Thesis for the Degree of Doctor of Philosophy

Design and evaluation of multi-epitope vaccine candidates utilizing bioinformatics approaches

생물정보학적 접근을 통한 다중 항원결정기 백신 후보물질의 설계 및 평가

February 2025

The Graduate School
Seoul National University
Biomodulation Major

Nguyen Thi Truc Ly

Design and evaluation of multi-epitope vaccine candidates utilizing bioinformatics approaches

Advised by Professor Heebal Kim

Submitting a Ph.D. Dissertation of
Biomodulation

December 2024

The Graduate School
Seoul National University
Biomodulation Major

Nguyen Thi Truc Ly

Confirming the Ph.D. Dissertation written by
Nguyen Thi Truc Ly
December 2024

Chairman: Cheol-Heui Yun (Seal)

Vice Chairman: Heebal Kim (Seal)

Committee Member: Sungho Won (Seal)

Committee Member: Woori Kwak (Seal)

Committee Member: Minseok Seo (Seal)

Abstract

Vaccines have long been a foundation in the prevention and control of infectious diseases, representing one of the most successful public health interventions in history. From the eradication of smallpox to the control of COVID-19, vaccines have saved millions of lives worldwide. However, despite these remarkable successes, several pathogens, including those responsible for bacterial and viral infections, continue to pose a significant threat to global health. With infectious diseases emerging and re-emerging becoming one of the leading causes of mortality globally, there is an urgent need to develop novel vaccine strategies that can address these evolving challenges.

Traditional vaccine approaches, which rely on live attenuated or inactivated pathogens, face significant limitations. These methods may not always elicit the desired immune response, especially in immunocompromised individuals, and there is always the risk of reversion to a virulent form in live vaccines. Moreover, in the context of rapidly mutating viruses, such as SARS-CoV-2, traditional approaches struggle to keep up with the antigenic diversity. Thus, there is a critical need for innovative approaches in vaccine design that can overcome these limitations and provide long-lasting protection against infectious diseases.

In this context, bioinformatics has emerged as a powerful tool to enhance vaccine development. Immunoinformatics, a specialized field within bioinformatics, focuses on analyzing immune responses and predicting epitopes that can stimulate both cellular and humoral immunity. This field has enabled the strategic design of vaccines by identifying specific segments of proteins, known as epitopes, that are recognized by the immune system. These epitopes can be used to create multi-epitope vaccines, which combine several immunogenic epitopes to elicit robust and targeted immune responses. Computational methods significantly reduce the time and cost involved in vaccine development while providing a platform for designing vaccines that are safer, more effective, and highly adaptable.

Multi-epitope vaccines represent a promising advancement in vaccinology. Unlike traditional vaccines that rely on whole pathogens or large subunits, multi-epitope vaccines are designed to incorporate only the most immunogenic regions of antigens. These epitopes stimulate both cytotoxic T lymphocytes and helper T lymphocytes, as well as B cells, creating a comprehensive and targeted immune response. Furthermore, by selecting epitopes from multiple antigens, multi-epitope vaccines address antigenic variability and offer broader protection against rapidly evolving pathogens.

This thesis explores the design and evaluation of multi-epitope vaccine candidates for three significant infectious diseases: Powassan virus, tuberculosis, and respiratory syncytial virus. These diseases represent critical global health challenges due to their rising incidence, significant mortality rates, and lack of effective vaccines. Employing advanced bioinformatics tools, including immunoinformatics and computational biology, potential vaccine candidates were designed and rigorously evaluated. The findings demonstrate the capability of computational vaccine technology to rapidly design adaptable and effective vaccines for emerging and persistent health threats.

Keywords: Multi-epitope vaccine, Immunoinformatics, Bioinformatics, Infectious diseases, Computational vaccinology

Student number: 2020-34548

Contents

List of Tables

List of Figures

Chapter 1 General Introduction

1.1. The global impact of infectious diseases

Infectious diseases are among the most critical challenges to global health, impacting millions of individuals and placing a substantial burden on healthcare systems worldwide. Caused by various pathogens, including viruses, bacteria, fungi, protozoa, and parasites, these diseases range from mild infections to severe pandemics (Mishra et al., 2022). Over the past century, vaccines have played a fundamental role in controlling and, in some cases, eradicating diseases such as smallpox, measles, polio, and more recently, COVID-19 (Henderson, 1980; Moss, 2017; Polack et al., 2020; Salk, 1955; Voysey et al., 2021). These vaccination efforts have prevented millions of deaths globally, emphasizing immunization's essential role in public health and disease prevention.

Nevertheless, infectious disease threats continue to evolve, with emerging and re-emerging pathogens posing significant challenges. Climate change is projected to expand tropical regions globally, which may increase the spread of vector-borne diseases like the Powassan virus (POWV) (Booth, 2018; Robert et al., 2020). The dynamics of climate change, including rising temperatures and rainfall variability, create favorable conditions for the spread of various vectors, particularly ticks and mosquitoes, which are known carriers

of diseases like POWV (Tidman et al., 2021; Wilson et al., 2020). Additionally, human trade and travel can facilitate the transmission of POWV from sylvatic to urban cycles, potentially leading to novel outbreaks among human and animal populations (Kraemer et al., 2019; Massengo et al., 2023). POWV is an emerging infectious disease that can cause severe neurological illness in humans (Della-Giustina et al., 2021). Clinical presentations vary from asymptomatic cases to severe neuroinvasive disease, with an incubation period of 1–5 weeks (Corrin et al., 2018; Hermance & Thangamani, 2017). Symptoms include fever, headache, vomiting, weakness, and, in severe cases, neurological complications such as encephalitis (inflammation of the brain) and meningitis (inflammation of the membranes surrounding the brain and spinal cord) can occur (Corrin et al., 2018; Telford & Piantadosi, 2023). Currently, no specific antiviral therapy or vaccine exists for POWV infections. Therefore, developing effective vaccines to mitigate the rising risk of POWV is a pressing global public health priority.

At the same time, airborne infections like respiratory syncytial virus (RSV) and tuberculosis (TB) consistently pose a threat to world health. TB affects over 10 million individuals annually and remains a leading cause of mortality in low and middle-income countries (Chakaya et al., 2022). Despite being the world's second leading cause

of death from a single infectious agent, after COVID-19, TB treatment and control efforts have been disrupted by the COVID-19 pandemic (WHO, 2023). The development of effective TB vaccines has been challenging, although promising innovations such as DNA vaccines and live attenuated and killed whole-cell vaccines (WCVs) offer hope. However, like the Bacillus Calmette-Guerin (BCG) vaccine, tools to fight TB are imperfect. The BCG vaccine is effective (70-80%) in preventing severe childhood TB forms, such as meningitis and miliary TB, but its efficacy in adult pulmonary TB varies and may be influenced by environmental exposure to mycobacteria (Katelaris et al., 2020; McShane, 2011). Research suggested that BCG's immune response differs across populations due to genetic and environmental factors, with potential heterologous effects that may enhance immunity against other pathogens (Darboe et al., 2017; Villanueva et al., 2023). These studies underscore the challenge in TB vaccination strategies and highlight the need for complementary vaccines.

Besides, airborne viruses such as respiratory syncytial virus (RSV) can cause severe illness across all age groups. The COVID-19 pandemic has led to delayed RSV outbreaks and increased reported cases across multiple countries (Chuang et al., 2023; Zhou et al., 2024). Annually, RSV is estimated to cause 33 million new cases of acute

lower respiratory tract infections in children under five globally, resulting in approximately 3 million hospitalizations and 120,000 deaths (Colosia et al., 2023; Mejias et al., 2019). In adults, RSV can exacerbate pre-existing cardiopulmonary conditions, particularly in elderly populations, leading to 60,000 to 160,000 hospitalizations and 6,000 to 10,000 deaths annually in the United States (Harris, 2023). Despite the high incidence of RSV infections, particularly among infants and older adults, current treatment options are limited, and no effective vaccine is available. This underscores the critical need for continued efforts to develop strategies for RSV prevention and control.

These three diseases, POWV, TB, and RSV, illustrate diverse challenges in infectious disease control. While POWV represents an emerging vector-borne threat, TB demonstrates the complexities of combating a persistent bacterial pathogen, and RSV highlights the ongoing struggle with airborne viral infections. Together, they underscore the urgent need for innovative vaccine strategies capable of addressing a wide range of pathogens. In this thesis, by focusing on immunogenic regions of pathogens, multi-epitope vaccines offer a strategic advantage, enabling targeted immune responses against multiple antigenic sites. This approach could improve protective efficacy against diverse pathogens. The application of computational

tools for epitope selection allows for rapid and accurate vaccine design, positioning multi-epitope vaccines as a highly adaptable approach to the evolving challenges in infectious diseases.

1.2. Limitations of traditional vaccine approaches

“If God did not exist, it would be necessary to invent him”, remarked the great French philosopher Voltaire (1694–1778). The same could be said about vaccines; we would have to create them if they didn't exist because they are so helpful (Flower, 2008). Indeed, vaccines have transformed public health. In countries with high coverage of vaccine programs, many of the diseases that were previously responsible for the majority of deaths have disappeared or near disappeared, such as polio (paralytic), diphtheria, rubella, tetanus, haemophilus influenzae type B, measles, mumps, hepatitis B, pertussis, etc. (Abbas et al., 2021).

A vaccine is a biological preparation that provides active acquired immunity to a particular infectious disease. It typically contains an agent that resembles a disease-causing microorganism and is often made from weakened or killed forms of the microbe, its toxins, or one of its surface proteins. Vaccines work by stimulating the immune system to recognize the agent as a threat, destroy it, and

remember it, thus enabling the immune system to respond more effectively if exposed to the actual pathogen in the future (Chen et al., 2022; Daddario-DiCaprio et al., 2006; Poland et al., 2020).

Traditional vaccine approaches, such as inactivated or live-attenuated vaccines, have been foundational in immunization strategies. Live-attenuated vaccines, for example, can confer long-lasting immunity after a single administration, as seen with the yellow fever vaccine, which protects for decades (Daddario-DiCaprio et al., 2006). However, these traditional methods have notable limitations. One significant challenge is the variability in vaccine efficacy across different populations, particularly in vulnerable groups such as the elderly or immunocompromised individuals. Studies have shown that the efficacy of inactivated influenza vaccines is often reduced in older adults, a phenomenon attributed to immunosenescence (Ramirez et al., 2016; Vilches et al., 2021). This reduced efficacy can lead to increased morbidity and mortality in these populations, highlighting a critical limitation of traditional vaccine formulations.

Moreover, the development of vaccines using traditional methods can be time-consuming and may not keep pace with emerging infectious diseases. For example, the rapid evolution of viruses like influenza complicates the creation of broadly effective vaccines, as

seen in the challenges faced during the COVID-19 pandemic (Chen et al., 2022; Poland et al., 2020). The traditional approach of using inactivated viruses may not provide the necessary rapid response required during outbreaks, as evidenced by the urgent need for effective vaccines against SARS-CoV-2 (Pormohammad et al., 2021).

Another limitation is the reliance on pre-existing immunity to vaccine vectors, which can reduce the effectiveness of viral vector-based vaccines. For example, vaccines based on adenovirus or modified vaccinia Ankara may face challenges due to high levels of pre-existing immunity in the population (Kapadia et al., 2005). This pre-existing immunity can lead to suboptimal immune responses, further complicating vaccine development and deployment.

While traditional vaccine approaches have been instrumental in controlling infectious diseases, they face significant limitations, including variable efficacy in vulnerable populations, challenges in rapid development for emerging pathogens, issues related to pre-existing immunity, difficulty in identifying suitable antigens, immune evasion, and technical challenges. These challenges underscore the need for technologies that offer the potential for rapid development, safety, high efficacy, easier production, stability in various storage and transport conditions, and more adaptability against infectious diseases.

1.3. The emergence of bioinformatics in vaccine development

The rise of bioinformatics has significantly transformed vaccine development, providing innovative tools and methodologies that enhance vaccine design, efficacy, and safety. Bioinformatics encompasses a range of computational techniques that facilitate the identification of potential vaccine targets, the prediction of immune responses, and the optimization of vaccine formulations. This approach allows precise targeting of pathogenic components, making vaccine design more rapid and adaptable.

One of the primary advantages of bioinformatics in vaccine development is its ability to identify and characterize immunogenic epitopes. For instance, studies have demonstrated the successful application of bioinformatics tools in the design of multi-epitope vaccines against various pathogens, including Plasmodium knowlesi, where immune-protective epitopes were identified and incorporated into vaccine candidates (A., 2021). This approach has been successfully applied in various studies, including the design of vaccines against hepatitis C and Brucella, where bioinformatics methods were employed to identify and validate immunogenic epitopes (Chen et al., 2021; Guest & Pierce, 2018; Guo et al., 2023). These

examples illustrate how bioinformatics can streamline the vaccine design process by enabling researchers to predict which epitopes are likely to elicit a strong immune response.

Besides, the rapid advancements in machine learning algorithms and artificial intelligence (AI), have significantly enhanced vaccine development processes (Khemasuwan & Colt, 2021). This integration facilitates various computational techniques, including molecular modeling, molecular docking, molecular dynamics simulations, and immune simulations, which collectively contribute to predicting the stability and efficacy of vaccine constructs. One of the notable advancements in this domain is the use of AlphaFold, a deep learning model developed by DeepMind, which has revolutionized protein structure prediction (Jumper et al., 2021). AlphaFold's ability to accurately predict the three-dimensional structures of proteins allows researchers to understand the vaccine structure better, thereby aiding in the design of effective vaccines (Varadi & Velankar, 2023). Furthermore, molecular docking and dynamics simulations further complement these efforts by providing insights into the interactions between vaccine candidates and immune receptors. For example, molecular docking studies can predict the binding affinity between vaccine constructs and receptors such as Toll-like receptors (TLRs),

which play a vital role in initiating immune responses (Anwar et al., 2023). This predictive modeling allows researchers to evaluate the potential efficacy of vaccine candidates before proceeding to experimental validation. Additionally, molecular dynamics simulations can assess the stability of these interactions over time, providing a deeper understanding of how vaccine constructs may behave in biological systems (Abraham et al., 2015).

Bioinformatics has not only improved vaccine efficacy and safety but has also significantly reduced the time and cost associated with vaccine development (Pyasi et al., 2021). Traditional vaccine development processes can be lengthy and resource-intensive; however, the application of computational methods allows for rapid screening of potential vaccine candidates from large datasets. For example, the integration of next-generation sequencing data with bioinformatics approaches has expedited the identification of potential vaccine targets for SARS-CoV-2, thereby facilitating the rapid development of COVID-19 vaccines (Mukherjee, 2020). This approach is essential for addressing emerging infectious diseases and enhancing timely vaccine responses.

Overall, bioinformatics has revolutionized vaccine development by enhancing the identification of immunogenic epitopes and

streamlining the vaccine development process. The integration of bioinformatics into vaccine research represents a powerful advancement towards more precise, efficient, and adaptable vaccine solutions, meeting the urgent demands of global health challenges. The case studies in this thesis demonstrate the transformative role of bioinformatics in vaccine design. From identifying immunogenic epitopes to predicting immune responses, these tools enable rapid and precise vaccine development across a spectrum of pathogens. This unified computational framework is applied consistently across the Powassan virus, *Mycobacterium tuberculosis*, and respiratory syncytial virus case studies, underscoring its broad applicability. Table 1.1 lists several servers, databases, and software with web addresses, brief descriptions, and comments available for vaccine development.

Table 1.1 Servers, databases, and software with web addresses, brief descriptions, and comments are available for vaccine development.

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
UniProt database	https://www.uniprot.org/	The UniProt database is a comprehensive, freely accessible resource for protein sequence and functional information. It provides curated data on proteins, including amino acid sequences, functional annotations, 3D structures, and post-translational modifications.	Strengths: - Extensive data: Covers a vast array of protein information, widely used in bioinformatics and life sciences. - Frequent updates: Continuously updated to ensure access to the latest protein data. Limitations: - Complexity: The breadth of data can be overwhelming for beginners; may require bioinformatics knowledge for effective navigation.	(Consortium et al., 2023)
MHC II server of IEDB (Immune Epitope Database and Analysis Resources)	http://tools.iedb.org/mhcii/	The MHC II Server of IEDB predicts peptide binding to histocompatibility complex class II (MHC-II) molecules, aiding vaccine and immunotherapy design by identifying potential T-cell epitopes. It offers multiple prediction methods, including consensus, SMM, and NN-align algorithms, for enhanced accuracy.	Strengths: - Multiple prediction methods: Offers a range of algorithms to improve prediction reliability. - Broad allele coverage: Supports diverse HLA alleles, facilitating global applicability. Limitations: Computational load: Processing time increases with large datasets or multiple alleles.	(Fleri et al., 2017)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
NetMHCII 2.3	https://services.healthtech.dtu.dk/services/NetMHCII-2.3/	NetMHCII 2.3 is a server designed to predict binding of peptides to MHC-II molecules. It uses neural networks trained on experimental binding data to estimate binding affinities for different MHC-II alleles. This is essential for identifying helper T-cell epitopes.	Strengths: - Accurate prediction: Employs advanced neural networks for high binding prediction accuracy. - Extensive allele coverage: Covers a broad range of MHC-II alleles, supporting diverse population studies. Limitations: Processing times may be longer for large datasets.	(Jensen et al., 2018)
NetMHCIIpan-4.3 server	https://services.healthtech.dtu.dk/services/NetMHCIIpan-4.3/	NetMHCIIpan-4.3 is a server that predicts peptide binding to a wide range of MHC-II alleles, including both human and non-human alleles, by using a pan-specific method. This tool is widely used for epitope discovery in vaccine and immunotherapy research.	Strengths: - Comprehensive coverage: Supports prediction across a broad spectrum of MHC-II alleles, including rare ones, aiding in population-wide studies. - Advanced algorithm: Pan-specific model provides high accuracy in binding prediction. Limitations: Processing may be slower with large datasets due to complex algorithms.	(Nilsson et al., 2023)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
MHC I server of IEDB	http://tools.iedb.org/mhci/ Update next-generation Tools site: https://nextgen-tools.iedb.org/pipeline?tool=tc1	The MHC I Binding Prediction Tool from the IEDB is an online resource that predicts peptide binding to MHC-I molecules, essential for identifying cytotoxic T-cell epitopes in vaccine design. The tool supports multiple algorithms, enhancing prediction accuracy and flexibility.	Strengths: - Versatile prediction options: Offers multiple algorithms, including consensus and artificial neural networks, to improve accuracy. - Wide allele coverage: Includes many human and non-human MHC-I alleles, supporting diverse research needs. - Regular updates: Part of the frequently updated IEDB platform, constantly updated with new data and algorithms. Limitation: Complex calculations may lead to longer processing times for large datasets.	(Fleri et al., 2017)
NetMHCpan 4.0 Updated to NetMHCpan 4.1	https://services.healthtech.dtu.dk/services/NetMHCpan-4.0/ https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/	NetMHCpan 4.0 is a tool designed to predict peptide binding to a wide range of MHC-I molecules, using a pan-specific model. It is commonly used for epitope mapping across diverse MHC alleles, making it useful in immunology and vaccine development research. NetMHCpan 4.1 builds on version 4.0, providing enhanced accuracy for MHC-I binding predictions with updated algorithms and additional MHC allele coverage. This version is particularly useful for large-scale epitope screening projects.	Strengths: - Broad allele coverage: Covers a wide array of MHC-I alleles, including rare alleles, aiding population studies. - Accurate predictions: Utilizes advanced algorithms for reliable binding predictions. Limitations: May process slowly with large datasets due to complex calculations.	(Jurtz et al., 2017) (Reynisson et al., 2020)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
ABCpred	http://crdd.osdd.net/raghava/abcpred/	ABCpred is a web server designed to predict B-cell epitopes using an artificial neural network (ANN) model. It identifies linear B-cell epitopes based on amino acid sequences, which is crucial for developing vaccines and understanding antibody responses.	Strengths: - Efficient B-cell epitope prediction: Specifically optimized for linear B-cell epitope mapping, using ANN models. - User-friendly interface: Easy to use for researchers with minimal bioinformatics experience. Limitations: Limited to linear epitopes: Does not predict conformational epitopes, which are essential in some contexts.	(Saha & Raghava, 2006)
BepiPred - 2.0 Updated to BepiPred - 3.0	https://services.healthtech.dtu.dk/services/BepiPred-2.0/ https://services.healthtech.dtu.dk/services/BepiPred-3.0/	BepiPred - 2.0 is a web server for predicting B-cell epitopes, combining machine learning techniques with protein sequence data. It predicts both linear and potential conformational B-cell epitopes, which are essential for antibody response analysis in vaccine design. BepiPred - 3.0 is built on version 2.0, incorporating deep learning for enhanced accuracy in predicting both linear and conformational epitopes. It is widely applied in antibody and vaccine research.	Strengths: - Enhanced prediction accuracy: Uses improved algorithms over the previous version, providing higher accuracy. - Predicts both linear and conformational epitopes: More comprehensive for vaccine research. Limitations: Dependent on sequence data quality: Limited by the quality and completeness of input sequences.	(Larsen et al., 2006) (Clifford et al., 2022)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
VaxiJen v2.0	http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html	VaxiJen v2.0 is an antigenicity prediction server for identifying vaccine candidates. Unlike traditional alignment-based methods, it uses a machine learning approach to distinguish antigens from non-antigens, making it a useful tool in preliminary vaccine design.	Strengths: - Efficient antigenicity prediction: Uses machine learning, allowing for quick and accurate antigen identification. - Alignment-free: Does not rely on sequence alignment, making it faster and suitable for diverse proteins. Limitation: Focuses solely on antigenicity, requiring additional tools for comprehensive vaccine design.	(Irini A. Doytchinova & Darren R. Flower, 2007)
ToxinPred	http://crdd.osdd.net/raghava/toxinpred/	ToxinPred is a web server for predicting the toxicity of peptides, which helps researchers identify toxic and non-toxic peptides in vaccine design and drug development. It uses various machine-learning techniques to evaluate toxicity based on peptide sequence characteristics.	Strengths: - Comprehensive toxicity prediction: Identifies toxic properties in peptides efficiently. - User-friendly interface: Accessible to researchers with varied expertise. Limitation: Focuses on peptide-level toxicity, not applicable for larger proteins directly.	(Gupta, Kapoor, Chaudhary, Gautam, Kumar, Open Source Drug Discovery, et al., 2013)
AllergenFP	https://ddg-pharmfac.net/AllergenFP/index.html	AllergenFP is a server designed to predict allergenicity in proteins, using a fingerprint-based approach to identify allergenic properties. This tool is particularly useful in vaccine design and food safety research to minimize allergic reactions.	Strengths: - Fingerprint-based prediction: Uses molecular fingerprints for reliable allergenicity prediction. - Quick analysis: Provides fast and accessible allergen prediction. Limitation: Focuses only on allergenic properties, requiring other tools for comprehensive protein analysis.	(Dimitrov et al., 2013)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
AllerTop v.2.0	https://www.ddg-pharmfac.net/AllerTOP/index.html	AllerTop v.2.0 is a web server for predicting protein allergenicity using an amino acid composition-based approach. It classifies proteins as allergens or non-allergens, making it valuable in vaccine development and food safety to reduce allergenic risk.	Strengths: - Efficient allergenicity prediction: Uses amino acid composition and auto-cross covariance for accurate allergen classification. - Fast and accessible: User-friendly interface for quick allergenicity assessment. Limitation: Limited to allergen prediction, necessitating complementary tools for broader protein analysis.	(Dimitrov et al., 2014)
IFNepitope	http://crdd.osdd.net/raghava/ifnepitope/	IFNepitope is a web server that predicts interferon-gamma (IFN-γ) inducing epitopes, essential for designing vaccines that promote cell-mediated immunity. It uses machine learning models trained on experimentally validated data to identify potential IFN-γ epitopes.	Strengths: - Targeted epitope prediction: Focuses on IFN-γ inducing epitopes, crucial for cell-mediated immunity in vaccines. - Machine learning models: Enhances prediction accuracy using experimentally validated data. Limitation: Limited to IFN-γ epitopes, so additional tools are needed for broader immune response analysis.	(Dhanda, Vir, et al., 2013b)
IL4Pred	http://crdd.osdd.net/raghava/il4pred/	IL4Pred is a web server that predicts interleukin-4 (IL-4) inducing peptides, which are essential for understanding Th2 immune responses. This tool utilizes machine learning algorithms to identify peptides that may induce IL-4 production, aiding in vaccine design and allergy studies.	Strengths: - Specific to IL-4: Focuses on predicting IL-4 inducers, useful for Th2 immunity analysis in vaccine research. - Machine learning based: Leverages computational models for improved accuracy. Limitation: Limited to IL-4, requiring other tools for comprehensive immune response analysis.	(Dhanda, Gupta, et al., 2013b)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
IL-10Pred	http://crdd.osdd.net/raghava/IL-10pred/	IL-10Pred is a web server that predicts interleukin-10 (IL-10) inducing peptides, facilitating the study of anti-inflammatory responses in vaccine design and immunotherapy. It employs machine learning models to identify peptides that may induce IL-10, important for modulating immune tolerance.	Strengths: - Focus on IL-10: Specifically predicts IL-10 inducers, valuable for anti-inflammatory research. - Machine learning-based: Utilizes trained models for accurate predictions. Limitation: Limited to IL-10, so other immune aspects require additional tools.	(Nagpal et al., 2017a)
Population Coverage-Immune Epitope Database and Analysis Resources (IEDB)	http://tools.iedb.org/population/	The Population Coverage tool from IEDB estimates how well a set of epitopes covers different HLA (human leukocyte antigen) alleles across global populations. This is essential for vaccine design, ensuring broader population effectiveness by addressing allele diversity.	Strengths: - Global HLA coverage: Estimates epitope coverage across diverse populations. - Essential for vaccine development: Helps design vaccines with broader, global applicability. Limitations: - Dependent on HLA data: Results rely on the accuracy and availability of HLA allele data. - Requires complementary analysis: Other immune parameters need separate tools.	(Bui et al., 2006)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
ProtParam	https://web.expasy.org/protparam/	ProtParam is a tool available on the ExPASy platform for analyzing the physicochemical properties of proteins, such as molecular weight, amino acid composition, extinction coefficient, and stability. This is useful in assessing protein characteristics in research and vaccine design.	Strengths: - Detailed physicochemical analysis: Provides in-depth information on properties like stability and molecular weight. - User-friendly: Straightforward interface suitable for quick assessments. Limitations: - Limited to primary structure: Analyzes only the protein's amino acid sequence, not tertiary or quaternary structures. - Requires additional tools: Further tools are needed for comprehensive protein modeling and structural analysis.	(Gasteiger et al., 2005)
PROMOTIF program (PDBsum server)	http://www.ebi.ac.uk/pdbsum	PROMOTIF is a tool integrated within PDBsum for analyzing protein secondary structure motifs. It identifies structural motifs such as β-turns, γ-turns, β-strands, and helical geometries, providing detailed insight into protein structure for research and validation.	Strengths: - Detailed structural analysis: Identifies a wide range of secondary structure motifs, aiding in protein structure validation. - Useful for model validation: Often used to check structure accuracy in structural biology. Limitations: - No standalone access: Available only as part of PDBsum, limiting standalone use. - Focus on secondary structure: Limited to secondary structure analysis, not suitable for primary sequence analysis.	(Laskowski et al., 2018)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
SWISS-MODEL	https://swissmodel.expasy.org/	SWISS-MODEL is a web-based server for homology modeling of protein 3D structures. It uses sequence similarity to known structures in the Protein Data Bank (PDB) to predict structural models, supporting research in structural biology and protein engineering.	Strengths: - Automated homology modeling: Provides high-quality models based on sequence alignment with minimal input. - User-friendly interface: Accessible to both novices and experts in structural biology. Limitations: - Dependent on known templates: Requires similar structures in the PDB for accurate modeling. - Focus on homology modeling: Limited to homology modeling, not suitable for de novo predictions.	(Waterhouse et al., 2018)
I-TASSER	https://zhanggroup.org/I-TASSER/	I-TASSER (Iterative Threading ASSEmbly Refinement) is a web server for predicting 3D protein structures based on sequence-to-structure alignment and iterative structure assembly simulations. It also provides functional insights, such as ligand-binding sites and GO terms.	Strengths: - Comprehensive prediction: Provides both structural and functional predictions, making it versatile for protein studies. - Accurate for distant homologs: Effective when homologous templates are limited, using iterative assembly. Limitations: - Computationally intensive: Processing can take longer for larger proteins. - Dependent on template availability: Performance improves with related structures in the database.	(Yang et al., 2015)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
AlphaFold	https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb	AlphaFold is an AI-driven tool developed by DeepMind for predicting protein 3D structures with high accuracy based solely on amino acid sequences. It uses deep learning to model protein folding, accelerating research in structural biology, drug discovery, and vaccine development.	Strengths: - High prediction accuracy: Outperforms traditional methods, providing near-experimental accuracy for many proteins. - Broad application: Useful in various fields, from structural biology to drug and vaccine design. Limitations: - Focused on static structures: Does not address protein dynamics or interactions. - Limited in multimer modeling: Primarily models single-chain proteins, with limited accuracy for complexes.	(Jumper et al., 2021)
GalaxyRefine	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE	GalaxyRefine is a web server designed for protein structure refinement. It improves initial protein models by refining side chains and backbone structures, enhancing overall structural accuracy, particularly for homology-modeled proteins.	Strengths: - Improves model accuracy: Enhances both local and global structure through iterative refinement, useful for homology models. - User-friendly: Offers easy input for structure refinement. Limitations: - Time-consuming: Computationally intensive, particularly with large proteins. - Dependent on initial model quality: Performance varies with the accuracy of the initial model.	(Heo et al., 2013)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
ProSA-web	https://prosa.services.came.sbg.ac.at/prosa.php	ProSA-web is a tool for evaluating the quality of protein structures by analyzing Z-scores and identifying errors within 3D models. It is widely used for validating models, helping researchers assess whether a structure is within typical quality ranges for native proteins.	Strengths: - Effective quality assessment: Offers reliable Z-score analysis and error detection for model validation. - User-friendly interface: Simple input requirements, making it accessible for quick model checks. Limitations: - Limited to quality scoring: Does not provide structural refinement or optimization. - Requires complementary validation: Best used alongside other validation tools for comprehensive model evaluation.	(Wiederstein & Sippl, 2007)
PROCHECK	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/	PROCHECK is a protein structure validation tool that assesses stereochemical quality by evaluating parameters like bond angles and Ramachandran plots. It helps identify structural irregularities, making it valuable for validating homology-modeled and experimentally derived proteins.	Strengths: - Detailed quality metrics: Provides comprehensive assessments of protein stereochemistry, including bond angles and torsion angles. - Widely accepted: Commonly used for model validation in structural biology. Limitations: - Focus on stereochemistry: Limited to stereochemical parameters, without functional validation. - Manual interpretation: Requires users to interpret Ramachandran plots and other metrics for thorough validation.	(Laskowski et al., 1993)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
ERRAT	https://www.doe-mbi.ucla.edu/errat/	ERRAT is a protein structure validation tool used to identify potential errors by analyzing non-bonded atomic interactions. It generates an overall quality factor, useful for assessing the accuracy of homology-modeled and experimentally determined protein structures.	Strengths: - Error detection: Highlights regions with potential errors based on non-bonded interactions. - Quality factor: Provides an overall quality score, simplifying structural evaluation. Limitations: - Focused on non-bonded interactions: Limited to specific types of structural assessment, so it's best used with other validation tools. - Dependent on initial model quality: Effectiveness varies with initial model accuracy.	(Colovos & Yeates, 1993)
PatchDock	https://www.cs.tau.ac.il//~ppdock/PatchDock/	PatchDock is a molecular docking algorithm that identifies potential binding sites by matching complementary shapes. It is widely used for protein-protein, protein-ligand, and protein-DNA docking, aiding structural biology and drug design.	Strengths: - Shape-based docking: Matches 3D complementary shapes for effective docking predictions. - Versatile applications: Suitable for protein-protein, protein-ligand, and protein-DNA interactions. Limitations: - Focus on rigid body docking: This does not account for flexibility, which may limit accuracy in certain cases. - Requires post-processing: Results often need refinement with additional tools for best accuracy.	(Schneidman-Duhovny et al., 2005)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
ClusPro v2.0	https://cluspro.bu.edu/login.php	ClusPro v2.0 is a protein-protein docking server that generates multiple docking poses, clusters them, and ranks them based on cluster size and interaction energy. It's widely used for predicting protein interactions in structural biology and drug discovery.	Strengths: - Efficient clustering: Clusters and ranks docking poses based on cluster size, improving result accuracy. - Energy-based scoring: Utilizes energy scores for robust ranking. Limitations: - Rigid-body docking: Assumes rigidity in docking, which may affect flexibility-based results. - Requires additional refinement: Results may benefit from further refinement with other tools.	(Kozakov et al., 2017a)
GROMACS software	Linux/Unix operating system	GROMACS (GROningen MAchine for Chemical Simulations) is a versatile molecular dynamics simulation software package designed for biomolecular simulations of proteins, lipids, and nucleic acids. It's widely used in structural biology, drug design, and materials science.	Strengths: - Highly efficient: Optimized for fast simulations, especially on parallel processors. - Versatile and comprehensive: Supports complex biomolecular systems and various force fields. Limitations: - Steep learning curve: Requires expertise in MD simulations. - Computational resources: Simulations can be resource-intensive, especially for large systems.	(Abraham et al., 2015)

Table 1.1 (continued)

Server, Database, Software	Web address	Brief descriptions	Comments	Reference
C-immSim	http://www.cbs.dtu.dk/services/C-ImmSim-10.1/	C-ImmSim is an immune simulation tool that predicts immune responses to vaccine candidates or pathogens by simulating the interactions of immune cells, including T and B cells. This in silico approach aids in evaluating immunogenicity before laboratory testing.	Strengths: - Comprehensive immune simulation: Models complex immune interactions, helpful for early vaccine assessment. - Reduces lab testing: Provides preliminary insights into immune response. Limitations: - Dependent on simulation parameters: Results vary based on input parameters. - Does not replace experimental validation: Simulated predictions need laboratory confirmation.	(Rapin et al., 2011)

1.4. Multi-epitope vaccines: a promising strategy against infectious diseases

Multi-epitope vaccines represent a significant advancement in vaccinology, as they are designed to include multiple epitopes from different antigens to stimulate both humoral and cellular immune responses. By including epitopes that can elicit both T cells (cytotoxic T lymphocytes (CTLs) and helper T lymphocytes (HTLs)) and B cells and induce effective responses against targeted pathogens, these vaccines offer broader protection, more robust immune reactions, and greater adaptability to antigenic variability compared to single-epitope vaccines (Zhang, 2018).

