-mRNA-Vaccine-.

Author contributions

VDS: Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Visualization. FC: Writing – review & editing, Visualization. GR: Writing – original draft, Writing – review & editing, Visualization. FP: Writing – review & editing, Supervision, Funding acquisition.

Funding

The author(s) declared that financial support was received for this work and/or its publication. FP and VD are funded through the Programma di ricerca CN00000013 "National Centre for HPC, Big Data and Quantum Computing", finanziato dal Decreto Direttoriale di concessione del finanziamento n.1031 del 17.06.2022 a valere sulle risorse del PNRR–M4C2—Investimento 1.4—Avviso "Centri Nazionali" —D.D. n. 3138 del 16 dicembre 2021.

Acknowledgments

The authors acknowledge the use of generative AI tools (ChatGPT, OpenAI) to support the drafting and linguistic editing of this manuscript. All content generated was critically reviewed and approved by the authors.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI (ChatGPT, OpenAI) was used to assist with the following tasks:•Drafting and refining sections of the manuscript text •Improving scientific language clarity and grammar Generating concise summaries for the running title and keywords Structuring formal declarations. All AI-assisted content was critically reviewed, edited, and approved by the authors to ensure scientific accuracy and integrity.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.2025.1628583/full#supplementary-material

References

1. Wilson B, Geetha KM. Lipid nanoparticles in the development of mRNA vaccines for COVID-19. *J Drug Delivery Sci Technol*. (2022) 74:103553. doi: 10.1016/j.jddst.2022.103553

2. Cahn D, Duncan GA. High-density branched PEGylation for nanoparticle drug delivery. *Cell Mol Bioeng*. (2022) 15:355–66. doi: 10.1007/s12195-022-00727-x

3. Cui X, Vervaeke P, Gao Y, Opsomer L, Sun Q, Snoeck J, et al. Immunogenicity and biodistribution of lipid nanoparticle formulated self-amplifying mRNA vaccines against H5 avian influenza. *NPJ Vaccines*. (2024) 9:138. doi: 10.1038/s41541-024-00932-x

4. Tavares Luiz M, Viegas JSR, Abriata JP, Viegas F, de Carvalho Vicentini FTM, Badra Bentley MVL, et al. Design of experiments (DoE) to develop and to optimize nanoparticles as drug delivery systems. *Eur J Pharmaceutics Biopharmaceutics*. (2021) 165:127–48. doi: 10.1016/j.ejpb.2021.05.011

5. Musumeci T, Bonaccorso A, Carbone C, Russo G, Pappalardo F, Puglisi G. Design and optimization of PEGylated nanoparticles intended for Berberine Chloride delivery. *J Drug Delivery Sci Technol*. (2019) 52:521–30. doi: 10.1016/j.jddst.2019.05.012

6. Dorsey PJ, Lau CL, Chang T, Doerschuk PC, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. *J Pharm Sci*. (2024) 113:3413–33. doi: 10.1016/j.xphs.2024.09.015

7. Di Salvatore V, Russo G, Pappalardo F. (2024). Unintended risks of mRNA COVID-19 vaccines: A UISS simulation study on immune and organ health, in: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Piscataway, New Jersey, USA: Institute of Electrical and Electronics Engineers. pp. 6661–6. doi: 10.1109/BIBM62325.2024.10822378

8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8

9. Liaw A, Wiener M. Classification and regression by randomForest. *Forest*. (2001) 23:18–22.

10. Scrucca L. GA : A package for genetic algorithms in R. *J Stat Softw*. (2013) 53:1–37. doi: 10.18637/jss.v053.i04

11. Wickham H. ggplot2. *in Use R!*. Cham: Springer International Publishing (2016). doi: 10.1007/978-3-319-24277-4

12. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. (2007) 9:90–5. doi: 10.1109/MCSE.2007.55

13. Waskom M. seaborn: statistical data visualization. *J Open Source Softw*. (2021) 6:3021. doi: 10.21105/joss.03021

14. Pavlasova G, Mraz M. The regulation and function of CD20: an 'enigma' of B-cell biology and targeted therapy. *Haematologica*. (2020) 105:1494–506. doi: 10.3324/haematol.2019.243543

15. Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. *Annu Rev Immunol*. (2009) 27:591–619. doi: 10.1146/annurev.immunol.021908.132706

16. Minnich M, Tagoh H, Bönelt P, Axelsson E, Fischer M, Cebolla B, et al. Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol*. (2016) 17:331–43. doi: 10.1038/ni.3349

17. Russo G, Di Salvatore V, Sgroi G, Parasiliti Palumbo GA, Reche PA, Pappalardo F. A multi-step and multi-scale bioinformatic protocol to investigate potential SARS-CoV-2 vaccine targets. *Brief Bioinform*. (2022) 23. doi: 10.1093/bib/bbab403

18. Palladini A, Nicoletti G, Pappalardo F, Murgo A, Grosso V, Stivani V, et al. In silico Modeling and *In vivo* Efficacy of Cancer-Preventive Vaccinations. *Cancer Res*. (2010) 70:7755–63. doi: 10.1158/0008-5472.CAN-10-0701