One of the key advantages of multi-epitope vaccines is their ability to incorporate various epitopes from different antigens, which expands the scope of immune recognition of targeted pathogens. For instance, Chauhan et al. highlighted that multi-epitope vaccines can consist of overlapping CTL, HTL, and B-cell epitopes, thereby activating both cellular and humoral immune responses (Chauhan et al., 2019). This multifaceted approach not only enhances the immunogenicity of the vaccine but also reduces the risk of adverse effects associated with unwanted components compared to single-epitope vaccines or whole-pathogen vaccines.

Moreover, the ease of synthesis and stability in various storage conditions are also some of the advantages of multi-epitope vaccines (Zhang, 2018). The synthesis of multi-epitope vaccines is facilitated by their reliance on well-characterized peptides, which can be produced efficiently and cost-effectively. The ability to rationally engineer epitopes allows for the selection of the most immunogenic components, thereby optimizing the vaccine's effectiveness while minimizing production costs (Chauhan et al., 2019; Rashidi et al., 2022). This streamlined production process is crucial, especially when rapid responses are needed in the face of emerging infectious diseases or pandemics.

Additionally, the stability of multi-epitope vaccines is a notable advantage. Unlike traditional live attenuated vaccines or mRNA vaccines, which may require stringent storage conditions to maintain their efficacy (Rando et al., 2023; Uddin & Roni, 2021), multi-epitope vaccines can often be stored at ambient temperatures without significant loss of potency (Zhang, 2018). This stability is particularly beneficial for vaccines intended for use in regions with limited cold chain infrastructure (Kar et al., 2020). The chemical nature of the peptides used in these vaccines contributes to their robustness, allowing them to withstand various environmental conditions without

degrading (Maleki et al., 2022).

Furthermore, multi-epitope vaccines can be designed to include adjuvants that enhance their immunogenicity while maintaining stability. The incorporation of adjuvants not only boosts the immune response but also contributes to the overall shelf-life of the vaccine (Naz et al., 2023). This combination of stability and enhanced immune activation is critical for ensuring long-lasting protection against pathogens.

In summary, multi-epitope vaccines represent a promising strategy in vaccine development, offering a next-generation solution that leverages the power of immunoinformatics and bioinformatics. By incorporating multiple epitopes, these vaccines can elicit stronger and broader immune responses, making them valuable tools in the fight against infectious diseases. Additionally, their ease of synthesis, stability under various storage conditions, and favorable safety profile make them particularly compelling in modern vaccine development. Ongoing advancements in this field are expected to yield innovative vaccine candidates capable of addressing both current and emerging health challenges. For example, recent studies on tuberculosis have focused on designing multi-epitope vaccines, as outlined in Table 1.2.

Table 1.2 Details of previous MEVs design for *Mycobacterium tuberculosis*

Target protein	Selected epitopes	Adjuvants	References
Secreted exosome proteins: DnaK, GrpE, HBHA, LprA, LprG, LpqH, and MPT83	**MHC-I epitopes:** GEVKDVLLL, DAVITTPAY, RTTPSIVAF, ATKDGSHYK, and AVAGAAILV **MHC-II epitopes:** LTVAVAGAAILVAGL, EVKDVLLLDVTPLSL, NVNIAIGGAATGIAA, and IGARDDLMVNNAGLV **B-cell epitopes:** EAHAEEDRKRREEADV, DGTAVADTAENDQADQ, EGEDFDPVLHEAVQHE, LGYTSGTGQGNASATK, PGAASGPKVVIDGKDQ, and TDTRSRVEESRARLTK	RpfE (TLR4 agonist)	(Sharma et al., 2021b)
Ag85A, Mtb32A, Rv2608, and Rv2684	**MHC-I epitopes:** TIATFEMRY, SSPDVLTTY, PTVDYAFQY, STDTPWWAL, and VSIALAAIY **MHC-II epitopes:** PQLGFTLSGATPADA, IFYSHDTGIDWDVIF, KFLEGFVRTSNIKFQ, and TYGVDVVGYDRTQDV **B-cell epitopes:** GITGNGQIGFGKPANP, SRAGLTFNDFMLHLTP, YSDWYQPACGKAGCQT, and PALPLDPSAMVAQVGP	Griselimycin	(Bibi et al., 2021a)
ESAT-6-like protein and diacylglycerol acyltransferase	**B-cell derived T-cell epitopes:** DSGGYNANS, VSRADEEQQ, ADEEQQQAL, and AGVQYSRAD	RS09 (TLR4 agonist)	(Albutti, 2021a)
Rv2031, Rv2346, Rv2347, Rv3614, and Rv3615	**MHC-I, MHC-II, and B-cell epitopes:** DVLAAGDFWGGAGSVACQE, ARRMWASAQNISGAG, RIDHVELSARVAWMSES, HPRSLFPEF, WTADPIIGV, and ELFAAFPSFAGLRPT **MHC-I and MHC-II epitopes:** QTDSAVGSSW, HAMRDMAGR, HTAGVDLAK, and SSLHTAGVDLAKSLRIA	TpD and flagellin	(Ghandadi, 2022)

Table 1.2 (continued)

Target protein	Selected epitopes	Adjuvants	References
Rv1987	**B-cell epitopes:** LGLAITPVASAATARA, GLNIYVRRWRTALHAT, HATVSALIVAILGLAI, APGGSATGGLRGGLTG, APLTDWKLEFDLPAGE, VASAAT, TNSSTAPL, and APGGSATGGLRGGLTGSYSPPSS **MHC-I epitopes:** RATLSVTSTW, ATLSVTSTW, and YVRRWRTAL **MHC-II epitopes:** RTALHATVSALIVAI, RWRTALHATVSALIV, TALHATVSALIVAIL, and WRTALHATVSALIVA	NA	(Arega et al., 2024)
Membrane protein MmpL4	**MHC-I epitopes:** SVLLWQHILAIHLHWL, VVVRWPLPVLV, SIITVVLLLITVGVEL, and MIHAFAVPIILGW **MHC-II epitopes:** AALLGRWFWWPLRVRSRP, NAGLVFAVTMASMAVSDL, and VGTVVVRWPLPVLVA **B-cell epitopes:** FWWPLRVRSRPARTPTVPSE, LTRMHSLMAEMASTTHRMVGDTEEMKE, RPEGTTMDHTSIPFQISMQNAGQLQTIKYQ, YVTGPSALAADMHHSGDRSM, and LEAVGQERSVSLSPKDAPSF	L7/L12 Ribosomal protein (TLR4 agonist)	(Khan et al., 2023b)
Putative nitro-reductases: Rv2032, Rv3127, Rv3129, and Rv3131	**MHC-I epitopes:** AAAGTTANV, SYHAELFWW, AMAAAGTTA, RRRTDRRAY, NTQPWRWRV, WRWRVCPTS, ATLKRHCVR, FRTAEGAKL, TLKRHCVRV, HLRIAMTAA, MYWHLFEPA, and YWHLFEPAL **MHC-II epitopes:** LGWQAKVNR, LAVRAPSIH, WRWRVCPTS, LKRHCVRVI, FRTAEGAKL, WEITGRHFR, VRSYQNRRA, WRWVAESGS, WHLFEPALR, MRPELAAAS, VRVGLAPEM, and LRADAILLR	Phenol-soluble modulin alpha 4 (TLR4 agonist)	(Shiraz et al., 2021)

Table 1.2 (continued)

Target protein	Selected epitopes	Adjuvants	References
Rv0058, Rv0101, and Rv3343	**MHC-I epitopes:** LSDMRSGRM, PTNGQGRVY, LTADLSAAY, VSAPTIINY, FSIPVTFSY, VSESIPLNF, and YSTPALTLF **MHC-II epitopes:** HSDVMYRSVLALLML, LGLTVRYLTPHSKWS, TRIRLVLVSLGVSSF, AALFVLDSWLRPVPA, DQRGASLVVDWPASV, TRILRADTGAEVAFG, GRKEVFRLRLASGRE, and VMRLLSAEAKIKLSD **B-cell epitopes:** MMDIQLHEPTMWKHSP, CAAISAPLRPGSGMPP, and RGDYQGLLGFSSGANV	Griselimycin	(Khan et al., 2022)
Ag85a and low molecular weight T-cell antigen TB8.4	**MHC-I epitopes:** DINTPAFEWY **MHC-II epitopes:** GAAAQFNASPVAQSY, AAQFNASPVAQSYLR, AAAQFNASPVAQSYL, LPVEYLQVPSPSMGR, SAVVGLSMAASSALT, and PVEYLQVPSPSMGRD **B-cell epitopes:** DAVINTTCNYGQVVAA, YSDWYQPACGKAGCQT	50S ribosomal protein L7/L12 and pan-HLA DR-binding epitope	(Peng et al., 2023)
Rv1736c (nitrate reductase), Rv3872 (PE family-related protein), Rv2626c (Hypoxic response protein 1), Rv2031c (heat shock protein), Rv3878 (ESX-1 secretion-associated protein), and Rv2654c (Possible PhiRv2 prophage protein)	**MHC-II epitopes:** QLGETAAEL, MSHDPIAADI, ESHGVAAVLF, LTDRDIVIK, AGLRPTFDTR, and GELFWTVVPY **MHC-II epitopes:** VVLEFAATVDPEAGR and PVVLEFAATVDPEAG **B-cell epitopes:** AEVVAAARDEGAGASP, TAAAQYMREHDIGALP, DEMKEGRYEVRAELPG, and PSVGPDRPEYEPRGCP	Cholera toxin subunit B	(Jiang, Wang, et al., 2023)

Table 1.2 (continued)

Target protein	Selected epitopes	Adjuvants	References
Rv1698 and Rv197	**MHC-I epitopes:** ATDGTTALL, LTGTFLDAY, FSDTLLSSL, and STDAGNQGV **MHC-II epitopes:** VIAYGLLPGLALALT and VARFAAALAPRGSGT **B-cell epitopes:** LSNADPAAPTVEQAQRDTVL, and VNQTITVGKDAPTTAASSVR	Cholera toxin subunit B	(Rahmat Ullah et al., 2021)
PE_PGRS49 and PE_PGRS56	**MHC-II epitopes:** GTGWNGGKGDTG **B-cell, MHC-I, and MHC-II epitopes:** GIGGGTQSATGLG, FAGAGGQGGLGG, and GGAGGNGSLSS	RS09	(Ruaro-Moreno et al., 2023)
PE_PGRS17	**MHC-I epitopes:** GGVGGAGGAAGAVTTI, TGVAVNPVTGEVYVTN, VSLARAGTAGGAGRGP, and TVSVIDPTTNTVTGSP **MHC-II epitopes:** GGNVYVTNFGSGTVS, GEVYVTNFAGDTVSV, RFVLALSQAGSTYAV, FVLALSQAGSTYAVA, PVTGLVFVTNFDSNT, HGQHYQAISAQVAAY, QRFVLALSQAGSTYA, YQQRFVLALSQAGST, and TGLVFVTNFDSNTVS **B-cell epitopes:** AISAQVAAY, VTTITHASF, ALSQAGSTY, AVNPGGNVY, VSTAIAALF, and VIDPTTNTV	Griselimycin	(Moodley et al., 2022)
HtpX, Mce1a, Mce4C, Mce4D, OstA, , Rv1085c	**MHC-I epitopes:** SLAGNSAKV, FAVGMNVYV, and ILLKMCWPA HT epitopes: LVLVFALVVALVYLQ, PNLVAARFIQLTPVY, LVFALVVALVYLQFR, VLVFALVVALVYLQF, LTIGFFLHIPFPPVE, RPDLTIGFFLHIPFP, and DLTIGFFLHIPFPPV	NA	(Madan et al., 2021)

Table 1.2 (continued)

Target protein	Selected epitopes	Adjuvants	References
EspA, EspC, EsxA, EsxB, LppX, LprA, Mpt63 and PPE18	**MHC-I epitopes:** SLHTAGVDLAK, SQFNDTLNV, RVQGDNISVK, HVAVRTTGK, VLDPAAGVTQL, STNIRQAGVQY, TPAARALPL, APAAAAQAV, VWGLTVGSW, NKSEDAKFVY, KSEDAKFVY, GPSPTIVAM, IPGYPVAGQVW, TYIPVVGHAL, and FIIDPTISA **MHC-II epitopes:** DLTYIPVVGHALSAA, RIAAKIYSEADEAWR, TAGVDLAKSLRIAAK, STTKITGTIPASSVK, QGVPFRVQGDNISVK, STNIRQAGVQYSRAD, AVVRFQEAANKQKQE, AGIEAAASAIQGNVT, EQQWNFAGIEAAASA, QNGVRAMSSLGSSLG, AMFGYAAATATATAT, AAQVRVAAAAYETAY, DAKFVYVDGHLYSDL, STGKIYFDVTGPSPT, TTTKKYSEGAAAGTE, and ADGPVGAAAEQVGGQ **B-cell epitopes:** AVDASSGVEAAAGLGE, SLLGITSADVDVRANP, LKTQIDQVESTAGSLQ, GGSGSEAYQGVQQKWD, GARAGGGLSGVLRVPP, LVQIQIAPTKDTSVTL, GGCSTEGDAGKASDTA, GQVWEATATVNAIRGS, EDAERAPVEADAGGGQ, and YSEGAAAGTEDAERAP	Heparin-binding hemagglutinin	(Andongma et al., 2023)

1.5. Applications of multi-epitope vaccines in infectious diseases

The design and evaluation of multi-epitope vaccines presented in this thesis follow a consistent methodology, from epitope prediction to immune simulation. These steps, applied to the Powassan virus, *Mycobacterium tuberculosis*, and respiratory syncytial virus, not only highlight the adaptability of bioinformatics approaches but also reveal insights into the shared principles that govern immune responses across infectious diseases. By exploring the challenges and opportunities in designing multi-epitope vaccines for these diverse targets, this thesis underscores the versatility and potential of computational tools in tackling global health challenges.

Based on this consistent methodology, the thesis is organized into six chapters, each focusing on different aspects of multi-epitope vaccine design and evaluation.

- Chapter 1: General Introduction provides an overview of the global health challenges posed by infectious diseases, the limitations of traditional vaccine development, and the emergence of bioinformatics as a powerful tool for designing multi-epitope vaccines. This chapter sets the context for the studies presented in the following chapters.

- Chapter 2: Methodology Overview describes the bioinformatics and immunoinformatics approaches used throughout the thesis. Key computational tools and techniques, such as epitope prediction, epitope evaluation, molecular modeling, molecular docking, molecular dynamics simulations, and immuno-simulation, are explained in detail.
- Chapter 3 focuses on the design of a multi-epitope vaccine candidate for the Powassan virus, a tick-borne virus that causes severe neurological complications in humans. Its incidence rising due to climate change, thus becoming a growing public health concern. Currently, there are no vaccines or treatments for the Powassan virus. This chapter applies bioinformatics to predict B-cell and T-cell epitopes and evaluates the vaccine's potential to induce strong immune responses.
- Chapter 4: *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis, remains a significant global health threat, particularly in developing countries, despite the availability of the BCG vaccine. This study focuses on designing a multi-epitope vaccine targeting key proteins of MTB. The vaccine construct incorporates epitopes that stimulate both T-cell and

B-cell responses, which are critical for controlling intracellular pathogens like MTB.

- Chapter 5: This chapter presents a promising vaccine against respiratory syncytial virus (RSV), which poses a major health threat, particularly for infants and the elderly. Integrating immunoinformatics and computational approaches, this study analyzes RSV structural proteins across subtypes A and B, identifies T-cell and B-cell epitopes based on antigenicity, allergenicity, toxicity, and cytokine-inducing potential, offers theoretical support, and marks a significant advancement in vaccine development efforts for RSV.
- Chapter 6: General Discussion integrates the findings from the four studies and provides a broader discussion of the potential of multi-epitope vaccines in addressing infectious diseases. This chapter also explores the limitations of computational approaches and proposes future directions for experimental validation and clinical applications.

The research presented in this thesis has the potential to significantly impact both the fields of infectious disease control. The Powassan virus study (Chapter 3) lays the groundwork for epitope-based vaccine design in a viral context, which is further explored in

Mycobacterium tuberculosis (Chapter 4) through its application to a bacterial pathogen. These findings inform the design of vaccines for more complex pathogens, such as RSV (Chapter 5), showcasing the broad utility of the methods developed in this thesis.

The successful application of bioinformatics and immunoinformatics to the design of multi-epitope vaccines demonstrates the power of computational approaches in addressing some of the most pressing global health challenges. Firstly, the ability to rapidly design multi-epitope vaccines for emerging pathogens is critical for pandemic preparedness. The research in this thesis shows that bioinformatics tools can accelerate the identification of immunogenic epitopes and reduce the time needed to develop effective vaccines. This is especially important in the context of rapidly mutating viruses and other pathogens with high antigenic variability. Secondary, the computational framework developed in this thesis is highly flexible and can be adapted to a wide range of pathogens. This adaptability is particularly valuable in addressing emerging and re-emerging diseases that require rapid vaccine development. Finally, the ability to integrate bioinformatics with immunoinformatics allows for the design of vaccines that are not only effective but also safe, reducing the risk of adverse reactions.

In summary, this thesis sets the stage for the use of bioinformatics-driven multi-epitope vaccines in combating infectious diseases. The ability to predict, design, and evaluate vaccine candidates using computational tools represents a major advancement in the field of vaccinology. The research presented here not only addresses current health challenges but also provides a foundation for future vaccine development efforts.

Chapter 2 Methodology Overview

This thesis utilizes bioinformatics approaches, specifically immunoinformatics and computational biology, to design multi-epitope vaccine candidates for infectious diseases. Computational vaccinology is a critical advancement in modern vaccine design, allowing for rapid prediction and evaluation of antigenic epitopes without the need for laboratory-based testing in the initial stages. This chapter outlines the methodologies applied across the studies, and a summary of the steps involved is presented in Figure 2.1.

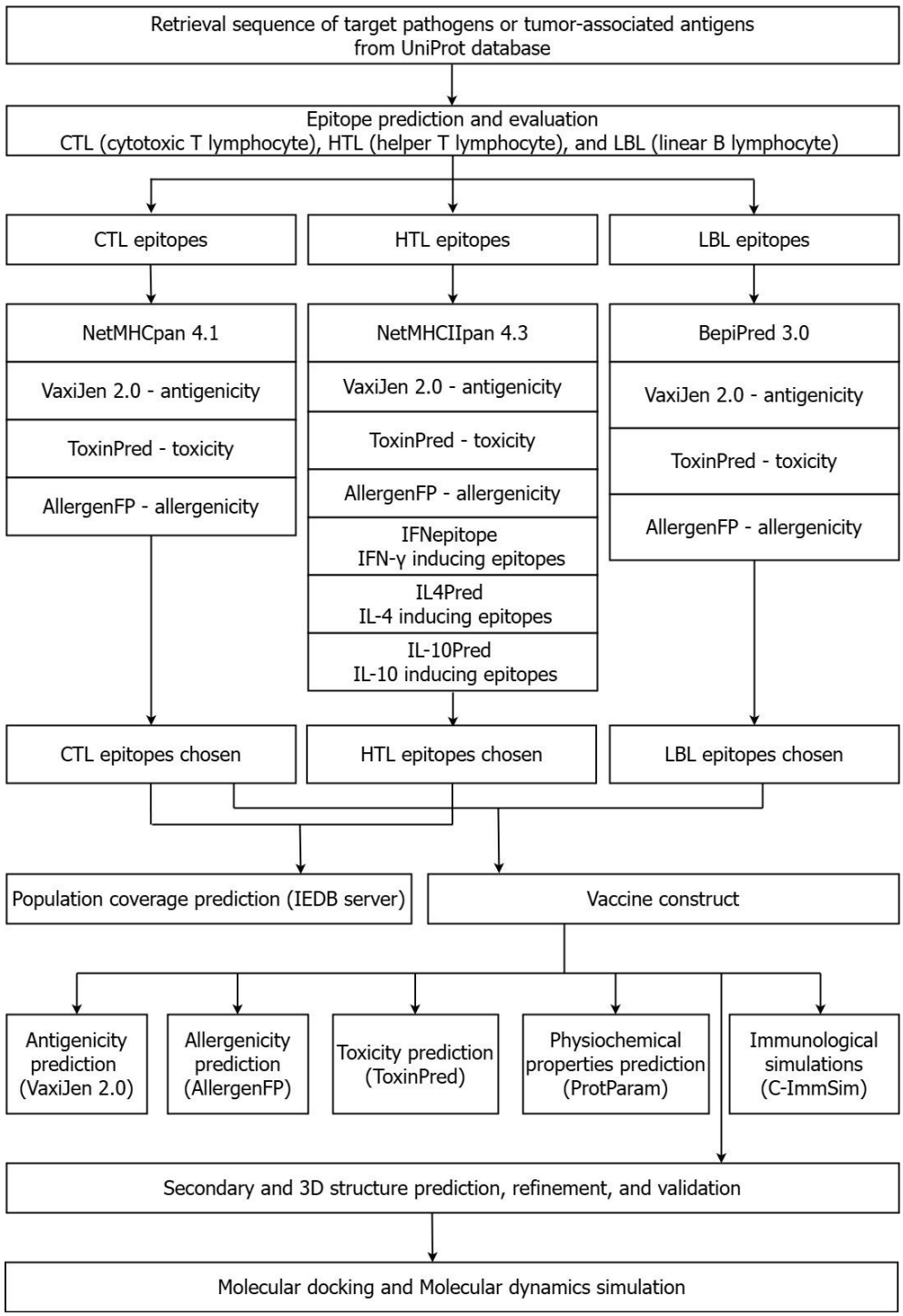


Figure 2.1 Schematic representation of the methodology used in multi-epitope vaccine design and evaluation.

2.1. Sequence retrieval

The protein sequences of target pathogens antigens were obtained from publicly available databases such as UniProt (Consortium, 2023)(Consortium, 2022). For each study, specific antigens that were critical for immune recognition were selected.

2.2. Epitope prediction

Predicted epitopes are critical in multi-epitope vaccine design as they drive the immune response.

Cytotoxic T lymphocyte (CTL) epitopes were predicted using the NetMHCpan 4.1 server, an advanced tool for predicting peptide binding affinities to MHC-I molecules across a wide range of species and alleles. Unlike earlier versions, NetMHCpan 4.1 uses a pan-specific model, which allows it to make accurate predictions even for rare or novel MHC alleles. It leverages deep learning to improve prediction accuracy, making it especially useful in identifying T-cell epitopes for vaccine development and immunotherapy research (Reynisson et al., 2020).

Helper T lymphocyte (HTL) epitopes were identified using the NetMHCIIpan 4.3 server, focusing on epitopes with strong binding affinity to MHC-II molecules. This server uses a pan-specific method

and provides high accuracy in binding prediction. This tool is widely used for epitope discovery in vaccine and immunotherapy research (Nilsson et al., 2023).

Linear B lymphocyte (LBL) epitopes were predicted using BepiPred 3.0. This version leverages deep learning methods for more accurate predictions compared to previous versions, enhancing accuracy in predicting both linear and conformational epitopes. BepiPred-3.0 is widely used in vaccine design, antibody development, and immunodiagnostics, where accurately identifying B-cell epitopes is essential for inducing antibody responses (Clifford et al., 2022).

2.3. Epitope evaluation

To ensure the vaccine candidates are immunogenic and safe, predicted epitopes were evaluated using:

VaxiJen v2.0, an online server for predicting the antigenicity. Unlike many other tools, VaxiJen does not require alignment with known epitopes; instead, it uses a physicochemical property-based approach to distinguish between antigenic (vaccine-relevant) and non-antigenic proteins. This makes it particularly fast and suitable for screening large numbers of proteins to identify potential vaccine candidates. VaxiJen v2.0 supports predictions for several pathogen

types, including viruses, bacteria, and parasites, making it a widely used tool in initial vaccine design (Irini A. Doytchinova & Darren R. Flower, 2007).

AllergenFP to predict allergenic epitopes. It employs a unique fingerprint-based approach, converting protein sequences into binary fingerprints, which are then compared with known allergens to assess potential allergenicity. AllergenFP is particularly valuable in vaccine development, where identifying and avoiding allergenic proteins is critical for safety. Its fast, alignment-free method makes it suitable for screening large datasets, ensuring that selected proteins are less likely to provoke allergic reactions (Dimitrov, Naneva, et al., 2013).

ToxinPred to ensure non-toxic properties of the epitopes. It uses machine learning models to assess various physicochemical properties, such as hydrophobicity and charge, to determine a peptide's toxicity. ToxinPred is particularly useful in vaccine and therapeutic peptide development, as it helps filter out toxic peptides early in the design process, reducing the risk of adverse reactions and ensuring safety in final vaccine formulations (Gupta, Kapoor, Chaudhary, Gautam, Kumar, & Raghava, 2013).

IFNepitope is a computational tool designed to predict and identify interferon-gamma (IFN-γ) inducing epitopes. IFN-γ is a

critical cytokine in immune responses, especially in activating macrophages and promoting antiviral immunity. IFNepitope uses machine learning algorithms to analyze protein sequences and predict which peptides are likely to induce IFN-γ production, making it valuable for designing vaccines and immunotherapies that aim to stimulate strong cellular immune responses (Dhanda, Vir, et al., 2013b).

IL4Pred is a bioinformatics tool used to predict interleukin-4 (IL-4) inducing peptides. IL4Pred employs machine learning techniques to analyze peptide sequences and predict their potential to induce IL-4 production. This tool is useful in vaccine design, especially for vaccines targeting pathogens that require a strong humoral response, as it helps identify epitopes likely to enhance IL-4 mediated immune responses (Dhanda, Gupta, et al., 2013a).

IL-10Pred is a bioinformatics tool designed to predict peptides that can induce interleukin-10 (IL-10) production. Using machine learning algorithms, IL-10Pred analyzes peptide sequences to predict their potential to stimulate IL-10 production. This tool is particularly valuable in designing therapeutic peptides and vaccines aimed at controlling inflammatory responses, autoimmune conditions, and chronic infections, where modulation of the immune response is essential (Nagpal et al., 2017a).

2.4. Population coverage analysis

The IEDB population coverage analysis tool was used to estimate population coverage for the selected T-cell epitopes in the vaccine candidate. In the IEDB tool, default values were used for the number of epitopes and query options, while the "World" option was selected for area/population (Bui et al., 2006b).

2.5. Vaccine construction

The chosen CTL, HTL, and LBL epitopes were assembled into a vaccine construct. The epitopes were linked using suitable linkers such as GPGPG, AAY, KKK, GGS, and EAAAK to preserve immunogenicity and structure as they are cleavable, flexible, and rigid. Additionally, adjuvants like tetanus and diphtheria toxoids (TpD), the final subunit of *Escherichia coli* type 1 fimbria (FimH), RS09, flagellin protein, the PADRE (Pan HLA DR-binding epitope) sequence, etc., were incorporated to enhance the immune response. Previous studies have demonstrated the effectiveness of TpD in protecting mucosal surfaces and promoting neutralizing antibody production (Chan et al., 2020; Li et al., 2018), while cholera toxin subunit B (Hou et al., 2014) and *E. coli* FimH have shown potential to enhance immunogenicity and stimulate cytokine production (Zhang et al., 2020; Zhang et al., 2022).

RS09 functions as an agonist of TLR4, while the flagellin protein acts as a TLR5 agonist (Forstnerič et al., 2017; Gupta et al., 2014; Shanmugam et al., 2012). The TLR agonists play a crucial role in activating innate and adaptive immunity. Incorporating PADRE into the vaccine can elevate its immunogenicity and efficacy (Ma et al., 2020). PADRE exhibits a high-affinity binding ability to various MHC class II molecules, facilitating the generation of antigen-specific CD4+ T-cell responses (Ghaffari-Nazari et al., 2015). Additionally, it has also been reported to elicit CD8+ T-cell responses (Ma et al., 2020).

2.6. Vaccine's physiochemical properties

The ProtParam web tool was used to evaluate the physicochemical properties of the vaccine. This tool is provided by the ExPASy (Expert Protein Analysis System) platform which analyzes various physicochemical properties of proteins based on their amino acid sequences. It calculates key parameters such as molecular weight, theoretical isoelectric point (pI), amino acid composition, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). These properties are essential for understanding protein stability, solubility, and overall behavior of vaccine design (Gasteiger et al., 2005).

2.7. Immune simulations

In silico immune simulations were performed using C-ImmSim, an online server designed to simulate the immune response to antigens and vaccine candidates. It uses a computational approach to model both humoral and cellular immune responses over time. C-ImmSim can simulate processes like the generation of antibodies, activation of T-cells, and cytokine responses, providing valuable insights into potential vaccine efficacy before experimental testing (Rapin et al., 2012).

2.8. Vaccine's secondary and tertiary structure prediction

The secondary structure of the vaccine was predicted using Gail Hutchinson's PROMOTIF v.3.0 program via the PDBsum server. PROMOTIF analyzes protein coordinate files to identify and detail structural motifs within a protein, providing information on secondary structures, β- and γ-turns, helical geometry, β-strand topology, β-hairpins, and more (Laskowski et al., 2018).

The tertiary structure of the vaccine construct was modeled using several servers, such as SWISS-MODEL, Iterative Threading ASSEmbly Refinement (I-TASSER), and AlphaFold via AlphaFold

Colab.

SWISS-MODEL is a web-based tool for homology modeling of protein 3D structures. It was used to generate structural models based on a known template, which is identified from proteins with similar sequences. It's widely used in structural biology, drug discovery, and vaccine design for visualizing protein functions, interactions, and ligand binding sites, aiding in research where experimental structures are unavailable (Waterhouse et al., 2018).

I-TASSER is a computational tool for protein structure prediction and function annotation. It uses a combination of threading (identifying structural templates), iterative fragment assembly, and atomic-level refinement to predict 3D structures of proteins from their amino acid sequences. Widely used in structural biology and bioinformatics, I-TASSER is valuable for research in drug discovery, vaccine design, and protein engineering, especially when experimental structures are unavailable (Yang et al., 2015).

AlphaFold is an AI-driven tool developed by DeepMind that predicts protein 3D structures with remarkable accuracy based solely on amino acid sequences. Leveraging advanced deep learning models trained on extensive experimental protein data, AlphaFold produces structural predictions comparable to those obtained by experimental

methods like X-ray crystallography or cryo-electron microscopy. This breakthrough has significantly accelerated research across structural biology, drug discovery, and vaccine development by offering rapid, accessible insights into protein folding, function, and interactions (Jumper et al., 2021).

While SWISS-MODEL, I-TASSER, and AlphaFold are all powerful tools for protein structure prediction, they each have distinct strengths. SWISS-MODEL is best for accessible, routine homology modeling when suitable templates are available. I-TASSER offers flexibility for proteins with limited homologs and provides additional functional insights, such as active site predictions. AlphaFold delivers unmatched accuracy, making it ideal for complex or novel proteins where high-quality structure predictions are essential.

2.9. Three-dimensional structure refinement and validation

To assess the structural quality and stability of the vaccine's 3D model, the following tools were employed GalaxyRefine, ProSA-web, PROCHECK, and ERRAT.

GalaxyRefine is a computational tool used for protein structure refinement, primarily improving the quality of initial protein models

generated by homology modeling methods. Developed as part of the GalaxyWeb server suite, GalaxyRefine applies iterative structural relaxation through molecular dynamics simulations to optimize side-chain positioning and correct minor backbone errors. This results in more accurate models with enhanced structural quality, particularly in terms of stereochemistry and overall geometry (Heo et al., 2013).

ProSA-web, an online tool used for evaluating the quality of protein structures, is typically generated through computational modeling or experimental methods like X-ray crystallography. ProSA-web analyzes structural models by assessing the overall quality based on Z-scores, which indicate how closely the model aligns with high-quality protein structures in the Protein Data Bank (PDB) (Wiederstein & Sippl, 2007).

PROCHECK is a widely used computational tool for assessing the stereochemical quality of protein structures. It generates detailed reports on structural features such as bond angles, bond lengths, and dihedral angles. By comparing these parameters to those in high-quality structures, PROCHECK identifies any structural anomalies or errors, helping refine protein models for better accuracy (Laskowski et al., 1993).

ERRAT is a structural analysis tool that evaluates the quality of

protein models by identifying potential errors in the non-bonded interactions between atoms. It examines the atomic interactions within a protein structure and calculates an “error value” to highlight regions where the geometry deviates significantly from known high-quality structures. ERRAT is especially useful for validating protein models derived from X-ray crystallography or homology modeling (Colovos & Yeates, 1993). Often used alongside other tools like PROCHECK and ProSA, ERRAT supports the improvement of model accuracy before further analysis or application in research fields such as drug design, vaccine design, and protein engineering.

2.10 Molecular docking and molecular dynamics simulation

To assess the interaction between the vaccine and immune receptors, molecular docking was performed using PatchDock or ClusPro v2.0. The stability of the interactions was confirmed through molecular dynamics simulations using GROMACS, which analyzed key metrics such as root mean square deviation (RMSD) of backbone residues, root mean square fluctuation (RMSF) of C-alpha, radius of gyration (Rg), and solvent accessible surface area (SASA).

PatchDock is a molecular docking algorithm used for predicting

protein-protein, protein-small molecule, and protein-DNA/RNA interactions. PatchDock identifies and aligns geometrically complementary shapes, allowing it to predict how molecules may interact based on surface compatibility. After initial docking, PatchDock refines the results using scoring functions to estimate the binding energy and rank possible interactions (Schneidman-Duhovny et al., 2005).

ClusPro v2.0 is a widely used online server for protein-protein docking, designed to predict the most likely 3D structures of protein complexes. It operates by generating multiple docking models and then clustering them to identify stable conformations. ClusPro v2.0 offers different modes based on the nature of the interaction, such as balanced, electrostatic-favored, hydrophobic-favored, and van der Waals interactions, which allows to tailor the docking to the specific characteristics of the proteins involved (Kozakov et al., 2017b).

ClusPro offers greater accuracy and reliability, especially for protein-protein docking, through its clustering-based approach, making it suitable for in-depth studies. PatchDock, on the other hand, is fast and efficient for quick docking predictions based on shape complementarity, which is helpful in preliminary analyses or for users needing flexible applications like protein-ligand or protein-nucleic

acid docking.

GROMACS is a high-performance, open-source software suite for molecular dynamics (MD) simulations, primarily used to study biomolecular systems like proteins and lipids. It excels in simulating atomic interactions over time, supporting a variety of force fields (e.g., CHARMM, AMBER), and offering numerous tools for energy minimization and data analysis. GROMACS is widely used in fields like structural biology, drug discovery, vaccine design, and materials science to study protein folding, molecular interactions, ligand binding, and membrane dynamics, providing insights that are critical for designing new drugs/vaccines, materials, and therapeutic strategies (Abraham et al., 2015).

This chapter was published in *Scientific Reports*

as a partial fulfillment of Nguyen Thi Truc Ly's Ph.D program.

Chapter 3 Immunoinformatics and computational approaches driven designing a novel vaccine candidate against Powassan virus

3.1. Abstract

Powassan virus is an arthropod-borne virus (arbovirus) capable of causing severe illness in humans for severe neurological complications, and its incidence has been on the rise in recent years due to climate change, posing a growing public health concern. Currently, no vaccines to prevent or medicines to treat Powassan virus disease, emphasizing the urgent need for effective countermeasures. In this study, I utilize bioinformatics approaches to target proteins of Powassan virus, including the capsid, envelope, and membrane proteins, to predict diverse B-cell and T-cell epitopes. These epitopes underwent screening for critical properties such as antigenicity, allergenicity, toxicity, and cytokine induction potential. Eight selected epitopes were then conjugated with adjuvants using various linkers, resulting in designing of a potentially stable and immunogenic vaccine candidate against Powassan virus. Moreover, molecular docking, molecular dynamics simulations, and immune simulations revealed a stable interaction pattern with the immune receptor, suggesting the vaccine's potential to induce robust immune responses. In conclusion, my study provided a set of derived epitopes from Powassan virus's proteins, demonstrating the potential for a novel vaccine candidate against Powassan virus. Further in vitro and in vivo studies are warranted to

advance our efforts and move closer to the goal of combatting Powassan virus and related arbovirus infections.

3.2. Introduction

Emerging infectious diseases have posed significant challenges to public health worldwide in recent years. Among these, the Powassan virus (POWV) is a compelling subject of investigation due to its potential to cause severe neurological illness in humans. POWV, a member of the Flaviviridae family, belongs to a group of vector-borne viruses that include other lethal pathogens such as Zika, dengue, and West Nile virus (Kemenesi & Bányai, 2018). The Powassan genome is composed of about 11 kb of single-stranded, positive-sense RNA that encodes three structural proteins: the capsid (C) protein, the premembrane (prM) protein, and the envelope (E) glycoprotein at the 5′ end of the genome and seven nonstructural proteins in the 3' end of the genome (Hermance & Thangamani, 2017). Geographically, the virus is predominantly found in North America, with a higher incidence in the eastern parts of Canada, Northeastern and Great Lakes regions of the United States, and in far eastern Russia (Corrin et al., 2018; Hermance & Thangamani, 2017). There are two lineages of POWV: lineage 1, which includes POWV strain LB, and lineage 2, also called

deer tick virus (Krow-Lucal et al., 2018). The lineage 1 virus is primarily transmitted to humans through the bite of infected ticks, specifically the Ixodes scapularis tick (also known as the black-legged tick) in North America. Meanwhile, the lineage 2 virus, such as Ixodes cookei mostly transmits to humans in the far eastern region of Russia (Stone et al., 2022).

The virus can lead to various clinical presentations, from asymptomatic cases to severe neuroinvasive disease (Della-Giustina et al., 2021). The time between being bitten by an infected tick and the onset of symptoms (incubation period) can vary, but it's generally within 1–5 weeks (Corrin et al., 2018; Hermance & Thangamani, 2017). Many individuals who are infected with POWV may not show any symptoms. Those who develop symptoms can range from mild to severe, including fever, headache, vomiting, weakness, and confusion. In some cases, neurological symptoms such as encephalitis (inflammation of the brain) and meningitis (inflammation of the membranes surrounding the brain and spinal cord) can occur (Telford & Piantadosi, 2023). Given the potentially rapid onset of symptoms and the lack of specific antiviral therapies, understanding the factors contributing to disease severity and developing novel strategies to prevent and treat POWV is paramount. Preventing POWV infections

hinges on effective tick bite prevention strategies, including insect repellents, protective clothing, and thorough tick checks (Bogaty & Drebot, 2018). As for treatment, there is no specific antiviral therapy or vaccine for Powassan virus infections. Supportive care is the cornerstone of managing cases, and hospitalization may be required for severe presentations (Khan et al., 2019).

In response to the urgent need for effective countermeasures against POWV infections, this study takes an innovative immunoinformatics approach to design a potential vaccine candidate. Leveraging the power of computational tools and bioinformatics analyses, I aim to harness the immune system's capabilities by targeting the structural proteins of POWV. These proteins, including the capsid, envelope, and membrane proteins, play pivotal roles in virus entry, replication, and host interactions, making them attractive targets for vaccine design. Initial steps encompass the retrieval and analysis of structural protein sequences of POWV to identify B-cell epitopes and major histocompatibility complex class I (MHC-I) and MHC-II binding peptides. By employing algorithms that predict binding affinity to MHC molecules, I prioritize epitopes likely to be presented to T cells, a key step in generating cellular immunity. Then, the antigenicity, allergic potential, toxicity, and cytokine inducing potential

of the epitopes were predicted to ensure that our vaccine candidate triggers specific immune responses without causing adverse effects. I aim to narrow the list of potential epitopes exhibiting optimal immunological properties through these stringent screening processes. Subsequently, I design a vaccine candidate using the epitopes that satisfy the selection criteria by joining them to adjuvants using suitable linkers. Finally, I employ molecular modeling techniques to generate the vaccine candidate's three-dimensional structure, which is then used to perform molecular docking and molecular dynamics simulation analyses with toll-like receptor (TLR) molecules.

3.3. Materials and Methods

3.3.1. Retrieval and analysis of protein sequences

The polyprotein sequence of the Powassan virus was obtained from UniProt with the accession number Q04538. Since this research specifically focuses on the structural proteins, their sequence details were extracted from the PTM/Processing section within the UniProt entry. Additionally, the antigenicity and allergenicity of the structural proteins—namely, capsid, pre-membrane protein/membrane (prM/M), and envelope proteins—were evaluated using Vaxijen 2.0 and the

allergenFP tool, respectively (Dimitrov et al., 2013; Irini A. Doytchinova & Darren R. Flower, 2007).

3.3.2. Prediction and evaluation of B-cell and T-cell epitopes

The NetMHC 2.3 website predicted T-helper cell epitopes that could bind to MHC class II molecules (Jensen et al., 2018). Additionally, the NetMHCpan 4.0 website was utilized to forecast T-cytotoxic cell epitopes that could potentially bind to MHC class I molecules (Jurtz et al., 2017). The protein sequences were input in FASTA format, with selected peptide lengths of 9 and 15 (default), on the NetMHCpan 4.0 and NetMHC 2.3 web servers, respectively. The default parameters of the NetMHCpan 4.0 and NetMHC 2.3 web servers were applied to establish the thresholds for strong and weak binders. For the prediction of multiple linear B-cell epitopes, the IEDB B-cell epitope prediction website, employing the BepiPred 3.0 linear epitope prediction approach, was operated (Larsen et al., 2006). The Vaxijen version 2.0 web server was employed to determine the antigenicity of the B-cell and T-cell epitopes (Irini A. Doytchinova & Darren R. Flower, 2007). The allergenFP, ToxinPred, and IFNepitope web servers were used to evaluate the epitopes' allergic potential, toxicity,

and interferon-γ activation potential, respectively (Dhanda, Vir, et al., 2013b; Dimitrov et al., 2013; Gupta, Kapoor, Chaudhary, Gautam, Kumar, Open Source Drug Discovery, et al., 2013). The IL4Pred and IL-10Pred servers were utilized to assess the capability of the epitopes to induce interleukin-4 and interleukin-10 generation (Dhanda, Gupta, et al., 2013b; Nagpal et al., 2017a).

3.3.3. Formulation of the vaccine candidate and determination of the properties

The final epitopes, B-cell and T-cell, were joined to TpD adjuvant and *Escherichia coli* FimH protein with suitable linkers to formulate the final vaccine candidate. TpD is a universal adjuvant for $CD4^+$ T cells, possibly outperforming "PADRE", a peptide that binds to several HLA-DR (Human Leukocyte Antigen-DR) molecules promiscuously (Alexander et al., 1994). TpD has been shown in vitro to protect mucosal membranes, as well as to stimulate the generation of neutralizing antibodies (Chawla et al., 2023). These results are consistent in many mammalian species and are characterized by the development of durable $CD4^+$ T memory cells, the production of neutralizing antibodies, and the release of cytokines such as TNF-α

and interferon-gamma (IFN-γ). A T-helper cell (Th1)-dominant immunological response is indicated by this pattern (Fraser et al., 2014). To further enhance the immune stimulation of the vaccine, I included the final subunit of the type 1 fimbria of *Escherichia coli* (FimH), which has been shown to interact in a dependent way with TLR4. In comparison to lipopolysaccharide (LPS), this interaction regulates MHC class I and class II molecules more advantageously and securely and fosters the maturation, activation, and proliferation of peripheral and dendritic cells in the local area. Like TpD, FimH also promotes IFN-γ and TNF-α production (Zhang et al., 2020; Zhang et al., 2022). Interestingly, FimH's application is effective in promoting mucosal immunity (Zhang et al., 2020). To form rigid or flexible protein configurations based on the desired biological activity, these adjuvants were linked by EAAAK linker; MHC-I and MHC-II epitopes were linked by KKK linker; and B-cell epitopes were linked by EAAAK linker as well. The physiochemical properties were analyzed by Expasy ProtParam webserver, whereas the antigenicity and allergic potential of the formulated vaccine candidate were determined by Vaxijen 2.0 and AllergenFP webservers (Dimitrov et al., 2013; Irini A. Doytchinova & Darren R. Flower, 2007; Gasteiger et al., 2005). The presence of a transmembrane helix in the vaccine candidate was

predicted by DeepTMHMM webserver (Hallgren et al., 2022). Furthermore, signal peptide in the vaccine construct was predicted by SignalP 4.1 tool (Nielsen, 2017). Finally, BLASTp analysis is performed to identify the homologous protein of the vaccine candidate in humans (taxon id: 9606; *Homo sapiens*) (Johnson et al., 2008).

3.3.4. In-silico immune simulation of the vaccine candidate

To assess the immune response induced by the formulated vaccine construct, I conducted an in-silico immune simulation using the C-ImmSim software (Rapin et al., 2011). I employed the default software settings except for the time step. Generally, it is recommended to maintain a minimum interval of four weeks between consecutive vaccine doses, although in certain situations, a longer gap may also be considered (Castiglione et al., 2012; Robinson et al., 2017). Thus, I evaluated the immune response profile for the vaccine construct by administering three vaccine doses at four-week intervals. Time steps of 1, 84 (approximately 4 weeks), and 168 (approximately 8 weeks) were utilized in the simulation.

3.3.5. Modeling, refinement, and validation tertiary structure of the vaccine candidate

The three-dimensional (3D) structure of the MEV construct was modeled using the AlphaFold (David et al., 2022; Jumper et al., 2021). Subsequently, to enhance the model's quality, the generated PDB file of the final vaccine from AlphaFold was submitted to the GalaxyRefine server for protein structure refinement (Heo et al., 2013). This server initiates the reconstruction of all side chain structures, followed by iterative relaxation of the structure through brief molecular dynamics (MD) simulations after side chain repackage perturbations. Evaluation metrics provided by the GalaxyRefine server encompass global distance test-high accuracy (GDT-HA), root-mean-square deviation (RMSD), MolProbity (indicating crystallographic resolution), and Ramachandran favored score. RMSD measures the distance between atoms, with lower values indicating greater stability. An RMSD score falling from 0 to 1.2 is typically considered acceptable (Heo et al., 2013). To supplement refinement, I use the ERRAT program to verify protein structures obtained through crystallography. It assesses errors based on non-bonded atom–atom interactions compared to a database of reliable high-resolution structures (Colovos & Yeates). Additionally, the PROCHECK and ProSA-web servers were employed to assess the

validity and quality of the selected 3D structure (Laskowski et al., 1993; Wiederstein & Sippl, 2007). The Ramachandran plot analysis was conducted using the PROCHECK section of the UCLA-DOE LAB server. This plot illustrates the statistical distribution of backbone dihedral angles ϕ and Ψ, along with the percentage and count of residues in the most favored, additionally allowed, generously allowed, and disallowed regions, thereby delineating the modeled structure's quality (Laskowski et al., 1993). The ProSA-web server was also utilized to identify potential errors in the final vaccine structure (Wiederstein & Sippl, 2007).

3.3.6. Molecular docking and molecular dynamics simulation studies

For the docking of the vaccine with TLR4, we utilized the PatchDock webserver, adhering to default settings (Schneidman-Duhovny et al., 2005). It is a protein-protein docking algorithm and server that is used to predict the three-dimensional structure of protein complexes. The PatchDock algorithm employs a shape complementarity approach, where it matches complementary surface patches on the interacting proteins to generate potential docking poses (Schneidman-Duhovny et

al., 2005). By submitting two protein structures (vaccine and TLR4) to the PatchDock server, I can obtain predictions for their possible docking orientations and interaction interfaces. From this server, I got 10 models, which were ranked based on geometric shape complementarity and energy scoring.

Afterward, to evaluate the stability of the docked complex, molecular dynamics (MD) simulations were conducted using the GROMACS 2023 software on a Linux operating system (Abraham et al., 2015). Three replicas throughout 100 ns of the MD simulation were performed for all-atom to ensure the reliability of the data. The CHARMM27 all-atom force field was used to model the parameters of the proteins. There are 16,182 atoms and 1037 residues in total. The TLR4 protein complexed with the vaccine was placed in a cubic box $9 \times 9 \times 9$ and solvated with SOL water molecules. The simulation box contained 380,145 atoms in the system. Charge neutralization was achieved by adding 2 Cl^- ions. System energy minimization was performed using the steepest descent method, applying a position restraint of 1000 kJ/mol nm^2 on the heavy atoms of the protein. The equilibration process was carried out in a phased manner. Initially, a 1 ns NVT simulation was performed, followed by a 1 ns NPT simulation with restraints on the heavy atoms of the protein. Subsequently, a 10

ns equilibration without restraints on atoms was conducted using the NPT ensemble. The structure obtained after 10 ns of MD simulation was utilized as the starting structure for further equilibration and production simulations. Production simulations were run for 100 ns using the NPT ensemble, maintaining a temperature of 300 K with velocity rescaling and a coupling time of 0.1 ps. The pressure was maintained at 1 atm using a Parrinello-Rahman barostat with a coupling time of 2 ps. After the efficient completion of 100 ns MD simulations, calculations were performed for the RMSD of backbone residues, root mean square fluctuation (RMSF) of C-alpha, radius of gyration (Rg), and solvent accessible surface area (SASA). Besides, I conducted superimpositions of the complex structures extracted from selected snapshots during the simulation using GROMACS tools. Furthermore, to explore the interface connecting TLR4 and the vaccine construct, specific snapshots were analyzed using the COCOMAPS tool (Vangone et al., 2011).

3.4. Results

3.4.1. Retrieval and analysis of protein sequences

The 3415 amino acid long polyprotein sequence reviewed has protein

level evidence and an annotation score of 5/5. In the polyprotein sequence, the capsid protein ranges between 1 and 94 amino acids, prM/M protein ranges between 116 and 278 amino acids, and envelope protein ranges between 279 and 775 amino acids. All the structural proteins were predicted as antigenic by the Vaxijen 2.0 tool. Similarly, all the structural protein except capsid protein was determined as non-allergic by the AllergenFP webserver.

3.4.2. Prediction and evaluation of B-cell and T-cell epitopes

A total of 67 strong binding MHC-I epitopes were identified, including 12 epitopes from capsid protein, 17 from prM/M protein, and 38 from envelope protein. Similarly, 40 strong binding MHC-II epitopes were predicted from the capsid protein, 14 from the prM/M protein, and 52 from the envelope protein, taking the total of strong binding MHC-II epitopes to 106. Finally, from the structural protein, 19 B-cell epitopes were predicted. All these epitopes were then evaluated for their antigenicity, IL-4, and Interferon-gamma inducing ability, toxicity, and allergenicity (Supplementary Datasets). After all the analysis, 8 epitopes were predicted: 2 from MHC-I, 4 from MHC-II, and 2 from B-cell epitopes that were antigenic, non-toxic, non-allergic, and could

activate IL-4 and Interferon-gamma (Table 3.1).

Table 3.1 Final epitopes selected for the vaccine design

Type of epitope	Protein	Epitope
MHC-I	Envelope	DTVVMEVSY VEFGPPHAV
MHC-II	Capsid	MFWKTVPLRQAEAVL MFWKTVPLRQAESAL
	Envelope	GSTIGRMFEKTRRGL GSTIGRMFEKTRKGL
B-cell	Envelope	KHKDNQDWNS EFGPPHAV

3.4.3. Formulation of the vaccine candidate and determination of the properties

Eight chosen epitopes highly suitable for vaccine design were connected with adjuvants via stable linkers. As a result, the final vaccine candidate has 442 amino acids (Figure 3.1).

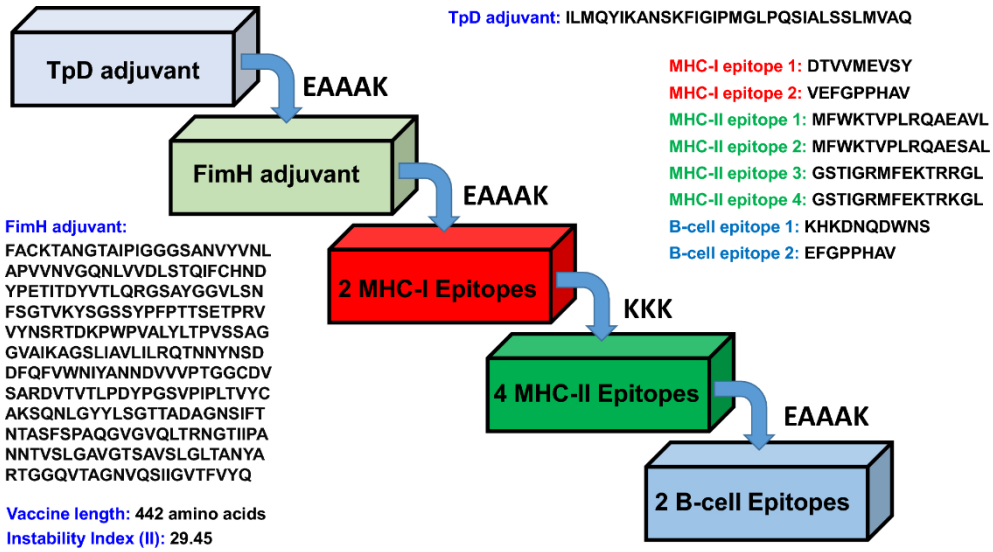


Figure 3.1 Schematic representation of the vaccine design with linkers, adjuvant, and epitopes sequentially and appropriately.

Furthermore, the vaccine candidate was predicted as antigenic by the Vaxijen 2.0 webserver and non-allergic by the AllergenFP webserver. The negative GRAVY score indicates that the vaccine candidate is hydrophilic, and the instability index of 29.45 below 40 shows that the vaccine candidate is stable (Table 3.2). Also, the final vaccine construct has no transmembrane helices or signal peptides. Remarkably, in the BLASTp analysis, the vaccine candidate appears distinct from human proteins, suggesting that it has the potential to be used safely in humans without triggering an autoimmune response in the host.

Table 3.2 Physiochemical properties of the formulated vaccine candidate

Features	Assessment	Remark
Molecular weight	47250.12	Average
Number of amino acids	442	-
Signal peptide	0	-
Transmembrane helix	0	-
Total number of atoms	6685	-
Extinction coefficient	54570	Average
Theoretical pI	9.65	Slightly basic
Aliphatic index (AI)	82.29	Thermostable
Instability Index (II)	29.45	Stable
Estimated half-life (mammalian reticulocytes, in vitro)	20 hours	Satisfactory
Estimated half-life (yeast cells, in vivo)	30 minutes	Satisfactory
Estimated half-life (*Escherichia coli*, in vivo)	>10 hours	Satisfactory
Grand Average of hydropathicity (GRAVY)	-0.099	Hydrophilic
Antigenicity	0.6346 (VaxiJen v2.0)	Antigenic
Allergenicity	Non-allergen (AllergenFP)	Non-allergen

3.4.4. In-silico immune simulation of the vaccine candidate

Figure 3.2 visually represents the anticipated immune response pattern resulting from the formulated vaccine, as determined through computational analysis. Successive administrations of the vaccine trigger an expansion in both the total B-cell population and B-memory cell populations, highlighting the stimulation of a robust secondary immune response (Figure 3.2a). Following each vaccination the total T-helper (TH) cell population and T-helper memory cells increased (Figure 3.2b). On the other hand, the T-cytotoxic (TC) memory cell

population rises after the first vaccination but notably decreases after the second and third doses (Figure 3.2c). Besides, the proliferation of natural killers (NK) was also observed, which is an essential mediator of T cell activation (Figure 3.2d). Furthermore when examining the impact of the initial dose (Figure 3.2e) in comparison to the subsequent second and third doses, it becomes evident that there is an increase in the concentrations of various antibodies, including IgM + IgG, IgM, IgG1 + IgG2, IgG1, and IgG2, indicating that immunization with the candidate vaccine leads to an augmented antibody response. Additionally, the Powassan vaccination using the formulated vaccine candidate can induce the generation of various cytokines, including IFN-γ, interleukin-10 (IL10), interleukin-12 (IL12), and transforming growth factor-beta (TGF-β) (Figure 3.2f). In comparison to the initial dose, the second dose of the vaccine results in increased populations of IFN-γ, IL-10 and IL-12 but showed a decrease in the population of TNF-α. After receiving the third dose of the vaccine construct, there is an overall decrease in the concentration of different cytokines and interleukins compared to the first and second doses.

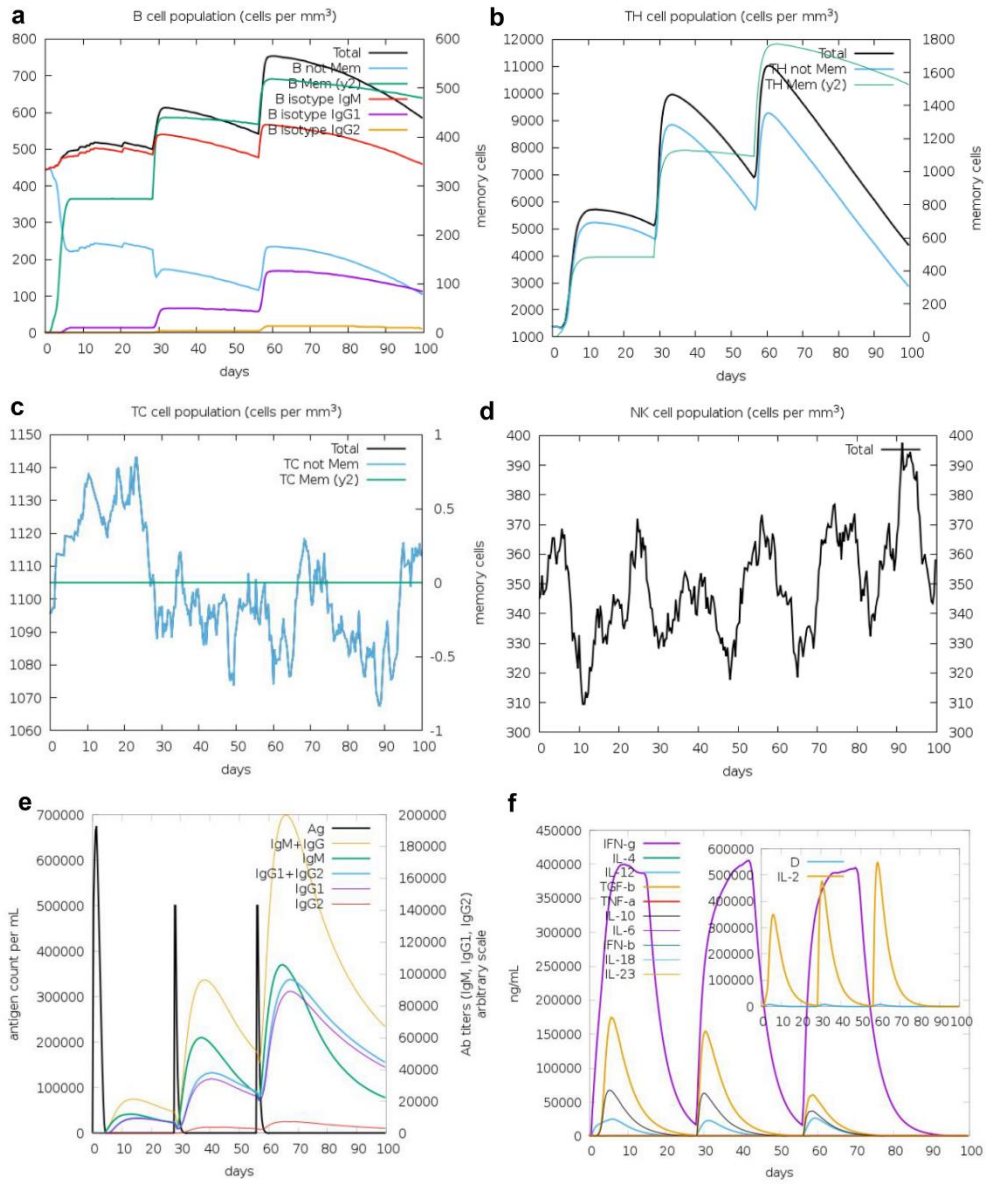


Figure 3.2 Powassan virus vaccine immune-simulation. (a) B-cell population. (b) T-helper cell population. (c) T-cytotoxic cell population. (d) Nature killer cell population. (e) Antigen count and antibody titer with specific subclass. (f) Concentration of cytokine and interleukins. Inset plot shows danger signal together with leukocyte growth factor IL-2.

3.4.5. Modeling, refinement, and validation tertiary structure of the vaccine candidate

The vaccine construct was modeled using the cutting-edge AlphaFold program (David et al., 2022; Jumper et al., 2021). Given that the AlphaFold prediction yielded a local distance difference test score (pLDDT) below 70 (Figure S1) and the Ramachandran plot analysis revealed only 79.8% of residues in most favored regions (Figure S2), I conducted further structural refinement employing the GalaxyRefine server (Heo et al., 2013). Refer to Table 3.3 for the refined model's assessment, Model 2 demonstrated the highest global distance test-high accuracy (GDT-HA) score of 0.9632 (the higher value, the more accurate), and the lowest root mean square deviation (RMSD) score of 0.397 (lower value indicating greater stability).

Table 3.3 Structure Information obtained from GalaxyWEB and ERRAT program

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored	ERRAT value
Model 1	0.9581	0.448	1.518	8.5	0.0	97.7	94.0789
Model 2	0.9632	0.397	1.498	8.0	0.3	97.7	95.7790
Model 3	0.9587	0.444	1.524	8.6	0.0	97.7	92.0630
Model 4	0.9570	0.460	1.488	8.9	0.6	98.0	92.6984
Model 5	0.9570	0.477	1.531	8.8	0.3	97.7	94.1558

Besides, among the 5 models, only Model 2 had an ERRAT value

greater than 95% (Table 3.3 and Figure 3.3a). According to the ERRAT program, good high-resolution structures generally produce values around 95% or higher. Furthermore, to validate the model's quality, I analyzed the Ramachandran plot and calculated a Z-score. The Ramachandran plot analysis revealed that 94.7% of residues were situated in the favorable core region, with 5.3% in the allowed and generously allowed regions, and 0.0% in the disallowed region (Figure 3.3b). Meanwhile, the Z-score had a value of -7.02 using the ProSA webserver (Figure 3.3c). Hence, I can infer that the tertiary structure of refined Model 2 (Figure 3.3d) exhibits a high level of quality suitable for subsequent docking studies.

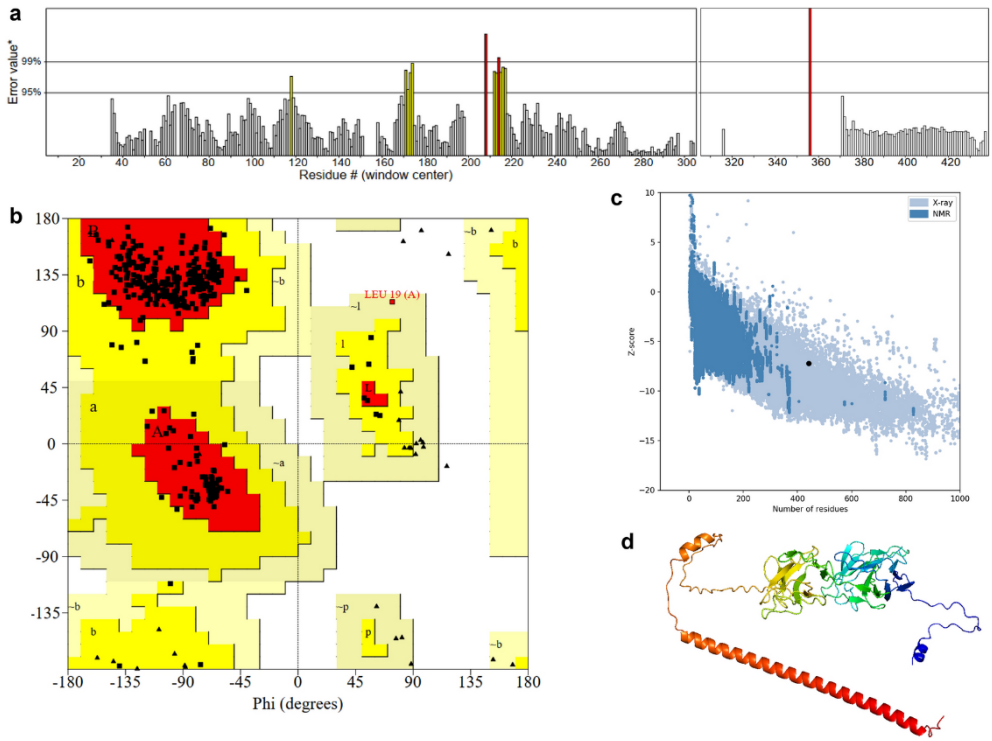


Figure 3.3 Tertiary structure of the vaccine candidate. (a) ERRAT program plot. (b) Ramachandran plot by PROCHECK webserver. (c) Z-score calculated by Pro-SA webserver. (d) Tertiary structure of refined Model 2 refined by GalaxyRefine, colored in rainbow from N-terminal (in blue) to C-terminal (in red).

3.4.6. Molecular docking of the vaccine candidate and immune receptor

Moving forward, I acquired the tertiary structure for the human TLR4 receptor (Uniprot ID: O00206) from the AlphaFold database. In the

case of the TLR4 structure, I retained only the extracellular domain encompassing amino acids 30–624, excluding other regions. For the docking of the vaccine with TLR4, I utilized the PatchDock webserver, adhering to default settings (Schneidman-Duhovny et al., 2005). The docked structure is illustrated in Figure 3.4, with four representative interactions highlighted in stick representation formed between the vaccine and TLR4 complex.