19. Huang D, Liu AYN, Leung K-S, Tang NLS. Direct measurement of B lymphocyte gene expression biomarkers in peripheral blood transcriptomics enables early prediction of vaccine seroconversion. *Genes (Basel)*. (2021) 12:971. doi: 10.3390/genes12070971

20. Ai J, Guo J, Zhang H, Zhang Y, Yang H, Lin K, et al. Cellular basis of enhanced humoral immunity to SARS-CoV-2 upon homologous or heterologous booster vaccination analyzed by single-cell immune profiling. *Cell Discov*. (2022) 8:114. doi: 10.1038/s41421-022-00480-5

21. Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, et al. Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity*. (2020) 53:1296–314.e9. doi: 10.1016/j.immuni.2020.11.017

22. Shi R, Liu X, Wang Y, Pan M, Wang S, Shi L, et al. Long-term stability and immunogenicity of lipid nanoparticle COVID-19 mRNA vaccine is affected by particle size. *Hum Vaccin Immunother*. (2024) 20. doi: 10.1080/21645515.2024.2342592

23. Haghighi E, Abolmaali SS, Dehshahri A, Mousavi Shaegh SA, Azarpira N, Tamaddon AM. Navigating the intricate *in-vivo* journey of lipid nanoparticles tailored for the targeted delivery of RNA therapeutics: a quality-by-design approach. *J Nanobiotechnology*. (2024) 22:710. doi: 10.1186/s12951-024-02972-w

24. Catenacci L, Rossi R, Sechi F, Buonocore D, Sorrenti M, Perteghella S, et al. Effect of lipid nanoparticle physico-chemical properties and composition on their interaction with the immune system. *Pharmaceutics*. (2024) 16:1521. doi: 10.3390/pharmaceutics16121521

25. Lin Y, Cheng Q, Wei T. Surface engineering of lipid nanoparticles: targeted nucleic acid delivery and beyond. *Biophys Rep*. (2023) 9:255. doi: 10.52601/bpr.2023.230022

26. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. (2010) 11:R106. doi: 10.1186/gb-2010-11-10-r106

27. Hassett KJ, Benenato KE, Jacquinet E, Lee A, Woods A, Yuzhakov O, et al. Optimization of lipid nanoparticles for intramuscular administration of mRNA vaccines. *Mol Ther Nucleic Acids*. (2019) 15:1–11. doi: 10.1016/j.omtn.2019.01.013

28. Bhattacharya S, Prajapati BG, Singh S. A critical review on the dissemination of PH and stimuli-responsive polymeric nanoparticular systems to improve drug delivery in cancer therapy. *Crit Rev Oncol Hematol*. (2023) 185:103961. doi: 10.1016/j.critrevonc.2023.103961

29. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

30. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Springer Texts in Statistics. New York, NY: Springer US (2021). doi: 10.1007/978-1-0716-1418-1

31. Turner JS, O'Halloran JA, Kalaidina E, Kim W, Schmitz AJ, Zhou JQ, et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature*. (2021) 596:109–13. doi: 10.1038/s41586-021-03738-2

32. Ernst L, Casals E, Italiani P, Boraschi D, Puntes V. The interactions between nanoparticles and the innate immune system from a nanotechnologist perspective. *Nanomaterials*. (2021) 11:2991. doi: 10.3390/nano11112991

33. Schoenmaker L, Witzigmann D, Kulkarni JA, Verbeke R, Kersten G, Jiskoot W, et al. mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability. *Int J Pharm*. (2021) 601:120586. doi: 10.1016/j.ijpharm.2021.120586

34. Hou X, Zaks T, Langer R, Dong Y. Lipid nanoparticles for mRNA delivery. *Nat Rev Mater*. (2021) 6:1078–94. doi: 10.1038/s41578-021-00358-0

35. Mamberti S, Pesce C, Avancini G, Somu G, Govinda N, Kundoor R, et al. On the retrograde transport of RNA-loaded lipid nanoparticles designed for brain delivery. *ACS Nanoscience Au*. (2025) 5:375–87. doi: 10.1021/acsnanoscienceau.5c00042

36. Liu W, Zhang M, Lv H, Yang C. Formulation-driven optimization of PEG-lipid content in lipid nanoparticles for enhanced mRNA delivery *in vitro* and *in vivo*. *Pharmaceutics*. (2025) 17:950. doi: 10.3390/pharmaceutics17080950

37. Hassett KJ, Benenato KE, Jacquinet E, Lee A, Woods A, Yuzhakov O, et al. Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines. *Mol Ther Nucleic Acids*. (2019) 15:1–11. doi: 10.1016/j.omtn.2019.01.013

38. Zhang H, Zhang L, Lin A, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature*. (2023) 621:396–403. doi: 10.1038/s41586-023-06127-z

Frontiers in Immunology

Explores novel approaches and diagnoses to treat immune disorders.

The official journal of the International Union of Immunological Societies (IUIS) and the most cited in its field, leading the way for research across basic, translational and clinical immunology.

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