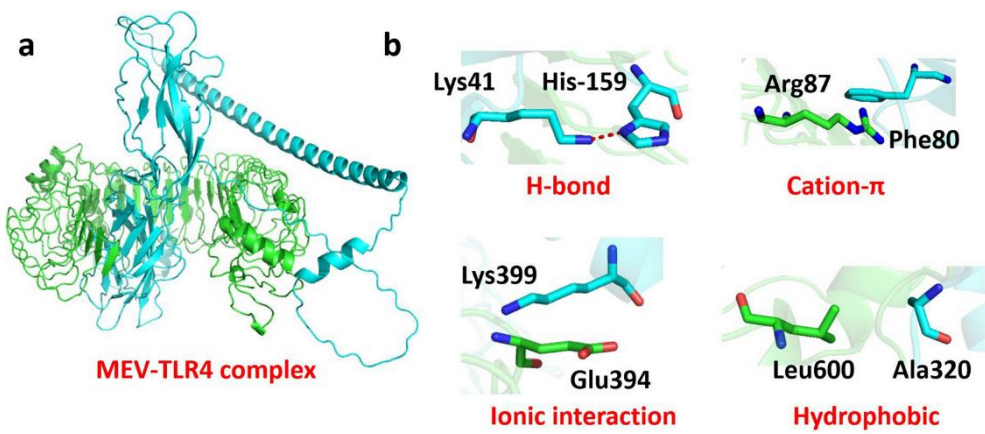
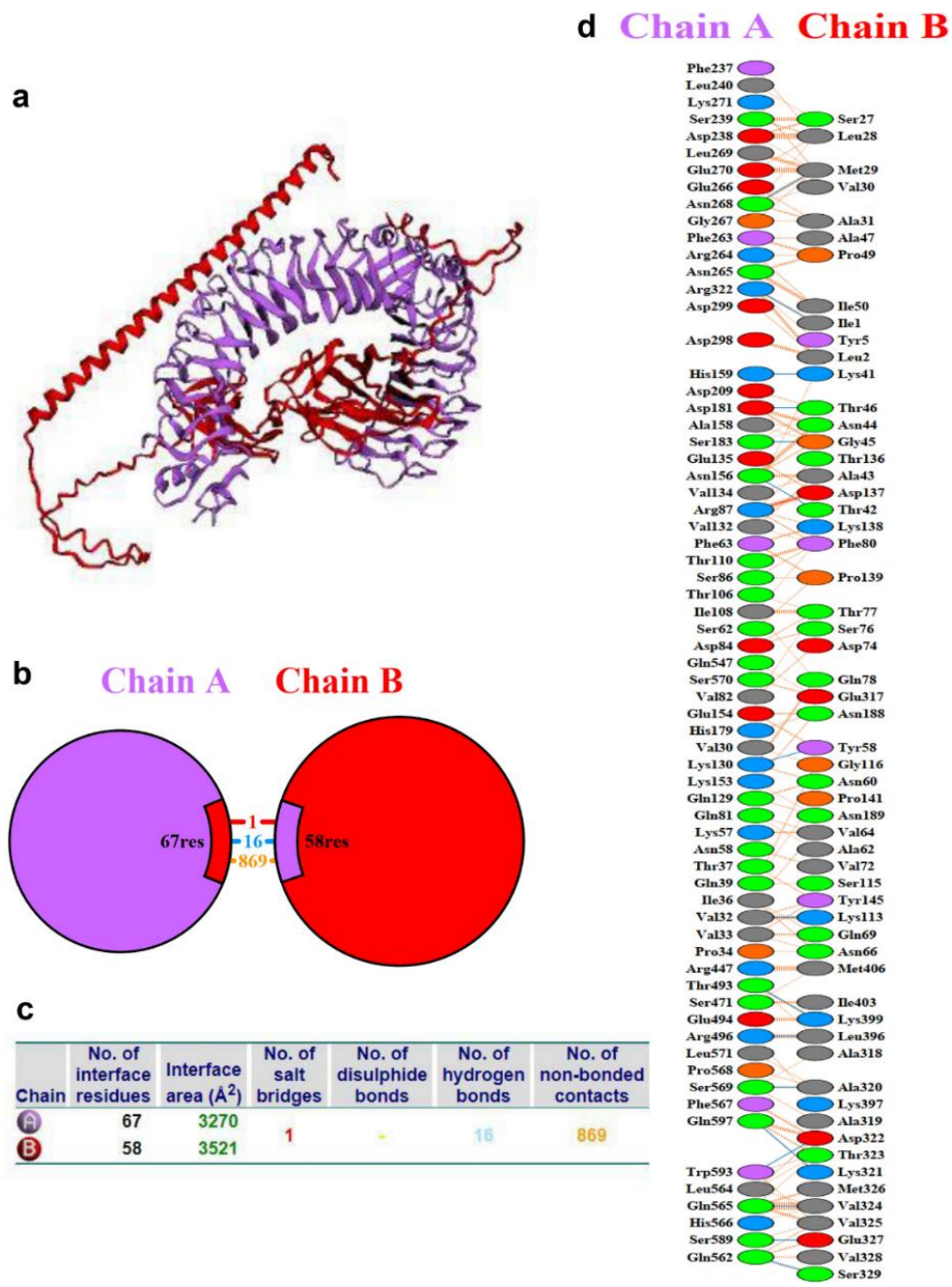


Figure 3.4 Docking and interaction analysis of the vaccine-TLR4 complex. (a) Molecular docking. (b) Four distinct inter-molecular representative interactions between vaccine and TLR4 complex are in sticks representation.

Besides, interacting residues between the vaccine and TLR4

were analyzed using the PDBsum web server (Laskowski et al., 2018). There are 67 and 58 residues interface, generating interface area 3270 and 3521 (Å) between TLR4 and vaccine, respectively. As a result, a total of 16 hydrogen bonds (blue lines), 1 salt bridge (red line), and 869 non-bonded contacts (orange dashed lines) were indicated in Figure 3.5.



Chain	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A	67	3270	1	-	16	869
B	58	3521				

Figure 3.5 Molecular docking between TLR4 (purple) and vaccine (red). (a) Three-dimensional representation of molecular docking. (b) Schematic diagram of interactions between protein chains. (c) Interface statistics. (d)

Residue interactions between TLR4 (chain A) and vaccine (chain B), salt-bridge (red line), hydrogen bonds (blue lines), and non-bonded contacts (orange dashed lines) between residues on either side of the vaccine-receptor interface.

3.4.7. Molecular dynamics simulations

To assess the stability of interactions in the vaccine and TLR4 complex, three replicas with different initial velocities throughout 100 ns of the MD simulations were run for all-atom using GROMACS 2023 software. Highly conserved results were achieved in all three replicates, as shown in Figure 3.6. For clarity, I will only report the results for the first replication. The root mean square deviation (RMSD) from the backbone of the complex, TLR4, and MEV were analyzed with average values of 0.786 ± 0.087 nm, 0.259 ± 0.046 nm, and 1.599 ± 0.375 nm respectively (Figures 3.6a–c). Besides, the root mean square fluctuation (RMSF) of the complexes was quantified to recognize the flexibility across the amino acid residues from the complexes. As depicted in Figure 3.6d, the RMSF values for TLR4 and MEV were calculated at 0.143 ± 0.057 nm and 1.109 ± 0.518 nm, respectively, indicating higher flexibility in the vaccine than TLR4. Subsequently, the radius of gyration in total (Rg) from the simulation

complex was also analyzed to examine the mobility and overall flexibility of the complex. Figure 3.6e shows that the complex had lower aberrations from 0 to 20 ns, and then the complex was stabilized and maintained the firm profile with a value of 3.565 ± 0.041 nm. Furthermore, the buried surface area (BSA) of the vaccine and TLR4 complex was calculated. The BSA for each residue of the docking complex was averaged over 100 ns to ensure the system had reached an energy-minimized state. Throughout the simulation, the BSA at the interaction interface between the MEV and TLR4 remained stable in all three repetitions with an average value of 66.460 ± 8.982 nm^2 (Figure 3.6f).

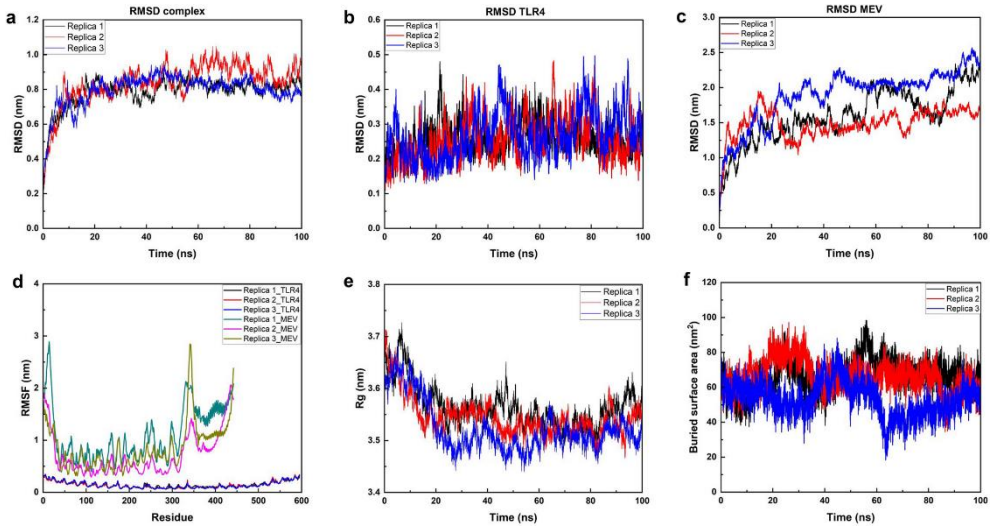


Figure 3.6 Molecular dynamics simulation studies. (a–c) RMSD backbone of the docking complex, TLR4, and MEV, respectively. (d) RMSF of the docking complex. (e) Rg in total of the docking complex. (f) BSA of the docking complex.

Additionally, the structural integrity's robustness was confirmed by aligning the entire complex involving the vaccine construct and the TLR4 receptor. Figure 3.7 illustrates a favorable alignment between these structures. The observed higher RMSD values in Figure 3.7a are attributed to the presence of unaligned, flexible regions, particularly the long and flexible loops within the vaccine construct. Despite these structural variations, the interaction pattern between the vaccine and TLR4 remains consistently stable.

Besides, to further explore the interface connecting TLR4 and the vaccine construct, like other researchers (Nguyen et al., 2022; Sharma et al.), specific snapshots were analyzed using the COCOMAPS tool as shown in Figure 3.7b. This analysis produced contact maps, visually representing the pairwise distances between residues of the vaccine construct and TLR4. In these contact maps, dots are color-coded: red, yellow, green, and blue signify distances below 7, 10, 13, and 16 Å, respectively.

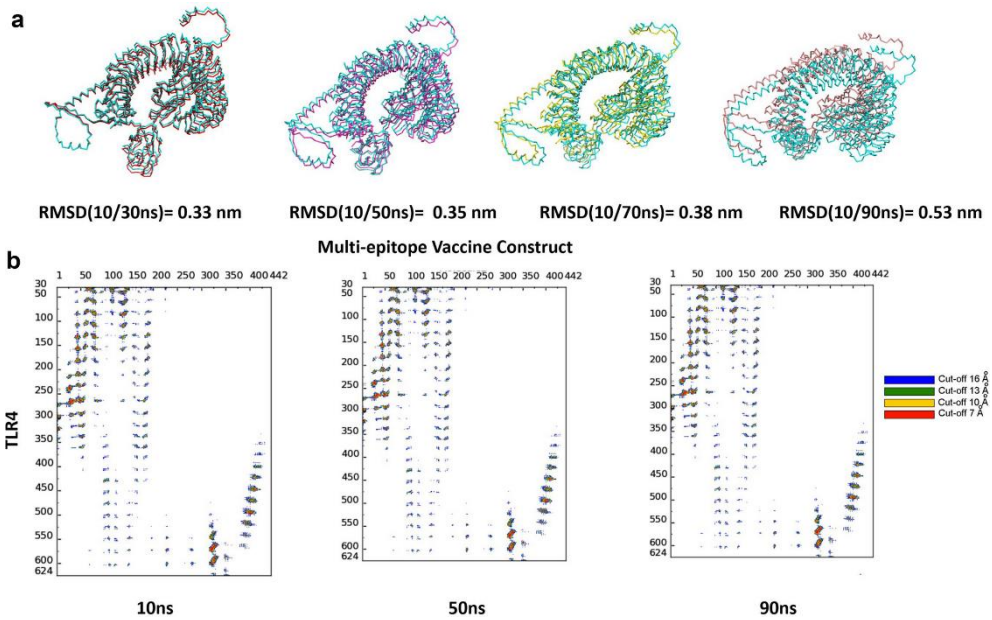


Figure 3.7 Dynamics and interaction patterns of TLR4-Vaccine complex. (a) Overlaying specific snapshots of the TLR4 and vaccine construct, along with their corresponding RMSD values in the initial simulation run. (b) Contact maps visually represent intermolecular interactions, indicating proximity between atom pairs within defined distances. Dots at the junction of two residues are color-coded, with red, yellow, green, and blue denoting closeness within 7, 10, 13, and 16 Å, respectively.

3.5. Discussion

POWV was discovered for the first time in Powassan, Canada, by an

encephalitis patient in 1958 (McLean Dm Fau - Donohue & Donohue). Usually, it causes sporadic infections, but since 2007, its cases have increased steadily in North America (Corrin et al., 2018). The frequency of POWV incidents stood at 0.7 per year between 1958 and 1998, increased to 1.3 cases annually spanning 1999 to 2005, and surged to 7.7 cases per year from 2006 to 2015 (Campbell & Krause, 2020). Moreover, the incidence rate of POWV has increased by 300% in the previous two decades (Choi et al., 2020). Unlike some other tick-borne diseases, POWV can be transmitted within 15 min after the bite of an infected tick (Della-Giustina et al., 2021). It leads to persistent neurological consequences in approximately 50% of documented cases and results in fatality in slightly over 10% of the cases (Campbell & Krause, 2020).

There are already six licensed vaccines for the flavivirus tick-borne encephalitis virus that can prevent the onset of neurological sequelae and severe disease; however, these vaccines are ineffective against POWV infection (Stone et al., 2022). No medicine or vaccine is currently approved for treating and preventing POWV disease. However, researchers continuously explore novel avenues to develop vaccine candidates or therapies for Powassan virus treatment. Antiviral molecules against flaviviruses, namely the adenosine

analogue NITD008 and NS5 capping enzyme inhibitor BG323, exhibited noteworthy reductions in POWV levels in vitro studies (Bullard et al., 2015; Yin et al., 2009). Intravenous immunoglobulin therapy has been utilized to manage POWV encephalitis for two patients. In both cases, the administration of intravenous immunoglobulin led to the survival of the patients with POWV infection. However, it's worth noting that one of the patients experienced considerable neurological complications following their recovery and discharge (Piantadosi et al., 2016). Recently, a vaccine candidate that uses the yellow fever virus vaccine strain as a vector expresses the prM/M and envelope proteins of POWV protected mice following lethal challenge and conferred a survival rate of 70% (Cheung et al., 2023).

Similarly, in another study, an mRNA vaccine encoding PrM/M and envelope proteins of POWV encapsulated in lipid nanoparticles generated neutralizing antibodies and protected mice from the lethal challenge of POWV (VanBlargan et al., 2018). In another study, Choi et al. developed a synthetic DNA vaccine consisting of POWV prM and envelope proteins, which elicited T cell and B cell immunity in mice and also protected the mice from POWV lethal challenge (Choi et al., 2020). Additionally, an immunogen displaying the domain III of the envelope glycoprotein of POWV presented on self-assembling protein

nanoparticle induced protective and neutralizing antibodies against POWV in mice (Malonis et al., 2022).

Developing a vaccine against POWV is crucial due to the potential severity of its neurological infections, including encephalitis and meningitis, often leading to long-term health consequences or even death. As an emerging tick-borne virus with increasing incidence over the years, the absence of specific antiviral treatments highlights the urgent need for preventive measures. A vaccine candidate against the POWV would provide a proactive solution for safeguarding public health and reducing healthcare burdens, aiding in curtailing the virus's geographical expansion and protecting vulnerable populations. Hence, in this study, a vaccine candidate has been developed using immunoinformatics approaches by targeting the structural proteins of POWV. Previously, several studies have targeted the Powassan structural proteins, mainly prM/M and envelope proteins, for developing vaccine candidates using strategies (Cheung et al., 2023; Choi et al., 2020; VanBlargan et al., 2018). Lately, there has been a notable rise in the adoption of immunoinformatics in creating vaccine candidates. This computational methodology offers a time- and cost-efficient means of developing innovative vaccine candidates. Another benefit of this method is its ability to identify numerous potential

vaccine candidates without the need for cultivating pathogenic organisms in a traditional laboratory setting (Mugunthan & Harish, 2021; Sami et al., 2021).

Moreover, comparable immunoinformatics techniques have been applied to formulate vaccine candidates targeting Monkeypox, Canine circovirus, Human cytomegalovirus, and Dengue virus (Akhtar et al., 2021; Akhtar et al., 2022; Kaushik, G, et al., 2022; Kaushik, Jain, et al., 2022). Previously, one study used the immunoinformatics approach to identify B-cell and T-cell epitopes only and did not evaluate the properties such as antigenicity, allergic potential, cytokine cytokine-inducing potential (Areeshi, 2018). Furthermore, the epitopes predicted in my study and the previous study were not similar, and the previous research did not design the final vaccine construct. The final epitopes selected in my study are antigenic, non-allergic, non-toxic and can induce various cytokine generation. Furthermore, the vaccine candidate formulated could be effective in generating an immune response in both the POWV lineages as both MHC-I epitopes and B-cell epitopes used in the final vaccine design are present in both lineage I and lineage II of the POWV. Furthermore, the two MHC-II epitopes from POWV lineage II that have high sequence similarity with the final two MHC-II epitopes used in the

vaccine construct have been added to the formulated vaccine construct to make the vaccine effective against lineage II virus as well.

To enhance the efficacy of the final vaccine construct, the final subunit of the *Escherichia coli* (*E. coli*) type 1 fimbria (FimH) and epitope of tetanus and diphtheria toxoid (TpD) adjuvants were used. Laboratory experiments have shown that TpD can stimulate the production of neutralizing antibodies and protect mucous membranes (Chan et al., 2020; Li et al., 2018). *E. coli* type 1 FimH has been reported to interact with TLR4, elicit the generation of IFN-γ and TNF-α, and play a role in the proliferation of local dendritic cells (Zhang et al., 2020; Zhang et al., 2022). Finally, the formulated POWV vaccine candidate was determined as antigenicity, toxicity, and allergenicity. Proteins with a molecular weight below 110 kDa are regarded as appropriate vaccine candidates (Sami et al., 2021). The formulated vaccine candidate had a molecular weight of 47.2 kDa, confirming its suitability as a vaccine candidate. Moreover, the vaccine candidate does not have homology with any human protein, thus minimizing the risk of autoimmune response to the hosts. The final vaccine protein exhibited an instability index of 29.45, suggesting its stability in biological conditions, as compounds with an instability index < 40 are considered stable. Additionally, according to

computational immune simulation, the formulated POWV vaccine candidate is expected to have the potential to trigger strong immune responses in those who receive it.

To delve deeper into the relationship between the vaccine candidate and TLR4, I conducted molecular docking and molecular dynamics simulations which verified the stability of these interactions. Three-dimensional representation of molecular docking between the vaccine candidate and TLR4 is shown in Figure 3.5a. Interestingly, the area of each circle is proportional to the surface area of the corresponding protein chain. The extent of the interface region on each chain is represented by the black wedge whose size signifies the interface surface area (Figure 3.5b). Statistics for this interface are given in Figure 3.5c with numbers of interface residues 67 (TLR4) and 58 (MEV), interface area of 3,270 $Å^2$ and 3,521 $Å^2$. Notably, interacting chains between two structures are joined by colored lines, each representing a different type of interaction, as per the key: salt-bridge (red line), hydrogen bonds (blue lines), and non-bonded contacts (orange dashed lines) (Figure 3.5d).

Furthermore, the root mean square deviation (RMSD) of the backbone and root mean square fluctuations (RMSF) of the C-alpha atoms for each residue were calculated to assess the structural

stability and flexibility. Figures 3.6a–c show that all of the docking complex's, TLR4's, and MEV's physical and chemical characteristics have reached equilibrium, meaning that their averages no longer vary with time. At the initial phase, the RMSD of the docking complex was lower, and then, it began to rise to 20 ns before reaching a stable plateau from 50 ns (Figure 3.6a). Meanwhile, because of the presence of adjuvants, the extremely flexible MEV construct has a significant RMSD (Figure 3.6c). Conversely, the TLR4 receptor appears to be extremely stable right from the beginning of the MD simulation (Figure 3.6b). Besides, the RMSF values for TLR4 and MEV were calculated at 0.143 ± 0.057 nm and 1.109 ± 0.518 nm, respectively, indicating higher flexibility in the vaccine than TLR4 (Figure 3.6d). I also studied the compactness of the receptor TLR4 interaction with the MEV using Rg. The receptor remains compact, and no unusual folding or unfolding was observed throughout the 100 ns. Finally, the following formula was used to determine the buried surface area (BSA) at the TLR4–MEV interface, where the solvent accessible surface area (SASA) values were obtained by the GROMACS program:

$$BSA_{interface} = (SASA_{TLR4} + SASA_{MEV}) - SASA_{TLR4\text{-}MEV}$$

At the interaction interface between the MEV and TLR4, the BSA averaged 66.460 ± 8.982 nm^2 in all three replicates of the

simulation (Figure 3.6f). Notably, Figure 3.7 highlights the interface's enduring stability across the sampled snapshots, as demonstrated by the inter-residue interactions. The above analyses indicate the stability of interface interactions between the TLR4 and the vaccine construct. In summary, the MEV exhibited robust interactions and stability with the TLR4 receptor, as demonstrated by our triplicate MD simulations and molecular docking investigations.

This chapter was published in *Synthetic and Systems Biotechnology*

as a partial fulfillment of Nguyen Thi Truc Ly's Ph.D program.

Chapter 4 Discovering peptides and computational investigations of a multi-epitope vaccine target *Mycobacterium tuberculosis*

4.1. Abstract

Mycobacterium tuberculosis is the causative agent of tuberculosis, a prevalent airborne infectious disease. Despite the availability of the Bacille Calmette-Guerin vaccine, its global efficacy remains modest, and tuberculosis persists as a significant global public health threat. Addressing this challenge and advancing towards the End *Mycobacterium tuberculosis* Strategy, I developed a multi-epitope vaccine based on immunoinformatics and computational approaches. Immunoinformatics screening of *Mycobacterium tuberculosis* protein identified immune-dominant epitopes based on major histocompatibility complex allele binding, immunogenicity, antigenicity, allergenicity, toxicity, and cytokine inducibility. Selected epitopes were integrated into a multi-epitope vaccine construct with adjuvant and linkers, forming a fully immunogenic vaccine candidate. Comprehensive analyses encompassed the evaluation of immunological and physicochemical properties, determination of tertiary structure, molecular docking with toll-like receptors, molecular dynamics simulations for all atoms, and immune simulations. Our MEV comprises 534 amino acids, featuring 6 cytotoxic T lymphocyte, 8 helper T lymphocyte, and 7 linear B lymphocyte epitopes, demonstrating high antigenicity and stability. Notably,

molecular docking studies and triplicate molecular dynamics simulations revealed enhanced interactions and stability of multi-epitope vaccine with the toll-like receptor 4 complex compared to toll-like receptor 2. In addition, the immune simulation indicated the capacity to effectively induce elevated levels of antibodies and cytokines, emphasizing the vaccine's robust immunogenic response. This study presents a promising multi-epitope vaccine against tuberculosis, exhibiting favorable immunological and physicochemical attributes. The findings provide theoretical support for tuberculosis vaccine development. My study aligns with the global initiative of the End *Mycobacterium tuberculosis* Strategy, emphasizing its potential impact on addressing persistent challenges in tuberculosis control.

4.2. Introduction

Tuberculosis (TB), a bacterial disease that primarily affects the lungs, is preventable and treatable, but 10 million people still catch it annually, and 1.6 million people died from TB in 2021, almost entirely in low and middle-income countries (Chakaya et al., 2022). TB has long been the world's deadliest infectious disease treatment, although it has suffered a setback and has been disrupted due to the COVID-19 pandemic (Zimmer et al., 2022). Like the Bacillus Calmette-Guerin (BCG)

vaccine, tools to fight TB are imperfect. However, there is “hopeful” innovation in vaccines like DNA vaccines, live attenuated and killed whole-cell vaccines (WCVs), multi-epitope vaccines (MEVs), etc. Among them, the therapeutic DNA vaccine is a promising vaccine, which is a promising strategy against tuberculosis. However, challenges with the delivery and expression of the DNA vaccine and potential issues with inducing an adequate immune response in all individuals may limit its efficacy (Li & Zhu, 2006). Live attenuated and killed whole-cell vaccines (WCVs) also offer promising vaccination strategies against tuberculosis. However, their efficacy may be compromised in immunocompromised individuals, and there is a risk of virulence reversion in live attenuated vaccines, leading to the potential for disease transmission (Scriba et al., 2016). MEVs are a type of vaccine that can be composed of cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and linear B lymphocyte (LBL) epitopes in a series or overlapping epitope peptides (Zhang, 2018). They are designed to induce multi antigenic immunity against significant complex pathogens with different strain variants (Khairkhah et al., 2022). MEVs can be used to prevent and treat tumors or viral infections (Mao et al., 2023; Truc Ly Nguyen & Heebal Kim, 2024a; Zhang, 2018). Besides, the advantage of epitope vaccines over

traditional subunit vaccines lies in the ability to combine immunodominant human HTL, CTL, and LBL epitopes from different antigens, thereby enhancing immunogenicity and reducing adverse effects. Currently, many MEVs based on epitope designs from multiple antigens against tuberculosis are being researched (Andongma et al., 2023; Bellini et al., 2023; Bibi et al., 2021b; Cheng et al., 2023; Cheng et al., 2022; Gong et al., 2021; Jiang, Peng, et al., 2023; Khan et al., 2023a; Nayak et al., 2023; Ruaro-Moreno et al., 2023; Sharma et al., 2021a). Among them, Jiang et al. selected 17 latent tuberculosis infection and regions of difference (LTBI-RD) antigens (Rv1511, Rv1736c, Rv1737c, Rv1980c, Rv1981c, Rv2031c, Rv2626c, Rv2653c, Rv2656c, Rv2659c, Rv2660c, Rv3425, Rv3429, Rv3872, Rv3873, Rv3878, and Rv3879) to identify immunodominant epitopes (Jiang, Peng, et al., 2023). Similarly, Bellini et al. designed and characterized a multistage peptide-based vaccine from 15 protein antigens associated with various activities of the MTB life cycle, including Rv3908c, Rv1886c, Rv1384, Rv1436, Rv3874, Rv0288, Rv0867c, Rv1174c, Rv1334, Rv0475, Rv0440, Rv0125, Rv1733c, Rv1039c, and Rv1039c (Bellini et al., 2023). In addition, a novel peptide-based vaccine was designed based on HTL, CTL, and B-cell epitopes predicted from 17 protective antigens of MTB. These 17 candidate

antigens were Ag85A, Ag85B, ESAT6, EspA, Mpt63, MTB32A, PPE18, RpfB, TB10.4, CFP10, MPT51, MPT64, MTB8.4, PPE44, PPE68, RpfA, and RpfB (Cheng et al., 2023). Notably, Bibi et al. showed that MEV might activate humoral and cellular immune responses and may be a possible tuberculosis vaccine candidate (Bibi et al., 2021b). In that work, a novel MEV designed by targeting Rv2608, Rv2684, Rv3804c (Ag85A), and Rv0125 (MTB 32A) which has been predicted to have different B-cell and T-cell epitopes. Another promising candidate is an MEV against MTB exploiting secreted exosome proteins (Voysey et al.) (Sharma et al., 2021a). However, the potential for immune evasion by MTB through antigenic variation may limit the vaccine's effectiveness, and there is a need for ongoing monitoring and updating of vaccine components to address this challenge.

Beyond these considerations, H37Rv is the most widely used MTB strain, and its protein Rv0256c, also known as PPE2 (Proline-Proline-Glutamate 2), has an essential role in immune activation and infection of the host (Pal et al., 2021). Rv0256c has been found to translocate to the nucleus of host cells and bind to the promoter region of inducible nitric oxide synthase (iNOS), suppressing iNOS gene transcription (Bhat et al., 2013; Bhat et al., 2017), ultimately protecting the mycobacterium from nitric oxide (NO) mediated killing.

Additionally, Rv0256c has been found to inhibit myeloid hematopoiesis and reactive oxygen species (ROS) production (Pal & Mukhopadhyay, 2021; Srivastava et al., 2019). In a previous study, Rv0256c induces a strong B cell response in tuberculosis patients (Abraham et al., 2014). These findings suggest that Rv0256c is a crucial protein contributing to MTB survival and pathogenesis. Therefore, in the present study, I aim to select the Rv0256c protein as the target sequence to design an MEV candidate. Through immunoinformatics techniques, I predicted CTL, HTL, and LBL epitopes. These epitopes were shown to be highly antigenic, nontoxic, and nonallergic. The potential for these epitopes to cause autoimmunity was also examined. In addition, the Toll-Like Receptor 4 (TLR4) agonist (RpfE) peptide was added to the vaccine design as an adjuvant to boost its immunogenicity. We further evaluated the vaccine construct's population coverage, antigenicity, allergenicity, toxicity, and physicochemical features. Afterwards, the tertiary structures of the vaccine construct were predicted, refined, and validated. The resultant tertiary structure was then docked with immune receptors TLR2 and TLR4. Furthermore, the stability of interactions was verified using molecular dynamics (MD) simulations for all atoms of the docking complexes in triplicate. Finally, to assess the immunogenicity and immunological response of the MEV, in silico

immune simulations were carried out. This study provides valuable insights for MTB vaccine development and contributes towards the End MTB Strategy.

4.3. Materials and Methods

4.3.1. Retrieval sequence, screening antigenicity, and allergenicity of target protein

The FASTA sequence of Rv0256c was obtained from the UniProt database (https://www.uniprot.org/) with accession number P9WI47. To screen for antigenicity, I employed VaxiJen v2.0 to predict the antigenicity of the Rv0256c protein, with a threshold value of 0.4 set up (http://www.ddg-pharmfac.net/VaxiJen/VaxiJen/VaxiJen.html) (Irini A Doytchinova & Darren R Flower, 2007). This server is focused on auto cross-covariance (ACC) transformation and alignment-independent prediction that maintains predictive accuracy of 70–89%. Protein Rv0256c showed antigenicity above the threshold value and was selected for further analysis (Irini A Doytchinova & Darren R Flower, 2007). For predicting allergenicity, the Rv0256c protein sequence was expanded for further analysis based on AllergenFP v.1.0, a bioinformatics tool for allergenicity prediction (https://ddg-

pharmfac.net/AllergenFP/) (Dimitrov et al., 2013). The result with a non-allergen property was selected for further analysis. To enhance humoral and cell-mediated immunity, both B-cell and T-cell antigens were predicted.

4.3.2. CTL epitopes prediction and assessment

The Immune Epitope Database and Analysis Resources (IEDB) MHC I server was used to predict CTL epitopes (http://tools.iedb.org/mhci/) (Fleri et al., 2017). This server predicts $CD8^+$ T-cell epitopes based on proteasomal C-terminal cleavage, MHC-I binding, and TAP transport efficiency. The 9-mer and 10-mer epitopes were predicted using the ANN 4.0 algorithm weight matrix, artificial neural networks, and IC50 value. I used the entire human HLA allele reference set. Finally, the predicted peptides were sorted as per the predicted IC50. As in the previous study, only epitopes with an IC50 over 500 were chosen (Al Tbeishat, 2022).

4.3.3. HTL epitopes prediction and assessment

The MHC II server of IEDB was used to predict HTL epitopes (http://tools.iedb.org/mhcii/) (Fleri et al., 2017). NN-align 2.3 (Net

MHC II 2.3) algorithm was used to predict the epitopes. The entire HLA human reference set was used. A 15mer epitope length was defined. Finally, the results were then arranged as per their modified ranks. Additionally, the ability to secrete IFN-γ, IL-4 and IL-10 of these chosen epitopes was predicted using the servers IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/) (Dhanda, Vir, et al., 2013a), IL4pred (https://webs.iiitd.edu.in/raghava/il4pred/) (Dhanda, Gupta, et al., 2013a), and IL-10 pred (http://crdd.osdd.net/raghava/IL-10pred/) (Nagpal et al., 2017b), respectively. All chosen epitopes showed the ability to secrete these cytokines.

4.3.4. B-cell epitopes prediction and assessment

An online server ABCpred was used to predict linear B-cell epitopes (http://crdd.osdd.net/raghava/abcpred/) (Saha & Raghava, 2006). The ABCpred uses an artificial neural network to predict linear B-cell epitopes in an antigen sequence. A 0.51 threshold was applied individually to each selected protein. Epitopes were chosen to have a 16mer length.

4.3.5. Antigenicity, allergenicity, and toxicity prediction of chosen epitopes

All chosen epitopes were tested to assess their antigenicity, allergenicity, and toxicity. For antigenicity prediction, CTL epitopes, HTL epitopes, and conformational B-cell epitopes amino acid sequences were submitted to the VaxiJen v2.0 server (Irini A Doytchinova & Darren R Flower, 2007). Both a threshold of 0.4 and bacteria were explicitly mentioned. Highly antigenic epitopes were selected for further analysis. Consequently, the allergenicity of the B cell, CTL, and HTL epitopes was predicted using the freely available allergenicity prediction tool AllerTOP v.2.0 (https://www.ddg-pharmfac.net/AllerTOP/index.html). AllerTOP v.2.0 predicts the allergens based on machine learning methods like auto and cross-covariance transformation, k nearest neighbors, and amino acid E-descriptors (Gonzalez-Galarza et al., 2019). All settings were left at their default values. Lastly, the toxicity of the B cell, CTL, and HTL epitopes was predicted using the freely server ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) (Gupta, Kapoor, Chaudhary, Gautam, Kumar, & Raghava, 2013). Only the epitopes identified as antigenic, non-allergenic, and non-toxic were retained for further research.

4.3.6. Population coverage of the epitopes

The diversity in MHC allele distribution reflects the world's geographical and cultural diversity. As a result, vaccination coverage is determined by the MHC alleles that its epitopes recognize. By using the population coverage tool in the IEDB database (http://tools.iedb.org/population/) (Bui et al., 2006), the epitopes and their corresponding MHC class I and class II alleles as input, we calculated the combined coverage of our T lymphocyte epitopes. These alleles genotypic frequencies employed in the IEDB database were obtained from the Allele Frequency Net Database (AFND) (http://www.allelefrequencies.net/) (Gonzalez-Galarza et al., 2019). Currently, AFND offers allele frequencies for 115 countries and 21 different ethnicities grouped into 16 different geographical areas.

4.3.7. Multi-epitope vaccine design and its properties evaluation

Highly antigenic, nonallergenic, and nontoxic epitopes were selected to design the vaccine. These best epitopes were linked through EAAAK, GPGPG, KK, and AAY linkers to construct a potential MEV. These linkers were incorporated into separate domains to let them act separately and enhance the vaccine's immunogenicity (Hajighahramani et al., 2017) as they are cleavable, flexible, and rigid (Chawla et al.,

2023). Apart from these epitopes and linkers, the TLR4 agonist (RpfE) peptide was added to the vaccine design as an adjuvant for increasing the immune response (Lee et al., 2014).

The ability to elicit humoral-cellular immune responses and the knowledge of a particular antigen associated with an immune response are referred to as immunogenicity and antigenicity, respectively. Therefore, the antigenicity and immunogenicity of a candidate vaccine are vital (Ilinskaya & Dobrovolskaia, 2016). ToxinPred and AllerTOP 2.0 were used to predict the toxicity and allergenicity of the vaccine construct. Allergenicity was checked to ensure that the vaccine did not exhibit any reactions (allergic) once injected into the body. The vaccine candidate should have a high level of antigenicity because this attribute defines an antigen's capacity to trigger an immune response and the development of memory cells. Antigenicity prediction was made using VaxiJen v2.0 and ANTIGENpro server (Magnan et al., 2010). Both methods are alignment-free. VaxiJen v2.0 functions by utilizing several physicochemical properties of the protein, while ANTIGENpro is a machine learning algorithm-based microarray analysis data-based server. Using the open web server ProtParam, many physiochemical features, including amino acid composition, Aliphatic Index (AI), molecular weight, Instability Index (II), Grand

Average of Hydropathicity (GRAVY), and theoretical isoelectric point (pI) were evaluated (Gasteiger et al., 2005).

4.3.8. Structure prediction, refinement, and validation of the vaccine construct

The vaccine's secondary structure motifs calculated with PDBsum (Laskowski et al., 2018) were computed using v.3.0 of Gail Hutchinson's PROMOTIF program (Hutchinson & Thornton, 1996). PDBsum is a web server (http://www.ebi.ac.uk/pdbsum) that offers structural details about the entries in the Protein Data Bank (PDB). Protein secondary structure, interactions between proteins and ligands and DNA, PROCHECK structural quality evaluations, and numerous more analyses are among the mostly image-based analyses. The PROMOTIF program examines a protein coordinate file and gives information regarding the structural motifs present in the protein. Currently, the program evaluates the following structural features: secondary structure, β- and γ-turns, helical geometry and interactions, β-strands and β-sheet topology, β-hairpins, etc. In order to illustrate each type of motif in the protein, PROMOTIF generates postscript files along with a summary page.

For predicting the three-dimensional (3D) structure of the

vaccine, I used several servers such as SWISS-MODEL (https://swissmodel.expasy.org/) - a fully automated protein structure homology-modeling server (Waterhouse et al., 2018), Iterative Threading ASSEmbly Refinement (I-TASSER) (https://zhanggroup.org/I-TASSER/) (Zhang, 2008), or a deep learning approach AlphaFold2 via ColabFold v1.5.5 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) (Jumper et al., 2021; Mirdita et al., 2021). Afterwards, the GalaxyRefine module of the GalaxyWEB server (http://galaxy.seoklab.org/) was used to refine the vaccine's 3D structures (Ko et al., 2012). Through the process of 3D structural refinement, the vaccine's near-accurate native structure was preserved while local mistakes were corrected and the accuracy of initially anticipated structures was improved. Subsequently, ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein & Sippl, 2007) and PROCHECK v.3.5 (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) (Laskowski et al., 1993) were two of the many publicly available tools utilized for 3D structural validation. ProSA-web is a web server that evaluates the overall quality and local model quality of 3D models based on the Z-score value (Akhtar et al., 2023; Sharma et al., 2023). Meanwhile, the PROCHECK program

analyzes the Ramachandran plot for the 3D structure of the vaccine construct to assess residual coverage in favored, allowed, and disallowed regions. A good quality structure would be expected to have over 90% of residues in the most favored regions.

4.3.9. Molecular docking and molecular dynamics simulations studies

The structure of TLR2 and TLR4 were downloaded from the Protein Data Bank (PDB) with ID 2Z7X and 3FXI, respectively. The 3D structure of MEV and immune receptors were docked using the ClusPro server (https://cluspro.bu.edu) – a widely used protein-protein docking server that predicts the 3D structures of protein complexes (Kozakov et al., 2017b). The following three steps are used by ClusPro to examine the molecular docking of vaccine with TLRs: (Rando et al.) rigid body docking by sampling billions of conformations, (2) grouping of the 1000 lowest energy structures generated to identify the largest clusters based on root-mean-square deviation (RMSD), (3) energy minimization for steric clash removal. Accordingly, the server provided 30 model complexes, out of which the model having the lowest binding energy (kcal/mol) was selected for dynamics.

Subsequently, PDBsum was used for analysis and to find interacting residues between the vaccine and TLR2, 4 (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) (Laskowski et al., 2018).

MD simulations were conducted using the GROMACS program on a Linux operating system to assess the stability of the complexes (Abraham et al., 2015). The CHARMM27 force field and spce water were employed to generate topology files, resulting in a system with 24,783 atoms from 1603 residues for the MEV-TLR2 complex and 26,709 atoms from 1736 residues for the MEV-TLR4 complex. Each complex was placed in a cubic box (12x12x12 for MEV-TLR2 and 11x11x11 for MEV-TLR4) to maintain integrity with 297,338 and 227,396 solvent molecules, respectively. To neutralize the charge, 27 Cl^- ions were added to the MEV-TLR2 complex, and 5 Na^+ ions were added to the MEV-TLR4 complex. The energy minimization utilized the steepest descent algorithm with 50,000 steps, stopping when the maximum force was <1000.0 kJ/mol/nm. Position restraints were applied during equilibration, including NVT equilibration at 300 K with 50,000 steps (100 ps) and NPT equilibration at 1 bar reference pressure with an additional 50,000 steps (100 ps). Production simulations for all-atom systems (916,797 atoms in MEV-TLR2 and

708,897 atoms in MEV-TLR4) were carried out using the NPT ensemble for 50,000,000 steps (100 ns). After completing the 100 ns MD simulation, analyses were performed, including calculating the root mean square deviation (RMSD) of backbone residues, root mean square fluctuation (RMSF) of C-alpha, radius of gyration (Rg), and solvent accessible surface area (SASA). Each complex was simulated in triplicate to ensure result accuracy, robustness, and dependability.

4.3.10. Immunological responses induced by the vaccine construct

C-IMMSIMversion 10.1 is an immune simulation tool that can assess the vaccine's immunological response (https://kraken.iac.rm.cnr.it/C-IMMSIM/) (Rapin et al., 2012). It predicts a position-specific scoring matrix used to understand immune response magnitude, which shows the result of vaccine dosage concerning different time intervals. This server describes a mammalian immune system's humoral and cellular response against vaccine construct. To quantify the impact of antigens and foreign particles on immune activity, an agent-based method based on the position-specific scoring matrix and machine learning techniques was applied. Except for the time steps, which were set at 1, 84, and 168, the simulation was run using the default parameters.

Injection occurs at time = 0 in time step 1, and each time step lasts for 8 hours. The entire simulation consisted of 1050 steps. Since most commercial vaccines prescribe a four-week delay between doses, three injections were anticipated to be needed at four-week intervals (Dey et al., 2023).

4.4. Results

4.4.1. Retrieval sequence, screening antigenicity, and allergenicity of target protein

After retrieving the FASTA sequence of the Rv0256c protein from the UniProt database, immunoinformatics analysis was used to predict the antigenicity and allergenicity of the protein. The Rv0256c protein was predicted as a probable antigen (0.4302) based on the VaxiJen v2.0 server with a threshold of 0.4. AllergenFP v.1.0 results indicated that the protein had non-allergenic properties as it held the highest Tanimoto similarity index of 0.94. Thus, I confirmed that the Rv0256c protein sequence could be considered for CTL, HTL, and B-cell epitope prediction.

4.4.2. CTL, HTL, and B-cell epitopes prediction and assessment

The Rv0256c protein sequence predicted CTL epitopes using the MHC-I tool from the IEDB server. From the IEDB server, the completed human HLA allele reference set was selected for epitope prediction. Epitopes with IC50 over 500 were chosen for further study. Further on, epitopes were checked for toxicity, antigenicity, and allergenicity. Only non-toxic, antigenic, and non-allergenic epitopes were chosen (Supplementary Data Sheet 1). An immunogenicity check was done using the IEDB server, and immunogenic epitopes with scores ≥0.3 were selected. Finally, six epitopes were included in the vaccine construct (Table 4.1).

Table 4.1 CTL epitopes for vaccine construction

Peptide	Length	Antigenicity scores[a]	Toxicity	Allergenicity	Immunogenicity scores[b]
LMATNFFGIN	10	0.9704	Non-Toxin	Non-Allergen	0.37602
FSGFDPWLPS	10	0.9231	Non-Toxin	Non-Allergen	0.33227
PANIAFALGY	10	0.6439	Non-Toxin	Non-Allergen	0.33119
VIQPFINWL	9	0.5510	Non-Toxin	Non-Allergen	0.31448
SPANIAFALG	10	0.5157	Non-Toxin	Non-Allergen	0.31063
GNPATIAFT	9	1.0466	Non-Toxin	Non-Allergen	0.30027

[a]Scores ≥ 0.4; [b]Score ≥ 0.3

The IEDB MHC II server was used for HTL epitope prediction. Epitopes were checked for toxicity, antigenicity, and allergenicity.

Only non-toxic, antigenic, and non-allergenic epitopes were chosen. Epitopes were checked and selected based on their ability to induce Il-4, Il-10, and IFN-γ (Supplementary Data Sheet 2). Finally, I included eight possible epitopes in the vaccine that induced the abovementioned cytokines (Table 4.2).

Table 4.2 HTL epitopes for vaccine construction

Peptide	Il-4 inducer	IFN-γ scores	Il-10 inducer	Toxicity	Antigenicity scores[a]	Allergenicity
AQARKAVGTGVRKKT	Yes	0.08	Yes	Non-Toxin	1.1876	Non-Allergen
ARKAVGTGVRKKTPE	Yes	0.02	Yes	Non-Toxin	1.2335	Non-Allergen
LNSAAQARKAVGTGV	Yes	0.41	Yes	Non-Toxin	0.8428	Non-Allergen
QAMFSGFDPWLPSLG	Yes	0.02	Yes	Non-Toxin	0.6122	Non-Allergen
QARKAVGTGVRKKTP	Yes	0.07	Yes	Non-Toxin	1.1250	Non-Allergen
SAAQARKAVGTGVRK	Yes	0.39	Yes	Non-Toxin	1.1115	Non-Allergen
VGDLNSAAQARKAVG	Yes	0.16	Yes	Non-Toxin	0.5454	Non-Allergen
AAQARKAVGTGVRKK	Yes	0.16	Yes	Non-Toxin	1.0585	Non-Allergen

[a]Scores ≥ 0.4

B-cell epitopes having a rank <10 predicted using the ABCpred webserver were selected for further studies. Furthermore, I selected only those epitopes that were non-toxic, non-allergenic, and antigenic to further incorporate into the vaccine design (Supplementary Data Sheet 3). I used the web servers ToxinPred, AllerTOP, and VaxiJen, in that order. Finally, seven epitopes were selected for the vaccine construct (Table 4.3).

Table 4.3 LBL epitopes for vaccine construction

Peptide	Antigenicity scores[a]	Toxicity	Allergenicity
TGVRKKTPEPDSAEAP	0.8432	Non-Toxin	Non-Allergen
PVAAIAPSIPTPTPTP	0.9063	Non-Toxin	Non-Allergen
TGSPQGAGTLGFAGTT	0.8953	Non-Toxin	Non-Allergen
RGYEYLDLDPETGHDP	0.9490	Non-Toxin	Non-Allergen
AQARKAVGTGVRKKTP	1.1161	Non-Toxin	Non-Allergen
APQIVKANAPTAASDE	0.7838	Non-Toxin	Non-Allergen
AWLVQASANSAAMATR	0.6719	Non-Toxin	Non-Allergen

[a]Scores ≥ 0.4

4.4.3. Population coverage analysis

The population coverage of the 14 T-lymphocyte epitopes (combined CTL and HTL) employed in this designed vaccine was evaluated using IEDB population coverage analysis. The IEDB database assessed the distribution of their 55 corresponding MHC alleles in 16 geographical areas and 101 countries. The region-wise coverage of alleles is

represented in Figure 4.1. Notably, our vaccine demonstrates a global coverage rate of 99.74%.

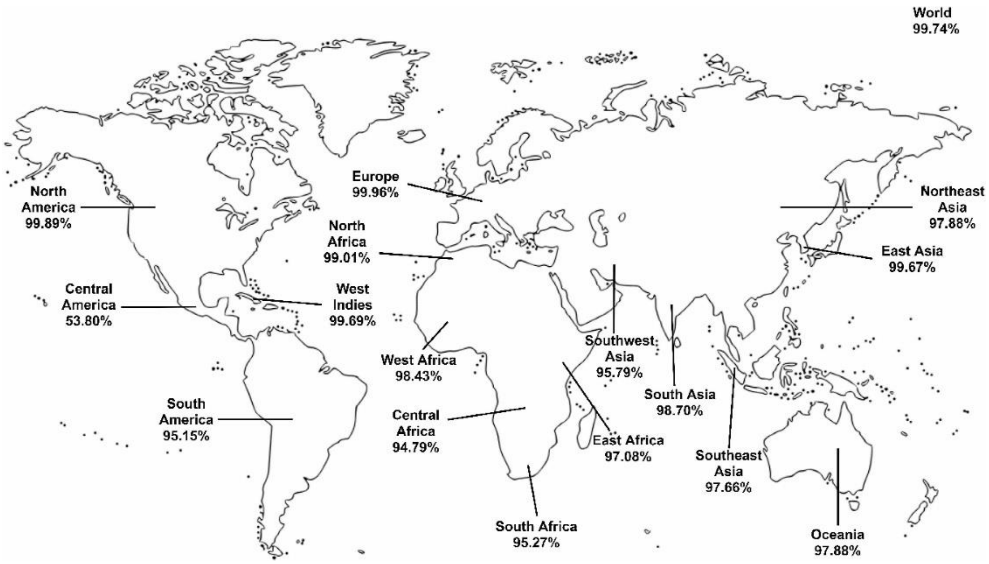


Figure 4.1 Population coverage for T-lymphocytes.

4.4.4. Multi-epitope vaccine design and its properties evaluation

For the MEV construction, highly antigenic, non-allergenic, and non-toxic epitopes were selected. According to the results in Table 4.1, Table 4.2, and Table 4.3, six CTL epitopes, eight HTL epitopes, and seven B-cell epitopes were selected. The selected epitopes were linked with amino acid linkers like EAAAK, GPGPG, KK, and AAY. Furthermore, to increase the immune response, the TLR4 agonist

(RpfE) peptide was added as an adjuvant to the N-terminal of the vaccine since, in their study, Lee et al. discovered the new finding that MTB directly binds TLR4 and initiates TLR4 signaling, which in turn causes DCs to produce IL-1 beta and express co-stimulatory and MHC antigen presentation molecules (Lee et al., 2014). As a result, the amino acid sequence of the constructed vaccine is mentioned in Figure 4.2.

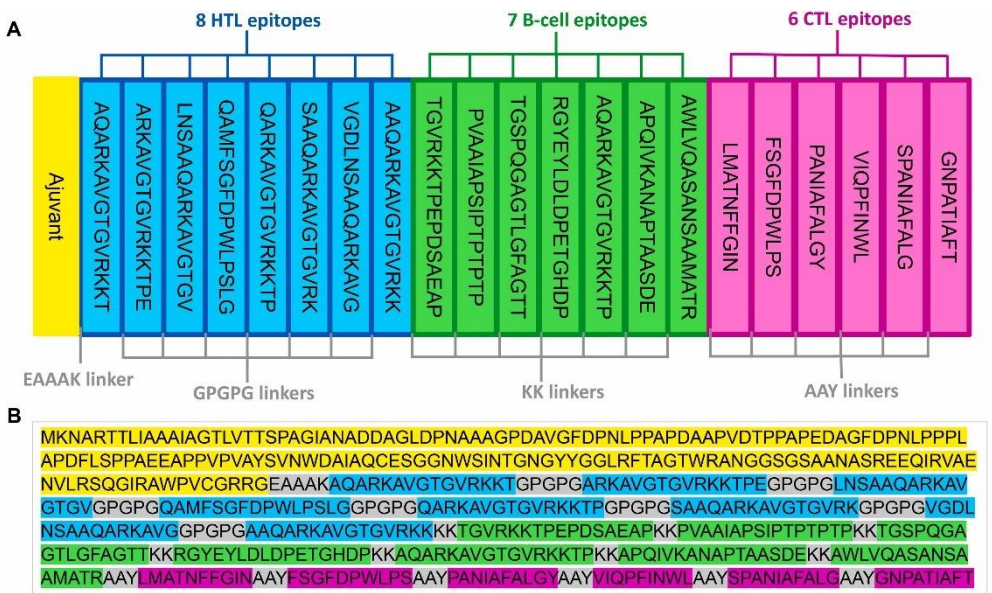


Figure 4.2 Structural details of MEV. Schematic representation (A) and sequence (B) of the final vaccine construct with linkers, adjuvant, and epitopes sequentially and appropriately.

The antigenicity, allergenicity, toxicity, and physicochemical analyses have been listed in Table 4.4. The vaccine was predicted to be antigenic by VaxiJen and ANTIGENpro with scores of 0.9363 and 0.9399, respectively. The vaccine was expected to be non-allergenic by AllerTOP. It was found to be non-toxic by ToxinPred. The ProtParam server determined the vaccine construct's molecular weight to be 53.80 kDa, and its 10.10 pI suggested that it had basic properties. Of those residues, there were 534 amino acids; 64 were positively charged, and 33 were negatively charged. The II was calculated to be 28.63 in terms of instability, indicating that the construct is stable following expression (a value above 40 predicts that the protein may be unstable). According to the AI calculation, the construct is thermostable, with a value of 62.77. The Grand Average of Hydropathicity (GRAVY), which indicates how hydrophilic a substance is, was estimated to be negative (-0.354). Based on these results, this MEV construct can be predicted as a potential vaccine candidate.

Table 4.4 Evaluation of antigenicity, toxicity, allergenicity, toxicity, and physicochemical properties of the vaccine construct

Features	Assessment	Remark
Number of amino acids	534	-
Molecular weight	53797.90	Average
Total number of atoms	7580	Average
Theoretical pI	10.10	Basic nature
Total number of negatively charged residues (Asp + Glu)	33	-
Total number of positively charged residues (Arg + Lys)	64	-
Aliphatic index (AI)	62.77	Thermostable
Instability Index (II)	28.63	Stable
Estimated half-life (mammalian reticulocytes, in vitro)	30 hours	Satisfactory
Estimated half-life (yeast cells, in vivo)	>20 hours	Satisfactory
Estimated half-life (*Escherichia coli*, in vivo)	>10 hours	Satisfactory
Grand Average of hydropathicity (GRAVY)	-0.351	Hydrophilic
Antigenicity	0.9363 (VaxiJen v2.0)	Antigenic
	0.9399 (ANTIGENPro)	Antigenic
Allergenicity	Non-allergen (AllerTOP v2.0)	Non-allergen
Toxicity	ToxicPred	Non-toxic

4.4.5. Structure prediction, refinement, and validation of the vaccine construct

The vaccine's secondary structure motifs are shown in Figure 4.3 which were computed using v.3.0 of Gail Hutchinson's PROMOTIF program. Particularly, among 534 residues, there are 0.7% β-strand, 25.5% α-helix, 1.1% 310-helix, 10.5% β-turn, 1.2% γ-turn, 0.7% β-hairpins, and 60.3% others.

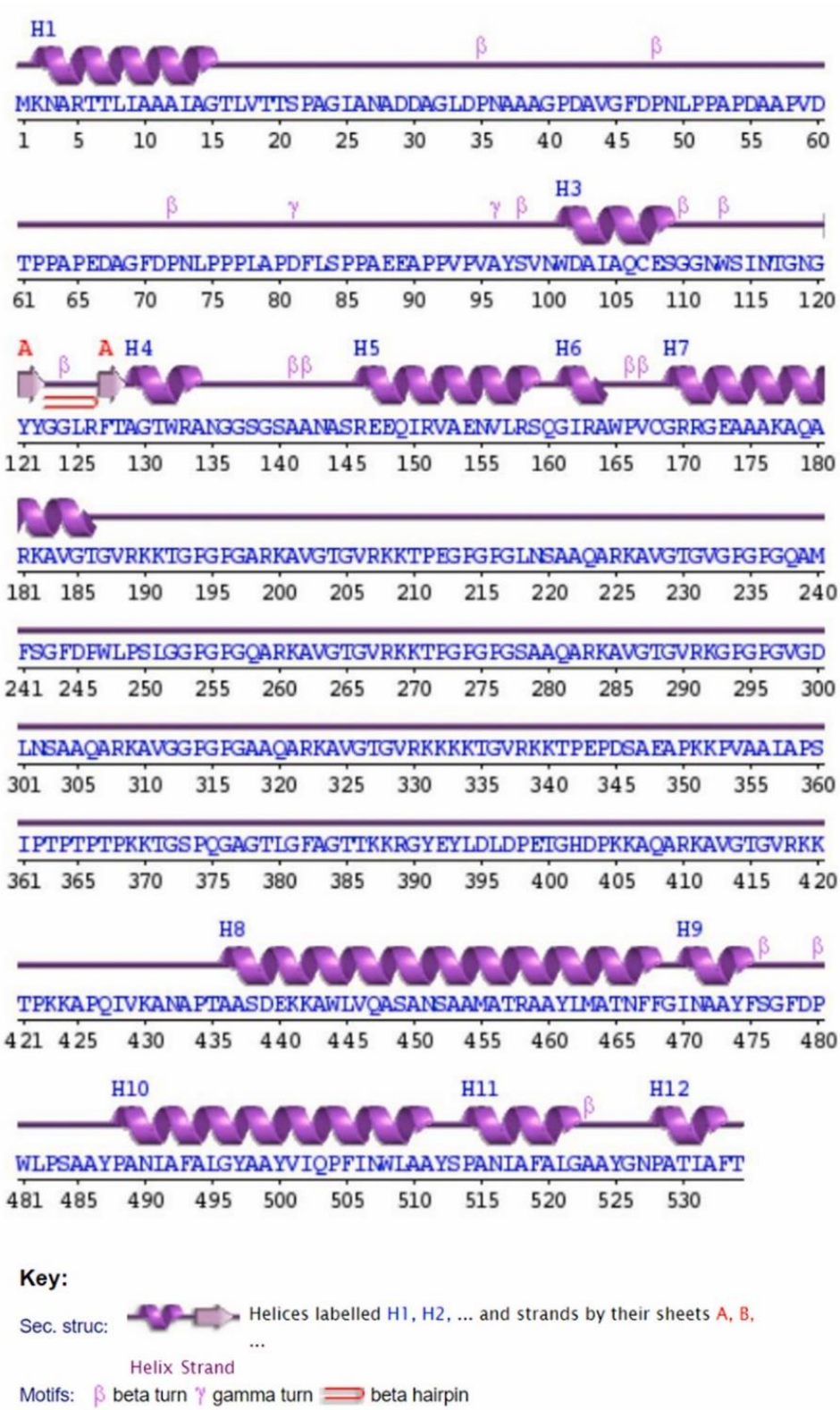

H1
β
β
MKNARTTLIAAAIAGTLVTTSPAGIANADDAGLDPNAAAGPDAVGFDPNLPPAPDAAPVD
1 5 10 15 20 25 30 35 40 45 50 55 60
β
γ
γ β
H3
β β
TPPAPEDAGFDPNLPPPLAPDFLSPPAEEAPPVPVAYSVNWDAIAQCESGGNWSINTGNG
61 65 70 75 80 85 90 95 100 105 110 115 120
A
β
A H4
ββ
H5
H6
ββ
H7
YYGGLRFTAGTWRANGGSGSAANASREEQIRVAENVLRSQGIRAWPVCGRRGEAAAKAQA
121 125 130 135 140 145 150 155 160 165 170 175 180
RKAVGTGVRKKTGPGPGARKAVGTGVRKKTPEGPGPGLNSAAQARKAVGTGVGPGPGQAM
181 185 190 195 200 205 210 215 220 225 230 235 240
FSGFDFWLPSLGGPGPGQARKAVGTGVRKKTPGPGPGSAAQARKAVGTGVRKGPGPGVGD
241 245 250 255 260 265 270 275 280 285 290 295 300
LNSAAQARKAVGGPGPGAAQARKAVGTGVRKKKKTGVRKKTPEPDSAEAPKKPVAAIAPS
301 305 310 315 320 325 330 335 340 345 350 355 360
IPTPTPTPKKTGSPQGAGTLGFAGTTKKRGYEYLDLDPEIGHDPKKAQARKAVGTGVRKK
361 365 370 375 380 385 390 395 400 405 410 415 420
H8
H9
β
β
TPKKAPQIVKANAPTAASDEKKAWLVQASANSAAMATRAAYLMATNFFGINAAYFSGFDP
421 425 430 435 440 445 450 455 460 465 470 475 480
H10
H11
β
H12
WLPSAAYPANIAFALGYAAYVIQPFINWLAAYSPANIAFALGAAYGNPATIAFT
481 485 490 495 500 505 510 515 520 525 530
Key:
Sec. struc:
Helices labelled H1, H2, ... and strands by their sheets A, B, ...
Helix Strand
Motifs: β beta turn γ gamma turn beta hairpin

Figure 4.3 Secondary structure prediction of the vaccine construct, representation in schematic "wiring diagram" including strands (pink arrows), helices (purple springs), and other motifs in red (e.g., β-hairpins, γ-turns, etc).

Subsequently, the 3D structure of the vaccine construct was predicted utilizing SWISS-MODEL, I-TASSER, and AlphaFold2. Firstly, when employing SWISS-MODEL, the value of Global Model Quality Estimate (GMQE) and QMEANDisCo Global of the two models were too low, only 0.25 and 0.09 for model 1 and model 2, respectively (Figure S1). As I know, GMQE and QMEANDisCo Global scores give an overall model quality measurement between 0 and 1, with higher numbers indicating higher expected quality. Therefore, I moved to the I-TASSER server to predict and acquire the top five final 3D models. Structure information, including C-score, estimated TM-score, estimated root-mean-square deviation (RMSD), number of structural decoys, and cluster density, of these 5 models is shown in Table S1. C-score is a confidence score for estimating the quality of predicted models. In general, the C-score is typically in the range of [-5, 2], where the higher value signifies the higher confidence of the model. TM-score and RMSD are usually used to measure the accuracy of structure modeling when the native structure is known. Based on that,

I chose model 1 (Figure S2A) with the C-score of −2.08, TM score of 0.47 ± 0.15, and estimated RMSD was 12.5 ± 4.3 Å. However, analyzing the Ramachandran plot of this model disclosed only 55.0% of residues in the most favored regions (Figure S2B). Hence, I conducted structural refinement using the GalaxyRefine server and got five refined models with the structure information displayed in Table S2. Refer to Table S2, model 2 demonstrated the highest global distance test-high accuracy (GDT-HA) score of 0.9148 (the higher value, the more accurate), and the lowest root mean square deviation (RMSD) score of 0.524 (lower value indicating greater stability) (Figure S2C). Nonetheless, when I applied the PROCHECK tool to validate the stereochemical quality of this model structure, there were only 76.0% of residues in the most favored regions (Figure S2D). A good quality model would be expected to have over 90% of residues in the most favored regions. For that reason, I could not use the prediction result obtained from I-TASSER. Consequently, I employed the deep learning approach of AlphaFold2 and predicted the top five 3D structure models of the vaccine via ColabFold v2.3.2 based on the local distance difference test (pLDDT) ranking (Figure S3). Among them, the 1st rank_model 2 with the best estimated reliability, was selected as the predicted structure (Figure S3A). Nevertheless, this

prediction yielded a pLDDT score below 50 (Figure S3B) and the Ramachandran plot analysis revealed only 44.2% of residues in most favored regions (Figure S3C), I performed further structural refinement employing the GalaxyRefine server which generated five refined models and their properties (Table 4.5).

Table 4.5 Structure information obtained from GalaxyWEB

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.593	27.4	7.7	45.3
MODEL 1	0.8095	0.862	1.211	1.9	0.3	96.2
MODEL 2	0.8071	0.885	1.295	2.3	0.0	95.9
MODEL 3	0.8062	0.882	1.454	2.8	0.0	94.4
MODEL 4	0.8038	0.883	1.376	2.2	0.3	94.4
MODEL 5	0.7949	0.911	1.214	1.4	0.3	95.1

As the result shows in Table 4.5, model 1 held the highest GDT-HA score of 0.8095, the lowest RMSD score of 0.862, and the lowest MolProbity score of 1.211 (lower MolProbity value indicates better model quality), and its 3D structure was represented in the cartoon in Figure 4.4A. Subsequently, to validate the quality of this 3D structure after the refining process, ProSA-web and PROCHECK tools were utilized. As shown in Figure 4.4B, the Z-score was determined to be -5.98. Notably, after refinement, the Ramachandran plot showed 92.8%, 5.0%, and 2.2% of residues were present in the favored, allowed, and disallowed regions, respectively (Figure 4.4C).

Additionally, corresponding quality scores assessed through QMEAN4 are presented in Figure 4.4D with a value of −5.58. Overview, model 1 (Figure 4.4A) was the best compared to other models and was selected as the vaccine candidate for further study, including molecular docking and simulations.

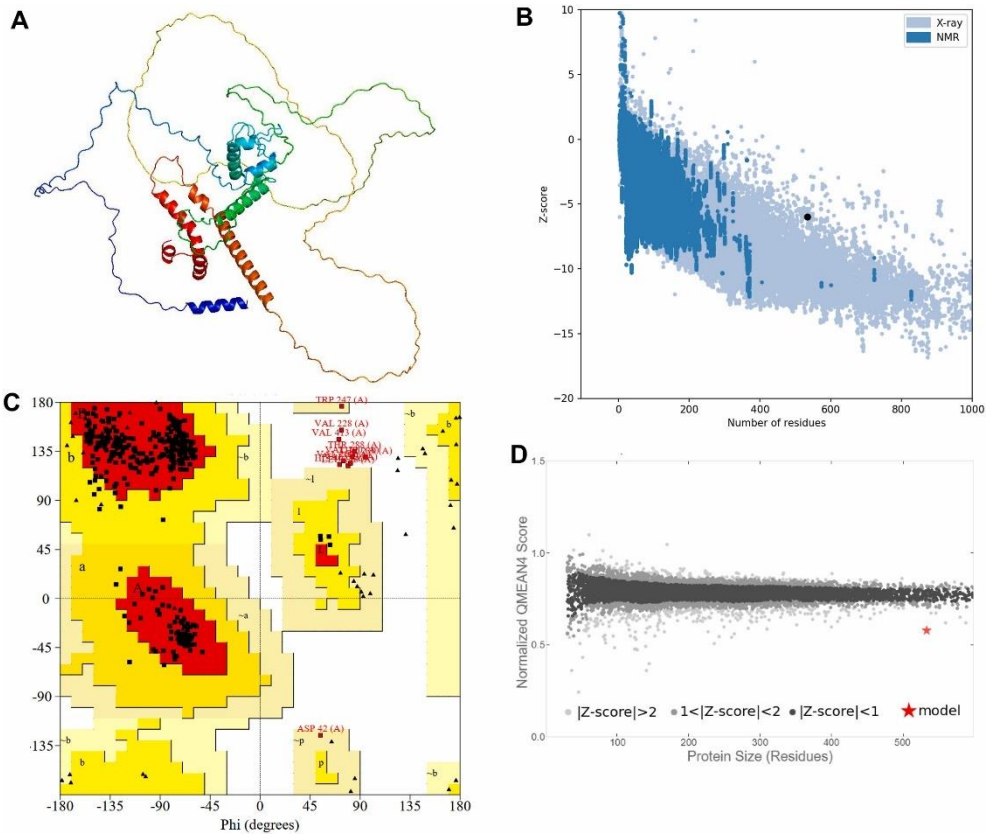


Figure 4.4 Prediction, refinement, and validation of the tertiary structure of the vaccine. (A) The 3D structure representation in the cartoon by PyMOL. (B) The Z-score was obtained from ProSA-Web. (C) Ramachandran plot gained from PROCHECK. (D) Normalized QMEAN score composed of four statistical potential terms (QMEAN4) of the vaccine.

4.4.6. Molecular docking of vaccine construct with immune receptors

The molecular docking was performed for two complex systems, MEV-TLR2 and MEV-TLR4, using the ClusPro server, which generated the top 30 models for each system. Among these models, the model with the lowest negative docking score was selected as the best-docked complex. Specifically, the models with -1376.3 kcal/mol (MEV-TLR2) and -1545.0 kcal/mol (MEV-TLR4) were selected for further analysis (Table 4.6).

Table 4.6 Molecular docking of the vaccine with TLR2 and TLR4

Target (PDB ID)	Center (kcal/mol)	Lowest energy (kcal/mol)		No. of interface residues	Interface area (Å^2)	No. of hydrogen bonds	No. of salt bridges
TLR2 (2Z7X)	-1079.1	-1376.3	Chain A	TLR2A: 8 Vaccine: 8	TLR2A: 414 Vaccine: 406	13	6
			Chain B	TLR2B: 46 Vaccine: 35	TLR2B: 1716 Vaccine: 1954	23	4
TLR4 (3FXI)	-1441.9	-1545.0	Chain A	TLR4A: 48 Vaccine: 46	TLR4A: 2298 Vaccine: 2278	44	8
			Chain B	TLR4B: 39 Vaccine: 30	TLR4B: 1486 Vaccine: 1641	24	5

A 3D representation of the surface and cartoon of the docking complex MEV-TLR2 was presented in Figures 4.5A and B, and MEV-TLR4 was shown in Figures 4.6A and B. In addition, interacting residues between MEV-TLR2 (Figures 4.5C–F) and MEV-TLR4 (Figures 4.6C–F) were visualized using PDBsum. Our results showed that 36 hydrogen bonds and 10 salt bridges were formed between the residues of MEV and two chains of TLR2 (Figures 4.5E and F). Similarly, 68 hydrogen bonds and 13 salt bridges were formed between the residues of the vaccine and two chains of TLR4 (Figures 4.6E and F). Based on these findings, MEV had excellent performance in strongly binding to TLR2 and 4 to produce a strong immune response.

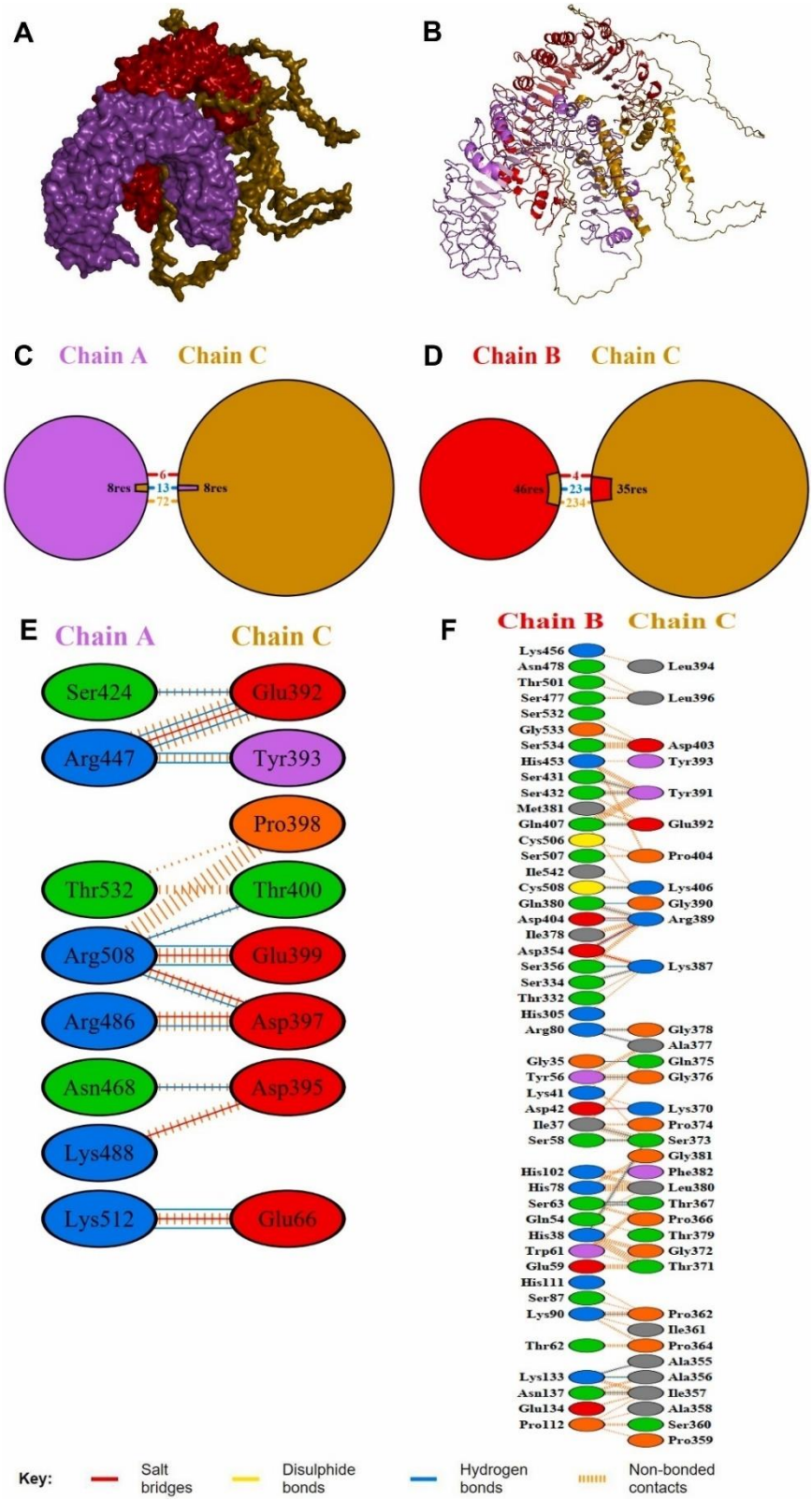

A
B
C
Chain A
Chain C
8res
6
13
72
8res
D
Chain B
Chain C
46res
4
23
234
35res
E
Chain A
Chain C
Ser424
Glu392
Arg447
Tyr393
Pro398
Thr532
Thr400
Arg508
Glu399
Arg486
Asp397
Asn468
Asp395
Lys488
Lys512
Glu66
F
Chain B
Chain C
Lys456
Asn478
Thr501
Ser477
Ser532
Gly533
Ser534
His453
Ser431
Ser432
Met381
Gln407
Cys506
Ser507
Ile542
Cys508
Gln380
Asp404
Ile378
Asp354
Ser356
Ser334
Thr332
His305
Arg80
Gly35
Tyr56
Lys41
Asp42
Ile37
Ser58
His102
His78
Ser63
Gln54
His38
Trp61
Glu59
His111
Ser87
Lys90
Thr62
Lys133
Asn137
Glu134
Pro112
Leu394
Leu396
Asp403
Tyr393
Tyr391
Glu392
Pro404
Lys406
Gly390
Arg389
Lys387
Gly378
Ala377
Gln375
Gly376
Lys370
Pro374
Ser373
Gly381
Phe382
Leu380
Thr367
Pro366
Thr379
Gly372
Thr371
Pro362
Ile361
Pro364
Ala355
Ala356
Ile357
Ala358
Ser360
Pro359
Key:
Salt bridges
Disulphide bonds
Hydrogen bonds
Non-bonded contacts

Figure 4.5 Molecular docking between TLR2 chains A (purple) and B (red) with the vaccine construct (chain C – dark yellow). (A–B) Three-dimensional representation of the docking complex in surface and cartoon, respectively. (C–D) Schematic diagram of interactions between TLR2 and the vaccine. (E–F) Residue interactions between TLR2 and the vaccine construct. Salt-bridges (red lines), hydrogen bonds (blue lines), and non-bonded contacts (orange dashed line) between residues on either side of the vaccine-receptor interface.

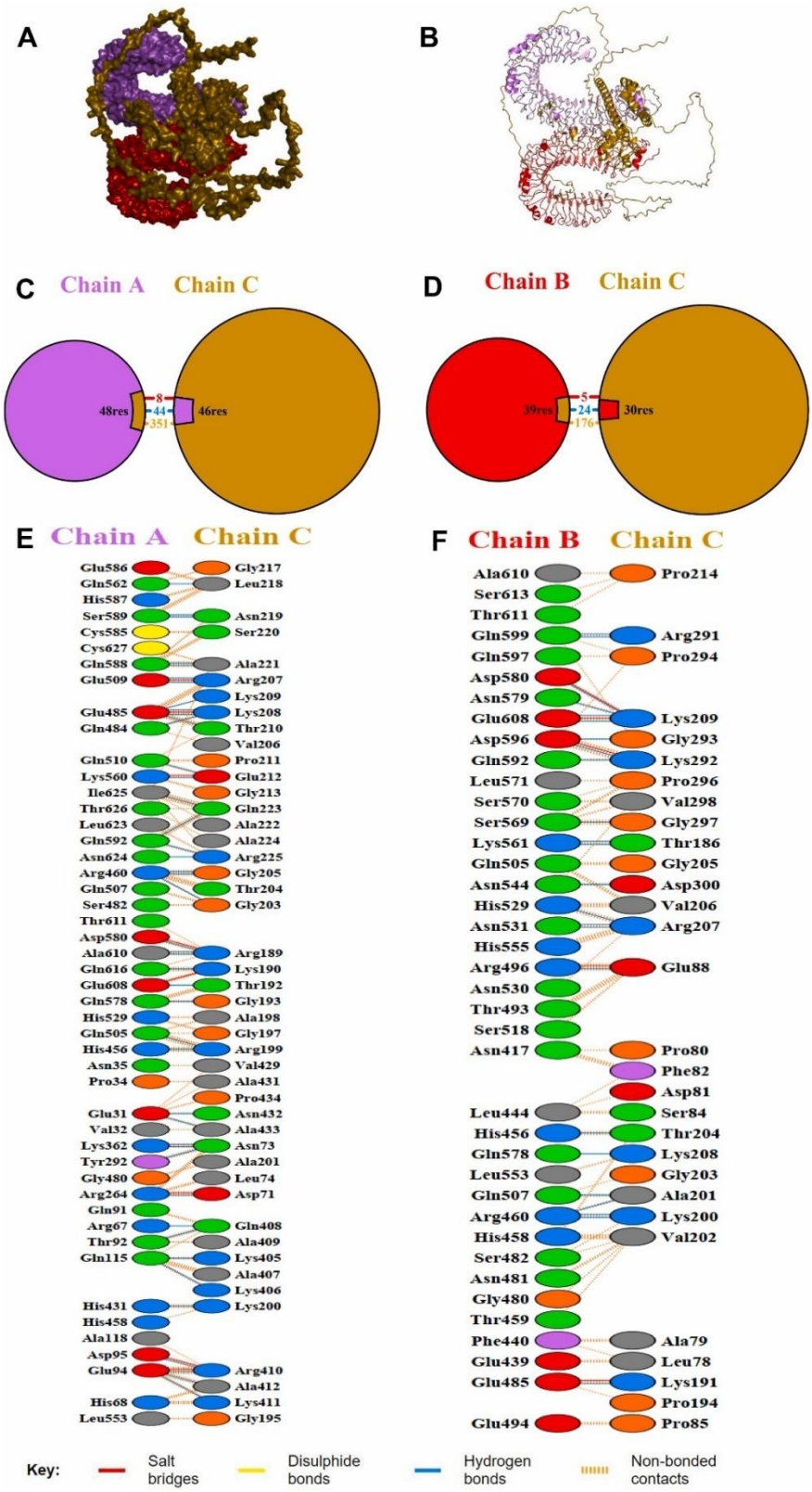

A
B
C
Chain A
Chain C
48res
8
44
351
46res
D
Chain B
Chain C
39res
5
24
176
30res
E
Chain A
Chain C
Glu586 Gly217
Gln562 Leu218
His587
Ser589 Asn219
Cys585 Ser220
Cys627
Gln588 Ala221
Glu509 Arg207
Lys209
Glu485 Lys208
Gln484 Thr210
Val206
Gln510 Pro211
Lys560 Glu212
Ile625 Gly213
Thr626 Gln223
Leu623 Ala222
Gln592 Ala224
Asn624 Arg225
Arg460 Gly205
Gln507 Thr204
Ser482 Gly203
Thr611
Asp580
Ala610 Arg189
Gln616 Lys190
Glu608 Thr192
Gln578 Gly193
His529 Ala198
Gln505 Gly197
His456 Arg199
Asn35 Val429
Pro34 Ala431
Pro434
Glu31 Asn432
Val32 Ala433
Lys362 Asn73
Tyr292 Ala201
Gly480 Leu74
Arg264 Asp71
Gln91
Arg67 Gln408
Thr92 Ala409
Gln115 Lys405
Ala407
Lys406
His431 Lys200
His458
Ala118
Asp95
Glu94 Arg410
Ala412
His68 Lys411
Leu553 Gly195
F
Chain B
Chain C
Ala610 Pro214
Ser613
Thr611
Gln599 Arg291
Gln597 Pro294
Asp580
Asn579
Glu608 Lys209
Asp596 Gly293
Gln592 Lys292
Leu571 Pro296
Ser570 Val298
Ser569 Gly297
Lys561 Thr186
Gln505 Gly205
Asn544 Asp300
His529 Val206
Asn531 Arg207
His555
Arg496 Glu88
Asn530
Thr493
Ser518
Asn417 Pro80
Phe82
Asp81
Leu444 Ser84
His456 Thr204
Gln578 Lys208
Leu553 Gly203
Gln507 Ala201
Arg460 Lys200
His458 Val202
Ser482
Asn481
Gly480
Thr459
Phe440 Ala79
Glu439 Leu78
Glu485 Lys191
Pro194
Glu494 Pro85
Key:
Salt bridges
Disulphide bonds
Hydrogen bonds
Non-bonded contacts

Figure 4.6 Molecular docking between TLR4 chains A (purple) and B (red) with the vaccine construct (chain C – dark yellow). (A–B) Three-dimensional representation of the docking complex in surface and cartoon, respectively. (C–D) Schematic diagram of interactions between TLR4 and the vaccine. (E–F) Residue interactions between TLR4 and the vaccine construct. Salt-bridges (red lines), hydrogen bonds (blue lines), and non-bonded contacts (orange dashed line) between residues on either side of the vaccine-receptor interface.

4.4.7. Molecular dynamic simulations of vaccine with immune receptors

In order to obtain the stability and dynamic behavior of the interactions between MEV with TLR2 and TLR4 receptors during 100 ns simulation, statistical parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), and solvent accessible surface area (SASA) were examined in triplicate (Figure 5.7). After 50 ns, the RMSD values for both remained stable, indicating that the complexes maintained a comparatively stable structure, according to Figures 4.7A and B, the complexes consistently retained moderate structural stability. During the simulation, the

average RMSD values of the MEV-TLR2 and MEV-TLR4 complexes were 1.767 ± 0.127 and 1.634 ± 0.151 nm, respectively.

Referring to the trajectories of Figures 4.7C and D, the MEV regions displayed the most significant fluctuations and an increasing trend in RMSF values, suggesting that the MEV regions were more dynamic or flexible than the TLRs regions. Besides, the Rg offers valuable information regarding the tendency of complex structures to expand during MDs. For the MEV-TLR2 complex (Figure 4.7E), the Rg values slightly increased from the start of the simulations until 10 ns, indicating compaction or tightening of the complex. The Rg values gradually decreased, then stable after 50 ns until they reached 100 ns. For the MEV-TLR4 complex (Figure 4.7F), the Rg gradually reduced from the beginning of the simulation up to 40 ns, then stabilized until it reached 100 ns. Throughout the simulation, the complexes maintained an overall relatively compact structure, as indicated by the average Rg value of 4.889 ± 0.187 and 4.663 ± 0.044 nm for MEV-TLR2 and MEV-TLR4, respectively.

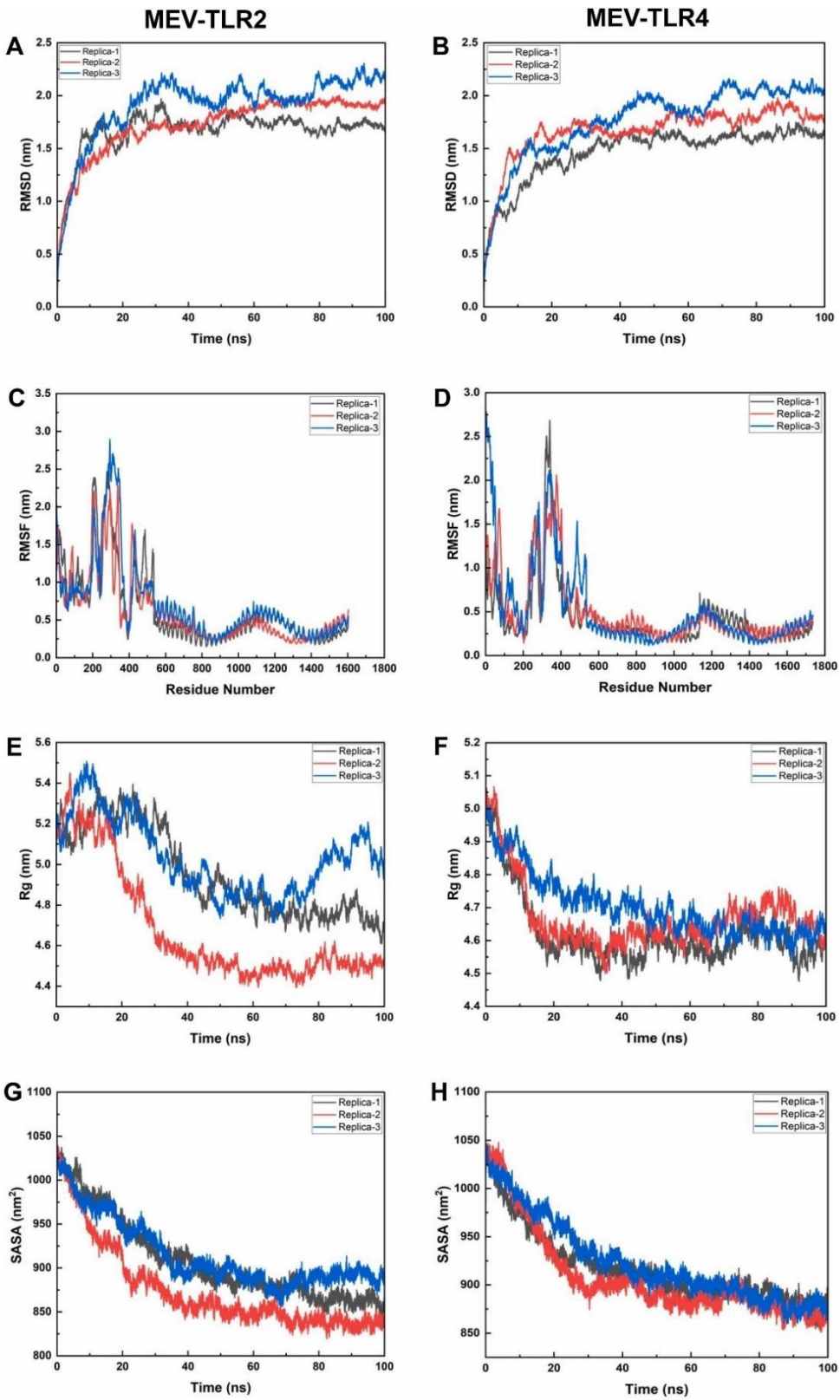

MEV-TLR2
MEV-TLR4
A
B
C
D
E
F
G
H
Replica-1
Replica-2
Replica-3
RMSD (nm)
RMSF (nm)
Rg (nm)
SASA (nm²)
Time (ns)
Residue Number

Figure 4.7 MD simulations of the docking complexes between MEV and TLR2/4. (A–B) Root mean square deviation plot for MEV-TLR2 and MEV-TLR4 complex, respectively. (C–D) Root mean square fluctuation of MEV-TLR2 and MEV-TLR4 complex, respectively. (E–F) Radius of gyration analysis between MEV-TLR2 and MEV-TLR4 complex, respectively. (G–H) Solvent-accessible surface area of MEV-TLR2 and MEV-TLR4 complex during MD simulations, respectively.

In addition, Figures 4.7G and H show the average SASA values of MEV-TLR2 and MEV-TLR4 complex were 889.117 ± 21.779 and 915.881 ± 7.935 nm^2, respectively. The SASA values gradually decreased throughout the simulation, indicating that the complexes got more compact or less exposed to solvent. Significant structural rearrangement may have occurred during the beginning of simulation until 40 ns (for MEV-TLR2) and 30 ns (MEV-TLR4) period, as evidenced by the significant and quick decrease in SASA during this time frame. After that, both were stable until the end of the simulation period.

4.4.8. Immune responses induced by the vaccine

The immunological simulation data was produced by the C-IMMSIMserver following three successive injections of the vaccine candidate. Figure 4.8A shows that a vaccine's initial exposure produces a relatively low immunoglobulin response, whereas a subsequent exposure produces an increased immunoglobulin response. Additionally, Figure 4.8A shows that immunoglobulins such as IgM and IgG are more abundant than other immunoglobulins such as IgM, IgG1, and IgG1 + IgG2. Details of the cytokine levels, especially IFN-g, increased significantly and were above 400,000 ng/ml, which are visualized in Figure 4.8B, while Figure 4.8C indicates responses of B lymphocytes. Our simulation study also identified helper T-cell and cytotoxic T-cell responses (Figures 4.8D and E), confirming the realistic character of the server-predicted immune response because helper T-cell activity is crucial for activating B-cells. During the self-memorization process following pathogen exposure, memory cells play a crucial role in preventing and regulating viral infection and reinfection. The successful injection of the vaccine candidate resulted in an increase in the regulatory components of the immune system, including DC cells, macrophages, NK cells, interleukins, and cytokines (Figures 4.8C-H). These results imply that the MEV is a highly

effective next-generation vaccine based on peptides that can stimulate a robust immune response against MTB infection.

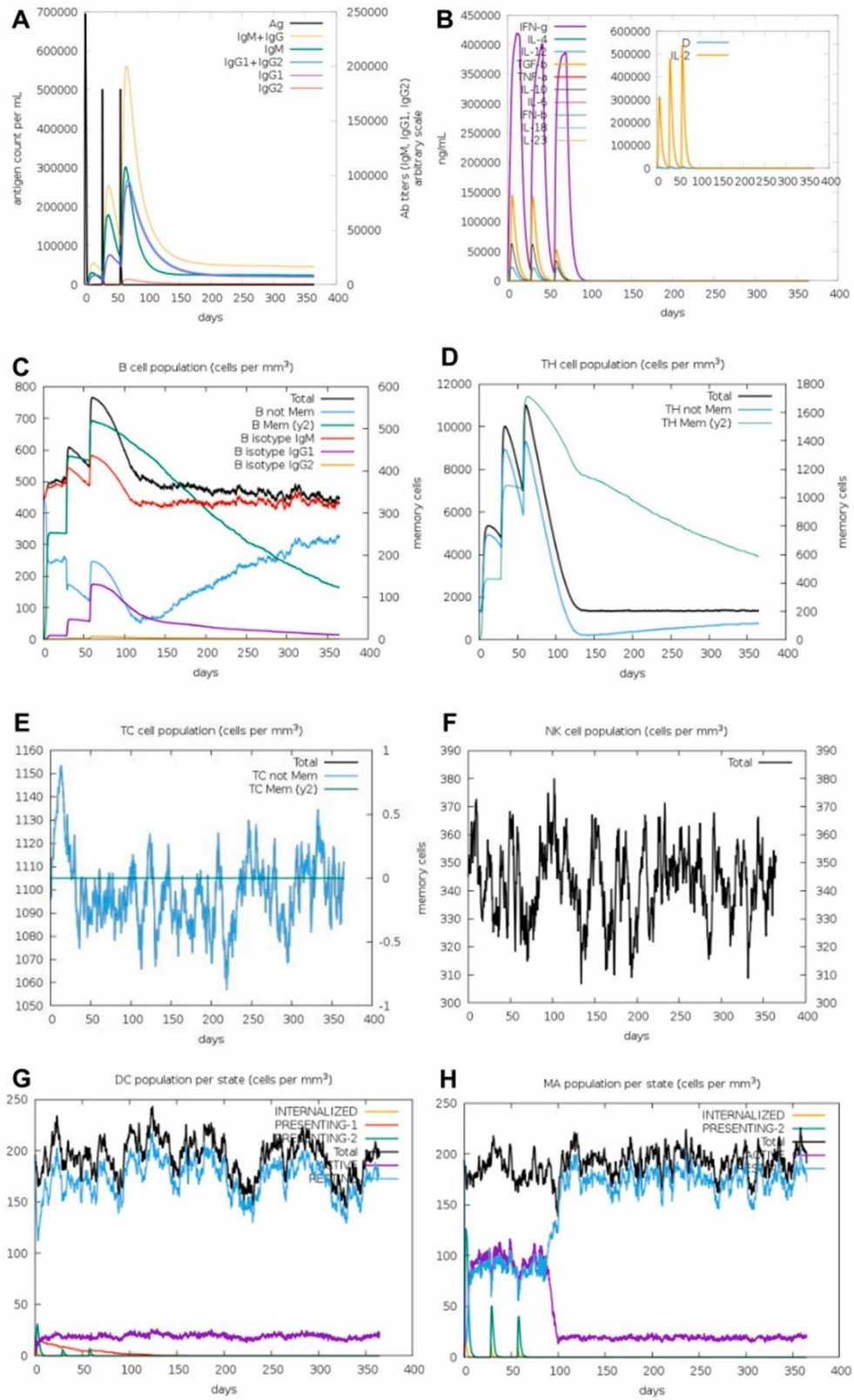

A
antigen count per mL
Ab titers (IgM, IgG1, IgG2) arbitrary scale
days
Ag
IgM+IgG
IgM
IgG1+IgG2
IgG1
IgG2
B
ng/mL
days
IFN-g
IL-4
IL-12
TGF-b
TNF-a
IL-10
IL-6
IFN-b
IL-18
IL23
D
IL-2
C
B cell population (cells per mm³)
memory cells
days
Total
B not Mem
B Mem (y2)
B isotype IgM
B isotype IgG1
B isotype IgG2
D
TH cell population (cells per mm³)
memory cells
days
Total
TH not Mem
TH Mem (y2)
E
TC cell population (cells per mm³)
memory cells
days
Total
TC not Mem
TC Mem (y2)
F
NK cell population (cells per mm³)
memory cells
days
Total
G
DC population per state (cells per mm³)
days
INTERNALIZED
PRESENTING-1
PRESENTING-2
Total
H
MA population per state (cells per mm³)
days
INTERNALIZED
PRESENTING-2
Total

Figure 4.8 The innate and adaptive immune responses induced by the MEV in the C-IMMSIMserver. (A) Immunoglobulin responses upon exposure to the vaccine. (B) Concentration of cytokines and interleukins. (C) B cell population. (D) T helper cell population. (E) T cytotoxic cell population. (F) Behavior of the population of Natural Killer cells. (G) Behavior of the population of Dendritic cells. (H) The population of macrophages after vaccination.

4.5. Discussion

Despite advances in vaccine technology, there are still no vaccines against some infectious diseases, including tuberculosis. The infectious pathogens underlying these diseases evade and alter host immune responses, making vaccine development difficult. In this study, I aim to design an MEV candidate against tuberculosis to contribute towards the End MTB Strategy. Several vaccines for MT have been developed to provide possible candidates for novel vaccine designs (Albutti, 2021b; Bibi et al., 2021b; Ruaro-Moreno et al., 2023; Sharma et al., 2021a). While a few new proteins with antigenic qualities were chosen for the current study to identify epitopes and develop the vaccine, most known antigenic proteins or proteins found in exosome vesicles were used in these previous studies to uncover antigenic

epitopes. Besides, although many vaccine candidates are available, each vaccine employs different algorithms and features. The in vitro and in vivo vaccination production process is far more complex, expensive, and time-consuming than the MEV. A range of laboratory medical studies are also required for the final epitope selection. A computational technique that saves time and predicts a peptide or epitope sequence that might be used to make a lab-based MEV is called in silico methodology.

The present study focuses on a PPE family protein, Rv0256c (PPE2), which induced a strong B cell response in tuberculosis patients (Abraham et al., 2014). I carried out a computer analysis using a range of immunoinformatics techniques to find 21 potent epitopes that could be helpful in the fight against tuberculosis. Therefore, using this method, I may reduce the cost and length of wet lab investigations. Finally, I speculate that the designed vaccine is extracellular, highly immunogenic, antigenic, nontoxic, and nonallergic; as such, it could be a promising MEV candidate for MTB based on computational analysis. Additional wet-lab validation is required to confirm the 21 epitopes' effectiveness as MEV in this study.

Firstly, I retrieved the protein sequence and evaluated antigenicity and allergenicity. Then, the immunoinformatic techniques

were used to screen and construct potential epitopes from the sequence of Rv256c protein for B and T cells. The antigenicity, allergenicity, and several physiochemical properties of the developed multi-epitope vaccination were then evaluated. The vaccine construct contains 534 amino acids, comprising six cytotoxic T lymphocyte, eight helper T lymphocyte, and seven linear B lymphocyte epitopes, along with adjuvants and linkers. The antigenicity score of MEV predicted by VaxiJen 2.0 and ANTIGENpro was 0.9363 and 0.9399, respectively. The MEV was predicted as stable and thermostable with an instability index and aliphatic index of 28.64 and 62.77, respectively. Afterwards, tertiary structure prediction, refinement, and validation were conducted. Ramachandran plot analysis reveals that 97.8% of the amino acid residues were in the most favored and allowed regions.

Subsequently, molecular docking and MD simulations were utilized in evaluating the complex stability. There were a total of 46 interaction sites in MEV-TLR2 and 81 interaction sites in MEV-TLR4 (Figures 4.5 and 4.6). These results showed that the interactions between MEV and TLR2, 4 were strong, and the docking effect was good, especially, the MEV-TLR 4 complex. Triplicate MD simulations were calculated to confirm the poses found by docking results. The MD calculation lets us establish if MD finds the most populated cluster

from docking. The static view provided by docking should be verified by using MD. In this study, an MD simulation could be helpful to confirm if the primary contacts found will be maintained during the MD to present more reliable results. In general, combining two in silico techniques (docking, MD) could improve the reliability of the results. I need to ensure that all the system's chemical and physical properties have reached an equilibrium where their averages no longer change as a function of time. A simple way to test this is by measuring the RMSD of the backbone concerning the start (Figures 4.7A and B). For the RMSD, the average is taken over the particles, giving time-specific values; for RMSF, the latter is averaged over time, giving each particle (residue) value. RMSF is a simple tool to measure the rigidity of the polypeptide chain. It calculates the deviations of the C-alpha atom's coordinates from their average position. The flexibility pattern reflects the location of secondary structure elements in the protein structure (Figures 4.7C and D). Besides, I studied the compactness of the receptors TLR2 and TLR4 interaction with the vaccine using Rg. The receptor remains compact, and no unusual folding or unfolding was observed throughout the 100 ns. The vaccine with TLR4 is more tightly packed than TLR2 (Figures 4.7E and F). Furthermore, I calculated the total solvent-accessible surface area to understand the

system's SASA of the binding region. I observed that initially, the area of structures contact by the solvent molecules was higher, and it decreased gradually with time and showed stable values, showing significant interaction can be assumed between the vaccine and TLR2, 4. In particular, the complex between the vaccine and TLR4 showed more stability than TLR2 (Figures 4.7G and H). To sum up, our molecular docking studies and triplicate MD simulations revealed superior interactions and stability of the vaccine with the TLR4 complex compared to the TLR2 complex. This aligns with the outcomes of a previous investigation (Alderwick et al., 2015). Alderwick et al. in that study, demonstrated that human immune cells that were infected with MTB express more TLR4, which played a role in the interaction with MTB and activated TLR4 related signaling, which in turn enhanced Th2 signaling and led to the development of tuberculosis disease.

Finally, I performed in silico immune simulation to characterize the immunogenicity and immune response of the vaccine. Figure 4.8 shows the outcomes of the immunological simulation after the MEV was administered. After the host immune system was exposed to MEV several times, there was a discernible rise in secondary and tertiary antibody levels, which were higher than primary antibody detection

levels. As a result, there was a quick drop in antigen concentration, which suggests quick clearance (Figure 4.8A). Interestingly, T-cells were found to have increased initially but then somewhat decreased. Moreover, cytokine levels, especially IFN-g, increased significantly and were above 400,000 ng/ml (Figure 4.8B). Figures 4.8C–H show the distribution of immune cell populations in different states. These immune cell populations showed a notable overall increase, exemplified by the maturation of memory cells. These results demonstrate that the vaccine designs can provide immunity against the MTB and strongly suggest the formation of immunological memory.

This chapter was published in *Infectious Disease Modelling*

as a partial fulfillment of Nguyen Thi Truc Ly's Ph.D program.

Chapter 5 Integrating immunoinformatics and computational epitope prediction for a vaccine candidate against respiratory syncytial virus

5.1. Abstract

Respiratory syncytial virus poses a significant global health threat, especially affecting infants and the elderly. Addressing this, the present study proposes an innovative approach to vaccine design, utilizing immunoinformatics and computational strategies. I analyzed respiratory syncytial virus's structural proteins across both subtypes A and B, identifying potential helper T lymphocyte, cytotoxic T lymphocyte, and linear B lymphocyte epitopes. Criteria such as antigenicity, allergenicity, toxicity, and cytokine-inducing potential were rigorously examined. Additionally, I evaluated the conservancy of these epitopes and their population coverage across various respiratory syncytial virus strains. The comprehensive analysis identified six major histocompatibility complex class I binding, five major histocompatibility complex class II binding, and three B-cell epitopes. These were integrated with suitable linkers and adjuvants to form the vaccine. Further, molecular docking and molecular dynamics simulations demonstrated stable interactions between the vaccine candidate and human toll-like receptors 4 and 5, with a notable preference for TLR4. Immune simulation analysis underscored the vaccine's potential to elicit a strong immune response. This study presents a promising respiratory syncytial virus vaccine candidate and

offers theoretical support, marking a significant advancement in vaccine development efforts. However, the promising in silico findings need to be further validated through additional in vivo studies.

5.2. Introduction

Respiratory syncytial virus (RSV) is a significant viral pathogen, particularly affecting the respiratory system, and is a leading cause of upper and lower respiratory tract infections, especially in infants and young children globally (Broor et al., 2018). Furthermore, RSV infects individuals with weakened immune systems or chronic lung/heart diseases and the elderly, where it can exacerbate the underlying diseases and account for the development of asthma, chronic obstructive pulmonary disease, and congestive heart failure (Falsey et al., 2005; Nam & Ison, 2019). First identified in 1956, RSV has since garnered attention due to its widespread prevalence, causing seasonal outbreaks and contributing to substantial morbidity and mortality worldwide (Borchers et al., 2013; Morris et al., 1956). RSV is ubiquitous and triggers outbreaks of respiratory infections, which tend to peak during the winter in temperate regions and the rainy season in tropical areas (Nam & Ison, 2019; Obando-Pacheco et al., 2018). The virus spreads through respiratory droplets, making it highly contagious

(Bergeron & Tripp, 2021). Virtually all children will have been infected with RSV by the age of two, and re-infections are common throughout life (Ruckwardt et al., 2019). In infants, RSV infections can cause crepitation, chest wall indrawing, hypoxemia, wheezing, and tachypnea, whereas, in adults, it can aggravate the underlying cardiopulmonary diseases (Broor et al., 2018; Ruckwardt et al., 2019). Premature birth, congenital heart disease, neuromuscular diseases, bronchopulmonary dysplasia, and male gender are the major risk factors of RSV infection (Messina et al., 2022). Among infants, RSV stands as the primary cause of hospitalizations worldwide and the second most prevalent cause of mortality in low- and middle-income nations. On a global scale, it has been approximated that RSV gives rise to 33 million new cases of acute lower respiratory tract infection in children under the age of five, leading to around 3 million hospitalizations and 120,000 deaths each year (Colosia et al., 2023; Mejias et al., 2019). In adults who contract RSV infection, mild cold-like symptoms are typical, and a few cases may progress to pneumonia or lung infection. Annually, approximately 60,000 to 160,000 older adults in the United States are hospitalized due to RSV infection, resulting in 6000 to 10,000 deaths (Harris, 2023).

While RSV infections are often mild and self-limiting, severe cases can lead to significant morbidity and mortality, particularly in

infants and older adults. Due to the health risks and mortality linked to RSV, there has been a persistent demand for effective therapeutics for either treating or preventing RSV infections and associated illnesses (Colosia et al., 2023). The SARS-CoV-2 pandemic has heightened the awareness within the scientific community to prepare for highly contagious human viruses proactively. Therefore, it is imperative to explore innovative strategies for developing potential therapies against RSV, aiming to avert potential calamities in the future. Hence, in this study, the immunoinformatics and computational approaches have been integrated to develop a potential vaccine candidate against RSV. Firstly, the structural proteins from both subtypes of RSV were employed to identify antigenic B-cell and T-cell epitopes. Subsequently, the epitopes' properties were predicted, including allergic potential, antigenic potential, toxic property, and their capacity to elicit interferon-gamma (IFN-γ) and interleukin-4 (IL4). Then, the conservancy of these epitopes was determined in different RSV strains belonging to subtype A and subtype B of RSV, followed by the population coverage analysis of the T-cell epitopes. Afterward, chosen epitopes were linked with adjuvants and linkers to formulate an RSV vaccine candidate, for which I predicted physicochemical properties, stability, antigenicity, toxicity, and allergenicity. Following

this, the vaccine candidate's tertiary structure was expected, then molecular docking and molecular dynamics simulation studies were conducted to elucidate its interactions with immune cells. Finally, an immune simulation study was carried out to assess how the designed RSV vaccine triggers immune responses in humans at various dosages. This research contributes to the ongoing efforts to address the challenges posed by RSV and underscores the potential of immunoinformatics in advancing vaccine design against respiratory pathogens.

5.3. Materials and Methods

5.3.1. Protein sequence retrieval and epitope prediction

Protein sequences of RSV structural proteins were retrieved from the UniProt database (Consortium, 2022). T-helper cell epitopes, capable of binding to MHC class II molecules, were predicted using the NetMHCII 2.3 server (Jensen et al., 2018). Similarly, T-cytotoxic cell epitopes with potential for MHC class I molecule binding were predicted using the NetMHCpan 4.0 web server (Andreatta & Nielsen, 2016). Protein sequences were input in FASTA format into the NetMHCpan 4.0 and NetMHCII 2.3 servers, with chosen peptide

lengths of 9 and 15 (default) respectively. Default parameters of the NetMHCpan 4.0 and NetMHCII 2.3 servers were applied to establish thresholds for strong and weak binders. For predicting multiple linear B-cell epitopes, the BepiPred 2.0 server was employed (Larsen et al., 2006). Subsequently, the antigenicity of predicted B-cell and T-cell epitopes was determined using the VaxiJen v2.0 web server (Irini A. Doytchinova & Darren R. Flower, 2007). The epitope's allergic potential, toxicity, and IFN-γ activation potential were assessed using the AllergenFP v1.0, ToxinPred, and IFNepitope web servers, respectively (Dhanda, Vir, et al., 2013b; Dimitrov et al., 2013; Gupta, Kapoor, Chaudhary, Gautam, Kumar, Open Source Drug Discovery, et al., 2013). Additionally, the IL4pred server was employed to evaluate the epitopes' capability to induce interleukin-4 (IL4) production (Dhanda, Gupta, et al., 2013b).

5.3.2. Conservancy analysis and population coverage analysis

The study aims to develop a vaccine candidate against both the subtypes of RSV. Hence, the conservancy of the predicted epitopes that were antigenic, non-toxic, non-allergic, and induced cytokine generation was performed by employing the Immune Epitope Database

and Analysis Resources (IEDB) tool (Bui et al., 2007). Within this tool, epitope and protein sequences from various RSV strains of both subtypes were inputted in FASTA format, and opting for the default settings for all other parameters. The IEDB population coverage analysis tool was utilized to ascertain population coverage for the selected T-cell epitopes in the design of the vaccine candidate (Bui et al., 2006). Notably, determining population coverage analysis for the final B-cell epitopes proved challenging due to the absence of web servers or software capable of predicting B-cell epitope population coverage. The IEDB population coverage analysis tool employed default values for the "number of epitopes" and "query by" parameters. The selection "World" was made for "select area(s) and population(s)", and the combined Class I and II options were chosen under the "select calculation option".

5.3.3. Vaccine candidate engineering and physiochemical properties prediction

The selected B-cell and T-cell epitopes were conjugated with flagellin, RS09 adjuvants, and the PADRE (Pan HLA DR-binding epitope) sequence through the use of GGS linkers, forming the final vaccine

construct. The GGS linker, chosen for its flexibility and biocompatibility, facilitated the integration of adjuvants and epitopes, enhancing the vaccine's immunogenic potential. The Expasy ProtParam tool was used to evaluate the vaccine construct's physicochemical properties, such as molecular weight, isoelectric point, aliphatic index, and stability. Antigenicity was predicted using the VaxiJen v2.0 tool, assessing the vaccine's potential to elicit an immune response (Irini A. Doytchinova & Darren R. Flower, 2007; Gasteiger et al., 2005).

5.3.4. Prediction, refinement, and validation of the tertiary structure of the vaccine candidate

The tertiary structure of the vaccine candidate was predicted using AlphaFold2's deep learning algorithm through ColabFold v1.5.5 (Jumper et al., 2021), with subsequent structural refinement performed on the GalaxyWEB server (Ko et al., 2012). Model quality and validation were thoroughly assessed using ProSA-web (Wiederstein & Sippl, 2007), PROCHECK program (Laskowski et al., 1993), and QMEAN4 (Benkert et al., 2010; Chawla et al., 2023), ensuring a comprehensive evaluation of the vaccine candidate's structure.

5.3.5. Molecular docking studies between the vaccine and immune receptors

The structure of the human TLR4 (UniProt ID: O00206) and TLR5 receptor (UniProt ID: O60602) were predicted by AlphaFold2. For both receptors, I retained only the extracellular domain encompassing amino acids 30–624 of TLR4 and 21–639 of TLR5, while excluding other regions. The 3D structure of the multi-epitope vaccine (MEV) candidate and immune receptors were docked using the ClusPro server (Kozakov et al., 2017b). ClusPro conducts molecular docking analysis through a multi-step process that includes rigid body docking, clustering of low-energy structures, and energy minimization. Initially, it employs rigid body docking, where billions of conformations are sampled. Subsequently, the 1000 lowest energy structures are grouped to identify the largest clusters, based on root-mean-square deviation (RMSD). Finally, energy minimization is conducted to eliminate steric clashes, refining the docking results (Truc Ly Nguyen & Heebal Kim, 2024b). Accordingly, the complex displaying the lowest binding energy (in kcal/mol) was selected for visualization. PDBsum was subsequently employed to analyze and identify interacting residues between the vaccine and TLR4/TLR5 receptors (Laskowski et al., 2018).

5.3.6. Molecular dynamics simulations of the vaccine-receptor complex

To investigate the stability of the vaccine-receptor complex, all-atom molecular dynamics (MD) simulations were performed using the GROMACS 2023 software on a Linux operating system employing the CHARMM27 force field (Abraham et al., 2015). The docking complex MEV-TLR4 was solvated in a cubic box (12 × 12 × 12) using the SPCE water model, surrounded by 325,909 solvent molecules. Subsequently, to neutralize the charge, 24 Na^+ ions were added. The system underwent energy minimization, using the steepest descent algorithm with 50,000 steps, and the minimization process ceased when the maximum force reached <1000.0 kJ/mol/nm. Following that, position restraints were applied during the equilibration process. NVT equilibration was executed at 300 K with 50,000 steps (100 ps), followed by NPT equilibration at 1 bar reference pressure with an additional 50,000 steps (100 ps). Afterward, a production simulation for all-atom (995,112 atoms) was conducted using the NPT ensemble for 50,000,000 steps (100 ns). Upon completing the 100 ns MD simulation, I computed the root mean square deviation (RMSD) of backbone residues, root mean square fluctuation (RMSF) of C-alpha, the solvent-accessible surface area (SASA) and the buried surface

area (BSA) for the system. To validate result accuracy and reliability, three replicate MD simulations were performed over 100 ns with a different starting velocity (Akhtar et al., 2022; Kaushik, G, et al., 2022). Additionally, superimpositions of the docking complex were created using selected snapshots from the MD simulations. Finally, the COCOMAPS tool (Vangone et al., 2011) was employed to examine interface connectivity between the MEV and TLR4 through specific snapshots.

5.3.7. Computational immune simulation of RSV vaccine candidate

An in silico immune simulation was performed using the C-IMMSIM online tool, applying default settings with the exception of the time step adjustment, to explore the immune response elicited by the RSV vaccine construct (Rapin et al., 2011). Although the recommended interval between two vaccine doses is typically four weeks, intervals ranging from 8 weeks to 6 months can also be considered, depending on the instance (Castiglione et al., 2012; Robinson et al., 2017). Accordingly, to assess the immune response to the RSV vaccine construct, three doses were administered at four-week intervals. The

time steps used in the simulation were 1, 84 (representing 4 weeks), and 168 (corresponding to 8 weeks).

5.4. Results

5.4.1. Protein sequence retrieval and epitope prediction

The UniProt IDs of the structural proteins of RSV strain A2 belonging to subtype A have been provided in Table 5.1. All retrieved proteins were verified at the protein level, each receiving a top annotation score of 5/5, with the exception of the M2-2 protein. A total of 749 strong-binding MHC-I epitopes, 1621 high-affinity MHC-II epitopes, and 36 linear B-cell epitopes were predicted from the RSV structural proteins (Supplementary Data Sheets). Then, these epitopes underwent further analysis for antigenicity (VaxiJen score above 0.4), allergenicity, toxicity, and ability to induce IFN-γ and IL-4. These comprehensive analyses led to the identification of 22 MHC-I epitopes and 72 MHC-II epitopes, with only 6 B-cell epitopes meeting the specified criteria.

Table 5.1 UniProt IDs of the structural proteins of RSV strain A2

Protein	UniProt ID	Identification method	Annotation score
N: Nucleoprotein	P03418	Evidence at protein level	5/5
P: Phosphoprotein	P03421	Evidence at protein level	5/5
M: Matrix protein	P0DOE7	Evidence at protein level	5/5
SH: Small hydrophobic protein	P0DOE5	Evidence at protein level	5/5
G: Major surface glycoprotein	P03423	Evidence at protein level	5/5
F: Fusion glycoprotein	P03420	Evidence at protein level	5/5
M2-1	P04545	Evidence at protein level	5/5
M2-2	P88812	Inferred from homology	2/5
L: RNA-directed RNA polymerase	P28887	Evidence at protein level	5/5

5.4.2. Conservancy analysis and population coverage analysis

Epitope conservancy was assessed across the S2 strain (A subtype), 9320 strain (B subtype), and B1 strain (B subtype) RSV strains. Sequences of all the structural proteins are available in UniProt for only these RSV strains, whereas other strains have only 1 or 2 protein sequences documented. Among the 22 MHC-I epitopes, 14 showed conservancy in all the selected strains. Likewise, 31 of the 72 MHC-II epitopes were conserved across these RSV strains. Three of the 6 B-cell epitopes were conserved in the tested RSV strains. To narrow down the epitope selection for the vaccine construct, a further parameter - a VaxiJen score above 1.1 - was applied. Consequently, 6 MHC-I and 5 MHC-II binding epitopes were selected for the final

construct, achieving a global population coverage of 52.07%, as detailed in Table S1.

5.4.3. Vaccine candidate engineering and physiochemical properties prediction

Following comprehensive analysis in the earlier sections, 6 MHC-I binding epitopes, 5 MHC-II binding epitopes, and 3 B-cell epitopes were shortlisted for the vaccine design (Table 5.2). Each selected epitope is antigenic, non-allergic, non-toxic, capable of inducing IFN-γ and IL-4, and conserved in all tested strains. These selected epitopes were then integrated with flagellin adjuvant, RS09 adjuvant, and PADRE sequence using appropriate linkers GGS, as outlined in the methodology, to form an RSV vaccine candidate composed of 545 amino acids (Table S2). Additionally, Table S3 presents the physicochemical parameters of the vaccine candidate, which is characterized by its antigenicity, non-allergenicity, non-toxicity, and stability.

Table 5.2 Epitopes for the design of the final RSV vaccine candidate

Protein	MHC type	Epitopes	Binding MHC alleles	VaxiJen score[a]
Nucleoprotein	MHC-I	EMKFEVLTL	HLA-B*08:01	1.1076
Matrix protein		EKDDDPASL	HLA-B*39:01	1.343
Fusion glycoprotein		ASISQVNEK	HLA-A*03:01	1.3852
RNA-directed RNA polymerase		DIRYIYRSL	HLA-B*08:01	1.3203
		ELEYRGESL	HLA-B*08:01	1.6945
		YHAQDDIDF	HLA-B*39:01	1.3715
Nucleoprotein	MHC-II	VFVHFGIAQSSTRGG	DRB1_0901	1.1723
Fusion glycoprotein		TDRGWYCDNAGSVSF	DRB3_0101	1.1909
RNA-directed RNA polymerase		TVVELHPDIRYIYRS	DRB1_0301	1.1211
		PWVVNIDYHPTHMKA	DRB1_0301	1.2262
		CPWVVNIDYHPTHMK	DRB1_0301	1.4633
Phosphoprotein	B-cell	ETFDNNEEESSYSYEEI NDQTNDNI	Not applicable	0.5306
RNA-directed RNA polymerase		ISNKSNRYNDNYN	Not applicable	0.8783
		SRPCEFPASIPAYRT	Not applicable	0.4767

[a]VaxiJen score above 1.1 for MHC epitopes and 0.4 for B-cell epitopes

5.4.4. Prediction, refinement, and validation of the tertiary structure of the vaccine candidate

The multi-epitope vaccine (MEV) construct was predicted using the cutting-edge AlphaFold2 via ColabFold v1.5.5. Given that, the AlphaFold2 prediction yielded a low score for specific residues (Figure S1), and the calculated scores for Ramachandran plots yielded 60.4% residues in core acceptable region, 35.6% in allowed and generously allowed regions and 4.0% residues in disallowed regions (Figure S2). To enhance the model's quality, I pursued further structural refinement

using the GalaxyWEB server. Among models 1–5, model 1 (Figure 5.1A) was determined to be the most accurately refined. The results in terms of improvement of model 1 over initial input models for backbone structure accuracy measured by GDT-HA, side-chain structure accuracy measured by RMSD, and physical correctness measured by MolProbity score were summarized in Table S4. Furthermore, to validate the model's quality, I calculated Z-Score with a value of −6.8 (Figure 5.1B), analyzed the Ramachandran plot (Figure 5.1C), and determined QMEAN4 value of −4.27 (Figure 5.1D). According to the Ramachandran plot, 94.5% of residues were situated in the favorable core region, 5.1% in the allowed and generously allowed regions, and only 0.4% in the disallowed region. Hence, the refined MEV model (Figure 5.1A) demonstrates a high quality, making it suitable for further studies.

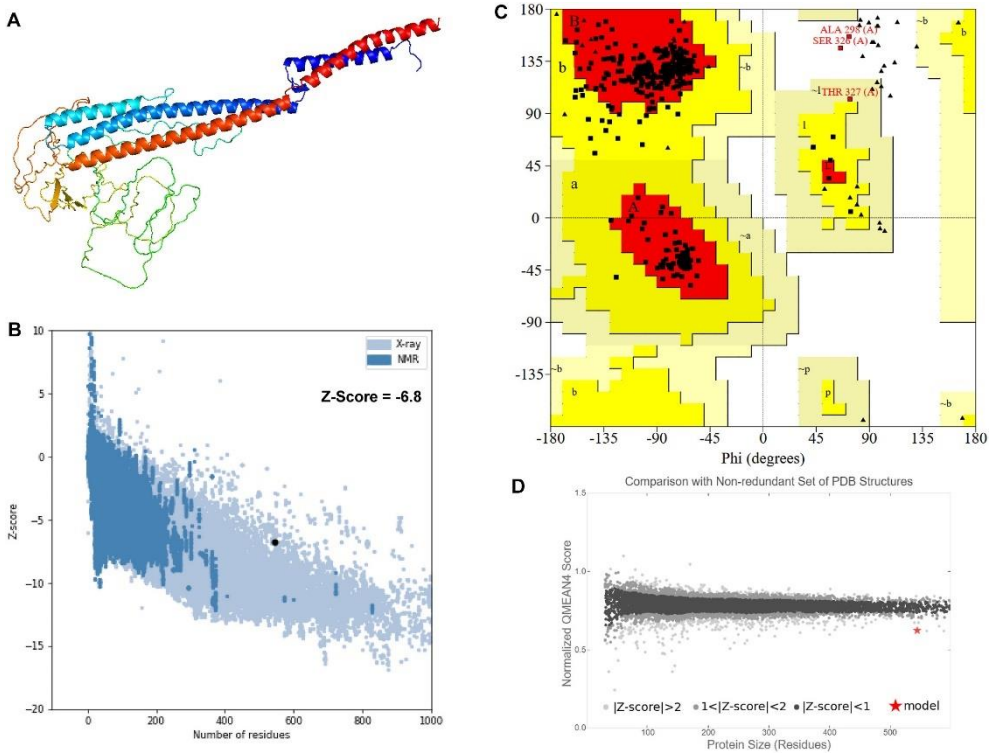


Figure 5.1 Tertiary structure of the vaccine construct. (A) Refined model representation in a cartoon by the GalaxyWEB server. (B) Z-Score was obtained from the ProSA-web of the refined model. (C) Ramachandran plot using PROCHECK program. (D) Normalized QMEAN score composed of four statistical potential terms (QMEAN4) of the vaccine construct.

5.4.5. Molecular docking of the vaccine candidate with immune TLR4 and TLR5 receptors

The tertiary structure of human TLR4 (UniProt ID: O00206) and TLR5 receptor (UniProt ID: O60602) were determined using AlphaFold2. The yield of high pLDDT values with confidence scores exceeding 90 for most residues indicated a high prediction confidence level for TLR4 (Figure S3) and TLR5 (Figure S4). The molecular docking was performed for two complex systems, MEV-TLR4 and MEV-TLR5, using the ClusPro server, which generated the top 30 models for each system. From these models, the ones with the lowest negative docking score were selected as the best-docked complexes. Specifically, the models with −1257.4 kcal/mol (MEV-TLR4) and −1542.7 kcal/mol (MEV-TLR5) were chosen for further analysis. Interactions between MEV and TLR4/TLR5 were illustrated in Figure 5.2, visualized using the PDBsum database with default threshold values of 3.5 Å. Analysis revealed the formation of 60 hydrogen bonds and 8 salt bridges between MEV and TLR4 (Figure 5.2A and Table S5), while MEV and TLR5 demonstrated only 19 hydrogen bonds (Figure 5.2B and Table S6). These findings indicate MEV's strong binding affinity to TLR4, suggesting its potential to elicit an immune response.

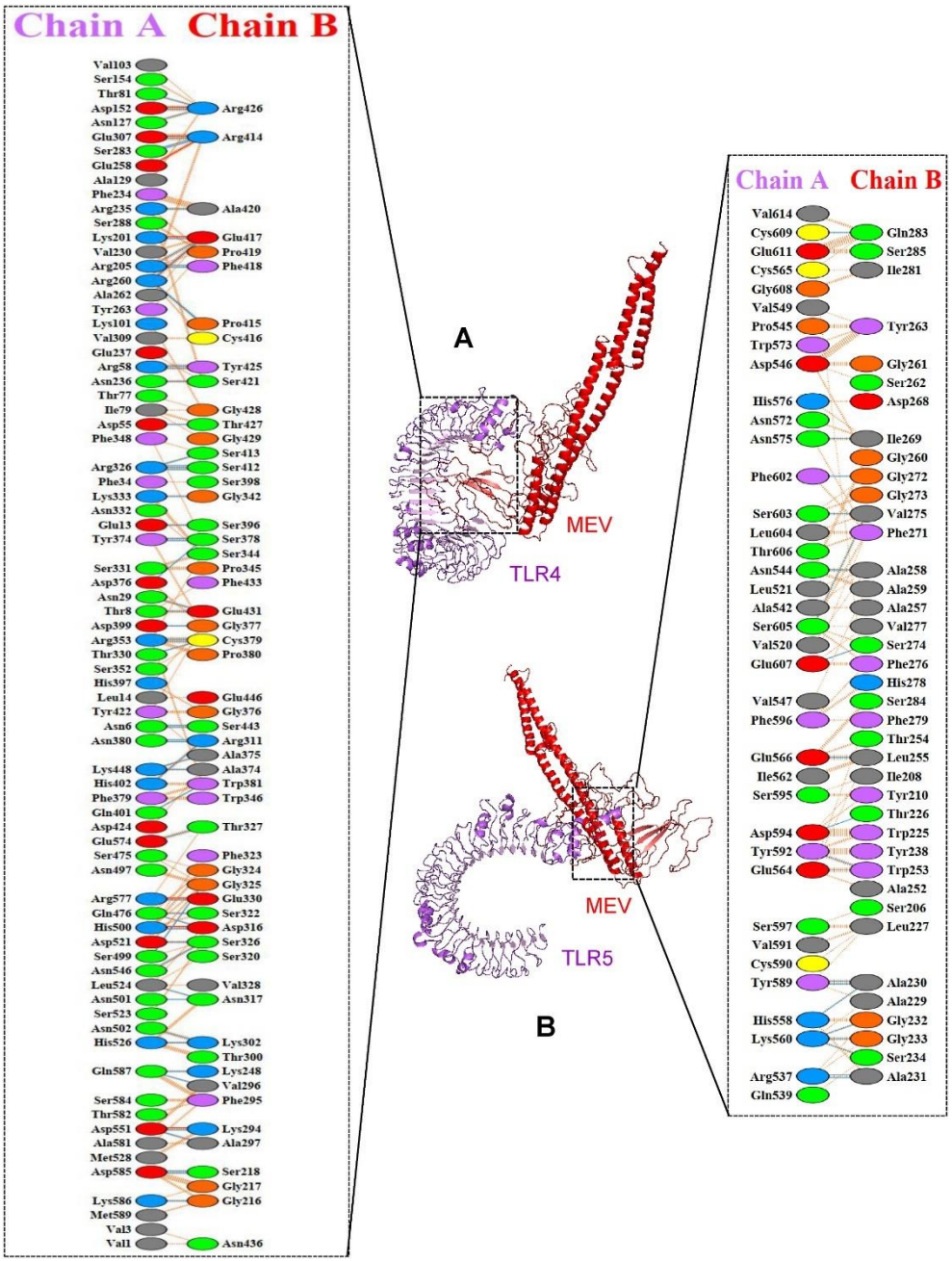


Figure 5.2 Molecular docking between the MEV (red) and TLR4/TLR5 (purple). (A) Docking complex representation in cartoon and residue interactions between TLR4 (chain A) and MEV (chain B). (B) Docking complex representation in cartoon and residue interactions between

TLR5 (chain A) and MEV (chain B). Salt bridges (red lines), hydrogen bonds (blue lines), and nonbonded contacts (orange dashed line) between residues on either side of the vaccine–receptor interface.

5.4.6. Molecular dynamics simulation studies

Following the molecular docking outcomes, the MEV–TLR4 complex, identified as the optimal docking configuration, was selected for MD simulation studies. Three replicates of MD simulations 100 ns were conducted for all-atom to investigate the stability of the vaccine-receptor complex using GROMACS 2023 software on a Linux operating system, which provides real-life environmental conditions for various biological models (T. L. Nguyen & H. Kim, 2024). RMSD from the backbone of the complex showed an average value of 0.84 ± 0.16 nm, achieving stability after 20 ns simulation time (Figure 5.3A). Besides, the RMSF was quantified to evaluate the flexibility across the amino acid residues within the complex. Figure 5.3B shows RMSF values of 0.53 ± 0.32 nm for the vaccine (residues 1 to 545) and 0.34 ± 0.08 nm for TLR4 (residues 546 to 1140), indicating the vaccine exhibits greater flexibility than TLR4. Significant fluctuations were observed particularly in the N- and C-terminal regions of flagellin within the vaccine. Another measure of docking complex MEV–TLR4 behavior is

the solvent-accessible surface area (SASA). The SASA is governed by the interactions (or lack of) of hydrophobic and hydrophilic amino acids with water. Figure 5.3C presents an average SASA value of 591.27 ± 16.35 nm^2, with the complex achieving a steady state after 40 ns and maintaining it throughout the simulation. Moreover, the buried surface area (BSA) at the MEV-TLR4 interface remained stable throughout the simulation, suggesting consistent interface interactions with a value of 45.61 ± 9.99 nm^2 (Figure 5.3D).

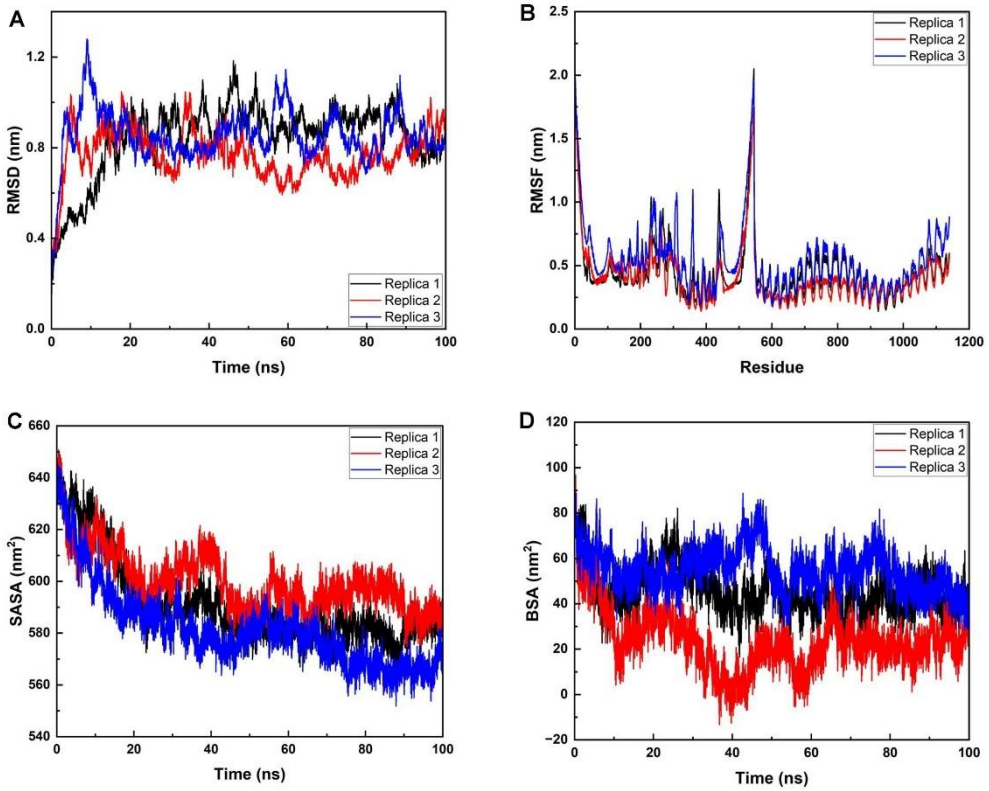


Figure 5.3 Molecular dynamics simulation of the vaccine and TLR4 complex. (A) Root mean square deviation (RMSD). (B) Root mean square fluctuation (RMSF). (C) Solvent-accessible surface area (SASA). (D) Buried surface area (BSA).

To further assess the complex's stability throughout the simulation, I calculated the RMSD of the MEV-TLR4 complex at various time steps. By comparing the stability of the MEV-TLR4 complex, I was able to determine their respective stability (Figure

5.4A). Subsequently, the COCOMAPS tool was employed to thoroughly study and visualize the contact points at the MEV–TLR4 interface (Figure 5.4B). Using intermolecular contact maps to find hot spot residues, COCOMAPS makes it possible to analyze and visualize the interaction interface in protein complexes (Vangone et al., 2011). While the study is presented for the chosen snapshots (Figure 5.4A), where the interaction pattern with the contacts is maintained, it is clear from Figure 5.4B that overall contacts remained steady (Akhtar et al., 2023; Kaushik, Jain, et al., 2022).

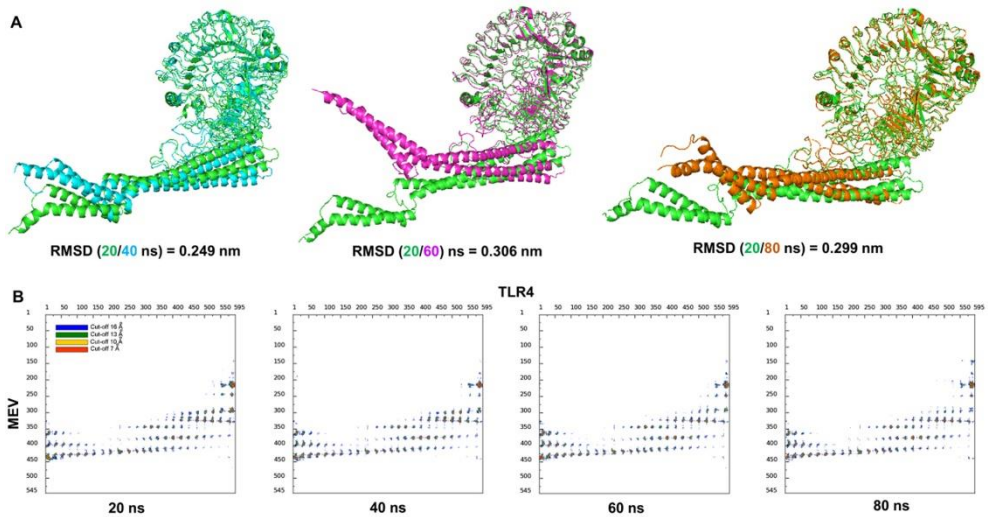


Figure 5.4 Stability of the docking complex MEV-TLR4. (A) Superimposition of selected snapshots of MEV-TLR4 and their respective RMSD values. (B) Contact maps showing the conservation of contacts between residues in MEV and TLR4.

Remarkably, the interface area and percentage of polar to non-polar residues at the interface of modeled complexes were found to be nearly the same across all instances (Table 5.3). Besides, across all model complexes, employing a cut-off distance of 5 Å to define two residues in contact, the distribution of hydrophilic-hydrophobic, hydrophilic-hydrophilic, hydrophobic-hydrophobic, and hydrogen bonds exhibited minimal variation in the four selected structures (Table 5.3). All the above analyses demonstrate that the interaction

patterns between TLR4 and MEV maintained stability throughout the simulation period.

Table 5.3 The stability of interface interactions between MEV and TLR4

Modeled complex	Interface area (A^2)	Polar residue (%)	Nonpolar residue (%)	Hydrophilic-Hydrophobic contacts	Hydrophilic-Hydrophilic contacts	Hydrophobic-Hydrophobic contacts	H-bonds
20 ns	2767.55	58.48	41.52	129	105	28	32
40 ns	2572.0	59.14	40.86	128	98	30	29
60 ns	2604.95	57.54	42.46	140	99	26	30
80 ns	2407.7	57.35	42.65	128	86	22	29

5.4.7. Computational immune simulation of RSV vaccine candidate

A computational immune simulation was conducted to assess the immune responses elicited by the designed RSV vaccine. Figure 5.5 displays the immune response patterns predicted from the computational analysis of the vaccine. Analysis of the initial dose (Figure 5.5A) compared to subsequent doses reveals a significant increase in antibody concentrations, encompassing IgM + IgG, IgM, IgG1 + IgG2, and IgG1. This signifies that immunization with the candidate vaccine instigates an enhanced antibody response. Additionally, successive vaccine doses result in an increase in the total B-cell population, B-memory cell populations, and B isotype IgM population, underscoring the stimulation of a robust secondary immune response (Figure 5.5B). Following each vaccination, there is an increase in the plasma B lymphocyte population as well (Figure 5.5C). Furthermore, the active TH cell population increases after each immunization (Figure 5.5D). However, the duplicating and resting TH cell population increases up to the second dose of the vaccine but decreases after the third dose of vaccination (Figure 5.5D). Moreover, the vaccine candidate stimulates the production of various cytokines,

including IFN-γ, interleukin-10 (IL10), interleukin-12 (IL12), and transforming growth factor-beta (TGF-β) (Figure 5.5E). Compared to the initial dose, the second dose of the vaccine results in an increased population of IFN-γ, IL-10, TGF-β, and IL-12. Following the third dose of the vaccine construct, there is an overall decrease in the concentration of different cytokines and interleukins compared to the first and second doses (Figure 5.5E).

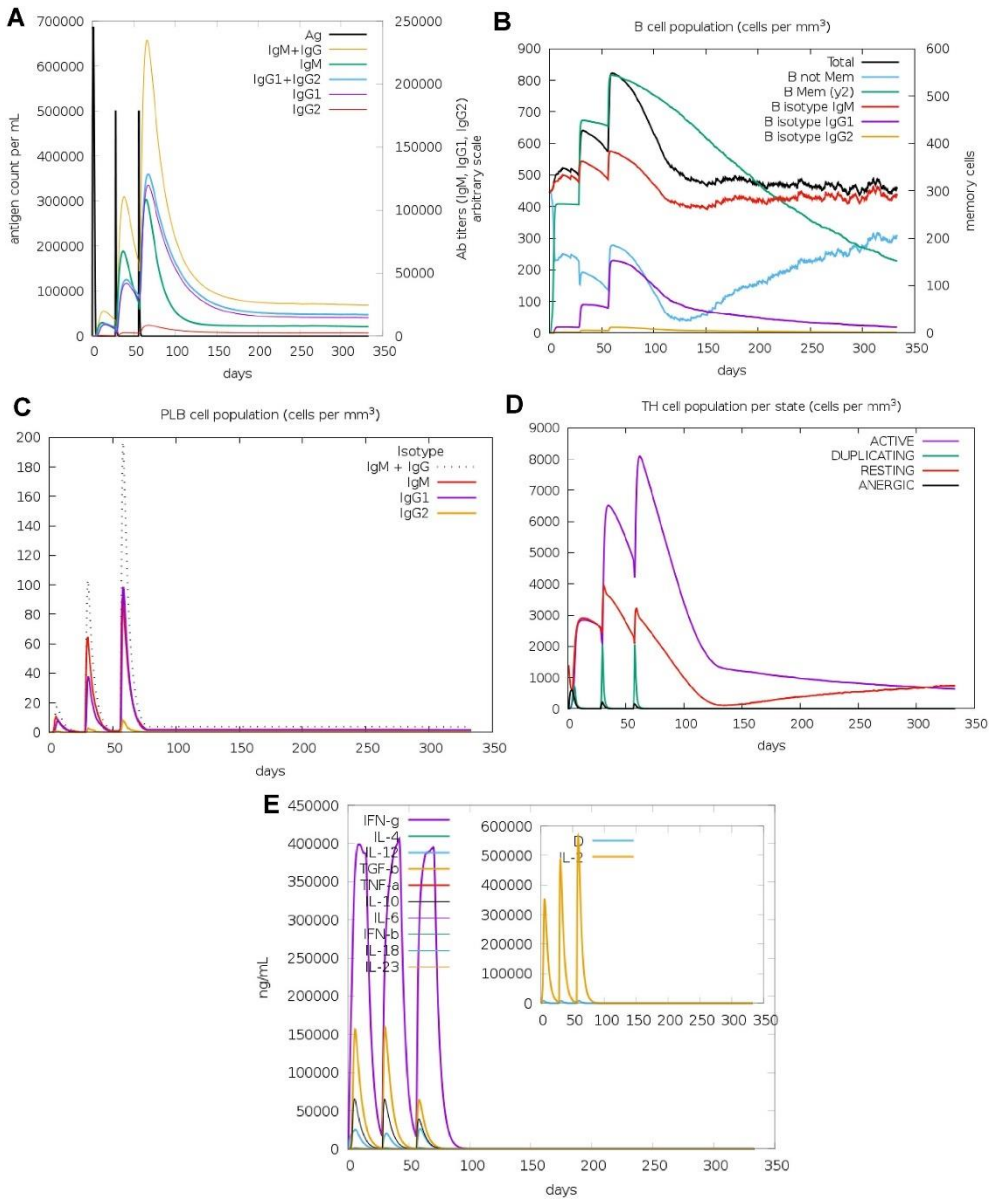


Figure 5.5 Immune simulation of the RSV vaccine construct. (A) Antigen and immunoglobulins. Antibodies are sub-divided per isotype. (B) B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM, IgG1 and IgG2. (C) Plasma B lymphocytes count sub-

divided per isotype (IgM, IgG1 and IgG2). (D) CD4+ T-helper lymphocytes count sub-divided per entity-state (i.e., active, resting, anergic, and duplicating). (E) Concentration of cytokines and interleukins. D in the inset plot is danger signal.

5.5. Discussion

RSV is an enveloped RNA virus belonging to the Paramyxoviridae family. There are two antigenically distinct RSV subtypes: A and B. These subtypes co-circulate within the same season, with one prevailing over the other. While specific studies indicate that subtype A is linked to heightened disease severity, others suggest that subtype B may have higher severity or that both subtypes exhibit comparable severity (Ciarlitto et al., 2019; Laham et al., 2017; Shen et al., 2022; Vandini et al., 2017). RSV possesses a single-stranded, negative-sense RNA genome, encoding for 11 proteins of which three are non-structural (NS1, NS2, M2–2), and 8 are structural proteins namely major surface glycoprotein (G), fusion glycoprotein (F), small hydrophobic protein (SH), nucleoprotein (N), phosphoprotein (P), matrix protein (M), M2-1 and RNA-directed RNA polymerase (L) (Kiss et al., 2014). Two major surface glycoproteins, the F and G proteins, play crucial roles in viral entry and fusion with host cells. The F protein

is particularly interested in vaccine development due to its role in entry into host cells, high conservation in RSV A and B subtypes, and the induction of protective immune responses. The G protein also bears antigenic determinants that trigger the production of neutralizing antibodies (Mejias et al., 2019). The L, N, M2-1, and P proteins envelop the viral RNA, creating a helical assembly known as the ribonucleoprotein complex, which protects the RNA (Cosentino et al., 2022; Kiss et al., 2014). The M2-1 protein helps regulate RSV organization and transcription process (Cosentino et al., 2022; Kiss et al., 2014). The M protein envelops the inner surface of the viral membrane, forming a protective layer around the viral genomic material (Conley et al., 2022). The deletion of the SH protein has slowed apoptosis in the infected cells and caused attenuation of the virus, suggesting the role of the protein in the pathogenesis of RSV (Li et al., 2015). Given these structural proteins' role in RSV's infection and pathogenesis, my study targeted them to identify epitopes for the potential vaccine design against RSV.

The US Food and Drug Administration (FDA) approved using the antiviral ribavirin in the aerosolized form for treating RSV infection in 1998. However, ribavirin offers limited clinical advantages in RSV and is not commonly prescribed as a routine treatment (Colosia et al.,

2023; Mir et al., 2021). Similarly, the FDA approved the monoclonal antibody palivizumab to prevent RSV-associated severe lower respiratory tract infections in pediatric patients in 1996 (Garegnani et al., 2021). Presently, extensive research, encompassing clinical trials, is underway to develop treatments or preventive measures for RSV. One of the extensively researched areas for developing therapies against RSV is the area of vaccine development. Different vaccine candidates, such as subunit-based, live attenuated, nucleic acid, and vector-based, have been explored to protect infants, children, and elderly patients (Bergeron & Tripp, 2021; Ruckwardt, 2023; Topalidou et al., 2023). Currently, the FDA has licensed two vaccines, Abrysvo and Arexvy, developed by Pfizer and GSK, respectively, and the 24 other candidates are under clinical trial stages (Topalidou et al., 2023). Furthermore, different groups have also explored the immunoinformatics approach to design a vaccine candidate against RSV (Dar et al., 2022; Moin et al., 2023; Naqvi et al., 2021; Tahir Ul Qamar et al., 2020). Nevertheless, my study's methodology, tools, and targeted proteins differ from those in previous research. Dar et al. focused on RSV's F and G proteins to predict antigenic epitopes (Dar et al., 2022). Moin et al. targeted the N, P, F, and G proteins to predict T-cell and B-cell epitopes for making vaccine candidates against both

RSV subtypes. Still, the computational tools utilized in my study differ from those in Moin et al.'s paper (Moin et al., 2023). Similarly, Naqvi et al. targeted the F, G, and SH proteins to predict only T-cell epitopes for the RSV vaccine candidate design (Naqvi et al., 2021). Additionally, Tahir et al. targeted the RSV F and G proteins to design a vaccine candidate following a methodology different from mine (Tahir Ul Qamar et al., 2020).

This study's immunoinformatics analysis identified 6 MHC-I, 5 MHC-II, and 3 B-cell epitopes within RSV's structural proteins. These selected epitopes were found to be antigenic, non-toxic, conserved among both the subtypes of RSV, non-allergic, and elicited the generation of IFN-γ and IL-4. The induction of the IFN-γ and IL-4 by the epitopes suggests the potential role of the epitopes in thwarting viral replication and activation of innate and adaptive immune activities (Cordeiro et al., 2022; Dittmer et al., 2001). The selected epitopes' conservancy suggests that my vaccine candidate could offer effective protection against both the RSV subtypes. Flagellin protein, RS09, and PADRE adjuvants were implemented to enhance the vaccine construct's efficacy. RS09 functions as an agonist of TLR4, while the flagellin protein acts as a TLR5 agonist (Forstnerič et al., 2017; Gupta et al., 2014; Shanmugam et al., 2012). The TLR agonists play a crucial

role in activating innate and adaptive immunity. Incorporating PADRE into the MEV can elevate its immunogenicity and efficacy (Ma et al., 2020). PADRE exhibits a high-affinity binding ability to various MHC class II molecules, facilitating the generation of antigen-specific $CD4^+$ T-cell responses (Ghaffari-Nazari et al., 2015). Additionally, it has also been reported to elicit $CD8^+$ T-cell responses (Ma et al., 2020). This attribute has led to its widespread utilization in the construction of MEVs. Proteins under 110 kDa are considered suitable for vaccine candidates (Shen et al., 2022), and my RSV vaccine's molecular weight is 57.22 kDa, affirming its appropriateness. Furthermore, the designed vaccine exhibits no homology with human proteins, thereby minimizing the risk of autoimmune responses in the vaccine recipients. The final vaccine protein displays an instability index of 38.74, suggesting stability in biological conditions, as compounds with an instability index below 40 are considered stable.

Moreover, molecular docking and dynamics simulations were performed to understand the MEV's interactions with TLR4 and TLR5, confirming these interactions' stability. Significantly, my results showed stronger and more stable interactions between the MEV and TLR4 than with TLR5. This aligns with the outcomes of previous investigations (Marzec et al., 2019; Monick et al., 2003). In that study,

Monick et al. demonstrated that TLR4 was more highly expressed and found in the membrane when lung epithelial cells (primary and transformed cell lines) were infected with RSV (Monick et al., 2003). Meanwhile, J. Marzec et al. suggested that TLR4 played a role in the pathogenesis of pulmonary RSV and the activation of cellular immunity caused by the inflammasome complex and vascular damage. Additionally, in line with computational immune simulation, the designed RSV vaccine construct is anticipated to have the potential to induce robust immune responses in recipients.

Chapter 6 General Discussion

This thesis aimed to design and evaluate multi-epitope vaccine candidates to combat infectious diseases using bioinformatics approaches. The computational techniques applied across the three case studies provided valuable insights into the potential of multi-epitope vaccines to induce robust immune responses. For *Mycobacterium tuberculosis* (MTB), the selected Rv0256c protein epitopes demonstrated strong antigenicity and stability, offering a promising approach to addressing other strains. Similarly, the vaccines designed for respiratory syncytial virus (RSV) and Powassan virus showed immunogenic potential through predicted interactions with immune receptors.

By adapting a common computational framework, this thesis highlights the potential of multi-epitope vaccines to address a wide range of global health challenges with precision and efficiency. The computational approaches employed consistently predicted highly antigenic, non-allergenic, and non-toxic epitopes across pathogens. The combination of T-cell and B-cell epitopes, along with appropriate adjuvants and linkers, contributed to the strong immunogenic profiles of these multi-epitope vaccines.

Additionally, one key finding was the broad applicability of multi-epitope vaccine design to both viral and bacterial pathogens.

The RSV and Powassan virus studies highlighted the importance of viral proteins in epitope selection, while the MTB study emphasized the value of intracellular antigens for effective immune targeting. This demonstrated the flexibility of bioinformatics tools in generating diverse vaccine candidates across a wide spectrum of infectious diseases.

The use of bioinformatics approaches played a pivotal role in the success of these studies. Immunoinformatics tools allowed for the rapid identification and evaluation of epitopes, which would have been time-consuming and resource-intensive through experimental methods alone. Tools such as IEDB, NetMHCIIpan, NetMHCpan, and BepiPred facilitated the efficient screening of large antigen datasets to identify the most promising candidates for vaccine development. Computational approaches, including molecular modeling, molecular docking, and molecular dynamics simulations, provided critical insights into the designed vaccine's 3D structures and their interactions with immune receptors such as Toll-like receptors (TLRs). The GROMACS simulations confirmed the vaccine's structural stability and flexibility under physiological conditions and revealed stable interactions between the vaccine and immune receptors. As bioinformatics tools continue to evolve, they will have a profound impact on clinical and

public health applications. The ability to rapidly design safe and effective vaccines will be essential in tackling future pandemics, positioning this research at the forefront of global health innovation.

Despite the promising results, several limitations must be acknowledged. First, the predictions made through computational tools require validation through experimental methods, including in vitro and in vivo testing. While in silico immune simulations indicated strong immune responses, actual biological systems may present unforeseen challenges, such as immune evasion by pathogens or unexpected toxicity. Additionally, the selection of epitopes depended on available databases and computational predictions, which might miss epitopes crucial for specific populations or contexts. Broader population coverage analysis and further refinement of epitope selection criteria will be necessary to address these limitations.

However, the impact of this research extends well beyond the specific vaccines developed in these studies. The computational framework presented here provides a flexible and scalable approach that can be applied to other infectious diseases, especially those with high antigenic variability, such as influenza and HIV. Quickly identifying and evaluating epitope candidates enables multi-epitope vaccine platforms to be rapidly adapted for emerging and re-emerging

pathogens, making this strategy highly relevant to pandemic preparedness.

Future research should focus on the experimental validation of the vaccine constructs developed in this thesis. In vitro studies are needed to confirm the antigenicity and immune-stimulating capacity of the selected epitopes, while in vivo studies will be critical for assessing the vaccines' safety and efficacy in animal models. Clinical trials would then be the next step, particularly for vaccine candidates that show promise in preclinical testing. Additionally, further refinement of the bioinformatics pipeline could enhance the precision of epitope prediction and vaccine design. The integration of artificial intelligence and machine learning tools into immunoinformatics could improve the accuracy of epitope selection, particularly for novel pathogens or less-studied antigens. Expanding the range of computational tools to include more advanced immune simulation platforms may also improve predictions of vaccine-induced immune responses.

In conclusion, this dissertation focuses on the design and evaluation of multi-epitope vaccine candidates to combat infectious diseases. The research integrates advanced bioinformatics techniques to predict and validate epitopes, design and optimize vaccine

constructs, and evaluate their potential efficacy and safety. The present study utilizes a comprehensive scientific approach, including immunoinformatics and computational vaccinology, to develop novel vaccine designs targeting the Powassan virus, *Mycobacterium tuberculosis*, and respiratory syncytial virus. The dissertation addresses the limitations of traditional vaccine development by utilizing computationally optimized epitopes with high immunogenic potential. The designs employed in the study effectively activate both cellular and humoral immune responses while minimizing the risk of adverse effects compared to single-epitope or whole-pathogen vaccines, as they include only the most immunogenic regions of antigens. The findings demonstrate significant advancements in the area of computational vaccine design, offering broad implications for addressing global health challenges.

References

A., A. (2021). Bioinformatics characterization of Plasmodium knowlesi apical membrane antigen 1 (PkAMA1) for multi-epitope vaccine design. *Tropical biomedicine*, *38*(3), 265-275. https://doi.org/10.47665/tb.38.3.067

Abbas, A. K., Lichtman, A. H., Pillai, S., Baker, D. L., & Baker, A. (2021). *Cellular and Molecular Immunology*. Elsevier. https://books.google.co.kr/books?id=RYyIAQAACAAJ

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1-2*, 19-25. https://doi.org/10.1016/j.softx.2015.06.001

Abraham, P. R., Latha, G. S., Valluri, V. L., & Mukhopadhyay, S. (2014). Mycobacterium tuberculosis PPE protein Rv0256c induces strong B cell response in tuberculosis patients. *Infection, Genetics and Evolution*, *22*, 244-249. https://doi.org/https://doi.org/10.1016/j.meegid.2013.06.023

Akhtar, N., Joshi, A., Singh, J., & Kaushik, V. (2021). Design of a novel and potent multivalent epitope based human cytomegalovirus peptide vaccine: An immunoinformatics approach. *Journal of*

Molecular Liquids, *335*, 116586. https://doi.org/10.1016/j.molliq.2021.116586

Akhtar, N., Kaushik, V., Grewal, R. K., Wani, A. K., Suwattanasophon, C., Choowongkomon, K., Oliva, R., Shaikh, A. R., Cavallo, L., & Chawla, M. (2022). Immunoinformatics-Aided Design of a Peptide Based Multiepitope Vaccine Targeting Glycoproteins and Membrane Proteins against Monkeypox Virus. *Viruses*, *14*(11), 2374. https://doi.org/10.3390/v14112374

Akhtar, N., Magdaleno, J. S. L., Ranjan, S., Wani, A. K., Grewal, R. K., Oliva, R., Shaikh, A. R., Cavallo, L., & Chawla, M. (2023). Secreted Aspartyl Proteinases Targeted Multi-Epitope Vaccine Design for Candida dubliniensis Using Immunoinformatics. *Vaccines*, *11*(2), 364. https://doi.org/10.3390/vaccines11020364

Al Tbeishat, H. (2022). Novel In Silico mRNA vaccine design exploiting proteins of M. tuberculosis that modulates host immune responses by inducing epigenetic modifications. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-08506-4

Albutti, A. (2021a). An integrated computational framework to design a multi-epitopes vaccine against Mycobacterium tuberculosis. *Scientific Reports*, *11*(1), 21929-21929.

https://doi.org/10.1038/s41598-021-01283-6

Albutti, A. (2021b). An integrated computational framework to design a multi-epitopes vaccine against Mycobacterium tuberculosis. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01283-6

Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*, *5*(8), a021113. https://doi.org/10.1101/cshperspect.a021113

Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H. M., Kubo, R. T., & Sette, A. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity*, *1*(9), 751-761.

Andongma, B. T., Huang, Y., Chen, F., Tang, Q., Yang, M., Chou, S.-H., Li, X., & He, J. (2023). In silico design of a promiscuous chimeric multi-epitope vaccine against Mycobacterium tuberculosis. *Computational and structural biotechnology journal*, *21*, 991-1004. https://doi.org/10.1016/j.csbj.2023.01.019

Andreatta, M., & Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system.

Bioinformatics (Oxford, England), *32*(4), 511-517. https://doi.org/10.1093/bioinformatics/btv639

Anwar, T., Ismail, S., Parvaiz, F., Abbasi, S. W., Al-Abbasi, F. A., Alghamdi, A. M., Al-Regaiey, K., Ul-Haq, A., Kaleem, I., Bashir, S., & Waheed, Y. (2023). Computational design of experimentally validated multi-epitopes vaccine against hepatitis E virus: An immunological approach. *PLoS ONE*, *18*(12). https://doi.org/10.1371/journal.pone.0294663

Areeshi, M. (2018). In silico Prediction of Epitope Based Vaccine Candidates against Powassan Virus Infection. *Egyptian Academic Journal of Biological Sciences. C, Physiology and Molecular Biology*, *10*(1), 39-47. https://doi.org/10.21608/eajbsc.2018.13654

Arega, A. M., Dhal, A. K., Pattanaik, K. P., Nayak, S., & Mahapatra, R. K. (2024). An Immunoinformatics-Based Study of Mycobacterium tuberculosis Region of Difference-2 Uncharacterized Protein (Rv1987) as a Potential Subunit Vaccine Candidate for Preliminary Ex Vivo Analysis. *Applied Biochemistry and Biotechnology*, *196*(4), 2367-2395. https://doi.org/10.1007/s12010-023-04658-9

Bellini, C., Vergara, E., Bencs, F., Fodor, K., Bősze, S., Krivić, D.,

Bacsa, B., Surguta, S. E., Tóvári, J., Reljic, R., & Horváti, K. (2023). Design and Characterization of a Multistage Peptide-Based Vaccine Platform to Target Mycobacterium tuberculosis Infection. *Bioconjugate Chemistry*, *34*(10), 1738-1753. https://doi.org/10.1021/acs.bioconjchem.3c00273

Benkert, P., Biasini, M., & Schwede, T. (2010). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343-350. https://doi.org/10.1093/bioinformatics/btq662

Bergeron, H. C., & Tripp, R. A. (2021). Immunopathology of RSV: An Updated Review. *Viruses*, *13*(12), 2478. https://doi.org/10.3390/v13122478

Bhat, K. H., Das, A., Srikantam, A., & Mukhopadhyay, S. (2013). PPE2 protein ofMycobacterium tuberculosismay inhibit nitric oxide in activated macrophages. *Annals of the New York Academy of Sciences*, *1283*(1), 97-101. https://doi.org/10.1111/nyas.12070

Bhat, K. H., Srivastava, S., Kotturu, S. K., Ghosh, S., & Mukhopadhyay, S. (2017). The PPE2 protein of Mycobacterium tuberculosis translocates to host nucleus and inhibits nitric oxide production. *Scientific Reports*, *7*, 39706-39706.

https://doi.org/10.1038/srep39706

Bibi, S., Ullah, I., Zhu, B., Adnan, M., Liaqat, R., Kong, W.-B., & Niu, S. (2021a). In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Scientific Reports*, *11*(1), 1249-1249. https://doi.org/10.1038/s41598-020-80899-6

Bibi, S., Ullah, I., Zhu, B., Adnan, M., Liaqat, R., Kong, W.-B., & Niu, S. (2021b). In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-80899-6

Bogaty, C., & Drebot, M. (2018). Powassan virus - an emerging public health concern. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, *190*(15), E472-E472. https://doi.org/10.1503/cmaj.171305

Booth, M. (2018). Climate Change and the Neglected Tropical Diseases. *Advances in Parasitology*, *100*. https://doi.org/10.1016/bs.apar.2018.02.001

Borchers, A. T., Chang, C., Gershwin, M. E., & Gershwin, L. J. (2013). Respiratory syncytial virus--a comprehensive review. *Clinical reviews in allergy & immunology*, *45*(3), 331-379. https://doi.org/10.1007/s12016-013-8368-9

Broor, S., Parveen, S., & Maheshwari, M. (2018). Respiratory Syncytial Virus Infections in India: Epidemiology and Need for Vaccine. *Indian Journal of Medical Microbiology*, *36*(4), 458-464. https://doi.org/10.4103/ijmm.ijmm_19_5

Bui, H.-H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *7*, 153-153. https://doi.org/10.1186/1471-2105-7-153

Bui, H.-H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *8*, 361-361. https://doi.org/10.1186/1471-2105-8-361

Bullard, K. M., Gullberg, R. C., Soltani, E., Steel, J. J., Geiss, B. J., & Keenan, S. M. (2015). Murine Efficacy and Pharmacokinetic Evaluation of the Flaviviral NS5 Capping Enzyme 2-Thioxothiazolidin-4-One Inhibitor BG-323. *PLoS ONE*, *10*(6), e0130083-e0130083. https://doi.org/10.1371/journal.pone.0130083

Campbell, O., & Krause, P. J. (2020). The emergence of human Powassan virus infection in North America. *Ticks and Tick-*

borne Diseases, *11*(6), 101540. https://doi.org/10.1016/j.ttbdis.2020.101540

Castiglione, F., Mantile, F., De Berardinis, P., & Prisco, A. (2012). How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Computational and mathematical methods in medicine*, *2012*, 842329-842329. https://doi.org/10.1155/2012/842329

Chakaya, J., Petersen, E., Nantanda, R., Mungai, B. N., Migliori, G. B., Amanullah, F., Lungu, P., Ntoumi, F., Kumarasamy, N., Maeurer, M., & Zumla, A. (2022). The WHO Global Tuberculosis 2021 Report – not so good news and turning the tide back to End TB. *International Journal of Infectious Diseases*, *124*, S26-S29. https://doi.org/https://doi.org/10.1016/j.ijid.2022.03.011

Chan, Y., Jazayeri, S. D., Ramanathan, B., & Poh, C. L. (2020). Enhancement of Tetravalent Immune Responses to Highly Conserved Epitopes of a Dengue Peptide Vaccine Conjugated to Polystyrene Nanoparticles. *Vaccines*, *8*(3), 417. https://doi.org/10.3390/vaccines8030417

Chauhan, V., Rungta, T., Goyal, K., Singh, M. P., Chauhan, V., Rungta, T., Goyal, K., & Singh, M. P. (2019). Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing

immunoinformatics approach. *Scientific Reports 2019 9:1*, *9*(1). https://doi.org/10.1038/s41598-019-39299-8

Chawla, M., Cuspoca, A. F., Akthar, N., Magdaleno, J. S. L., Rattanabunyong, S., Suwattanasophon, C., Jongkon, N., Choowongkomon, K., Shaikh, A. R., Malik, T., & Cavallo, L. (2023). Immunoinformatics-aided rational design of a multi-epitope vaccine targeting feline infectious peritonitis virus. *Frontiers in Veterinary Science*, *10*. https://doi.org/10.3389/fvets.2023.1280273

Chen, J., Chen, J., Xu, Q., Chen, J., Chen, J., & Xu, Q. (2022). Current Developments and Challenges of mRNA Vaccines. *Annual Review of Biomedical Engineering*, *24*(Volume 24, 2022). https://doi.org/10.1146/annurev-bioeng-110220-031722

Chen, Z., Zhu, Y., Sha, T., Li, Z., Li, Y., Zhang, F., & Ding, J. (2021). Design of a new multi-epitope vaccine against Brucella based on T and B cell epitopes using bioinformatics methods | Epidemiology & Infection | Cambridge Core. *Epidemiology & Infection*, *149*. https://doi.org/10.1017/S0950268821001229

Cheng, P., Jiang, F., Wang, G., Wang, J., Xue, Y., Wang, L., & Gong, W. (2023). Bioinformatics analysis and consistency verification of a novel tuberculosis vaccine candidate HP13138PB [Original

Research]. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/fimmu.2023.1102578

Cheng, P., Wang, L., & Gong, W. (2022). In silico Analysis of Peptide-Based Biomarkers for the Diagnosis and Prevention of Latent Tuberculosis Infection. *Frontiers in microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.947852

Cheung, A. M., Yip, E. Z., Ashbrook, A. W., Goonawardane, N., Quirk, C., Rice, C. M., MacDonald, M. R., & Hoffmann, H.-H. (2023). Characterization of Live-Attenuated Powassan Virus Vaccine Candidates Identifies an Efficacious Prime-Boost Strategy for Mitigating Powassan Virus Disease in a Murine Model. *Vaccines*, *11*(3), 612. https://doi.org/10.3390/vaccines11030612

Choi, H., Kudchodkar, S. B., Ho, M., Reuschel, E. L., Reynolds, E., Xu, Z., Bordoloi, D., Ugen, K. E., Tebas, P., Kim, J., Abdel-Mohsen, M., Thangamani, S., Weiner, D. B., & Muthumani, K. (2020). A novel synthetic DNA vaccine elicits protective immune responses against Powassan virus. *PLoS neglected tropical diseases*, *14*(10), e0008788-e0008788. https://doi.org/10.1371/journal.pntd.0008788

Chuang, Y.-C., Lin, K.-P., Wang, L.-A., Yeh, T.-K., & Liu, P.-Y. (2023). Impact of COVID-19 pandemic on RSV infection | IDR. *Infection*

and Drug Resistance, *16*. https://doi.org/10.2147/IDR.S396434

Ciarlitto, C., Vittucci, A. C., Antilici, L., Concato, C., Di Camillo, C., Zangari, P., & Villani, A. (2019). Respiratory Syncityal Virus A and B: three bronchiolitis seasons in a third level hospital in Italy. *Italian journal of pediatrics*, *45*(1), 115-115. https://doi.org/10.1186/s13052-019-0704-0

Clifford, J. N., Høie, M. H., Deleuran, S., Peters, B., Nielsen, M., & Marcatili, P. (2022). BepiPred-3.0: Improved B-cell epitope prediction using protein language models. *Protein Science*, *31*(12). https://doi.org/10.1002/pro.4497

Colosia, A., Costello, J., McQuarrie, K., Kato, K., & Bertzos, K. (2023). Systematic literature review of the signs and symptoms of respiratory syncytial virus. *Influenza and other respiratory viruses*, *17*(2), e13100-e13100. https://doi.org/10.1111/irv.13100

Colovos, C., & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. (0961-8368 (Print)).

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511-1519. https://doi.org/10.1002/pro.5560020916

Conley, M. J., Short, J. M., Burns, A. M., Streetley, J., Hutchings, J.,

Bakker, S. E., Power, B. J., Jaffery, H., Haney, J., Zanetti, G., Murcia, P. R., Stewart, M., Fearns, R., Vijayakrishnan, S., & Bhella, D. (2022). Helical ordering of envelope-associated proteins and glycoproteins in respiratory syncytial virus. *The EMBO journal*, *41*(3), e109728-e109728. https://doi.org/10.15252/embj.2021109728

Consortium, T. U. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523-D531. https://doi.org/10.1093/nar/gkac1052

Consortium, T. U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., . . . Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1). https://doi.org/10.1093/nar/gkac1052

Cordeiro, P. A. S., Assone, T., Prates, G., Tedeschi, M. R. M., Fonseca, L. A. M., & Casseb, J. (2022). The role of IFN-γ production during retroviral infections: an important cytokine involved in chronic inflammation and pathogenesis. *Revista do Instituto de Medicina Tropical de Sao Paulo*, *64*, e64-e64.

https://doi.org/10.1590/S1678-9946202264064

Corrin, T., Greig, J., Harding, S., Young, I., Mascarenhas, M., & Waddell, L. A. (2018). Powassan virus, a scoping review of the global evidence. *Zoonoses and Public Health*, *65*(6), 595-624. https://doi.org/10.1111/zph.12485

Cosentino, G., Marougka, K., Desquesnes, A., Welti, N., Sitterlin, D., Gault, E., & Rameix-Welti, M.-A. (2022). Respiratory syncytial virus ribonucleoproteins hijack microtubule Rab11 dependent transport for intracellular trafficking. *PLoS pathogens*, *18*(7), e1010619-e1010619. https://doi.org/10.1371/journal.ppat.1010619

Daddario-DiCaprio, K. M., Geisbert, T. W., Geisbert, J. B., Stroher, U., Hensley, L. E., Grolla, A., Fritz, E. A., Feldmann, F., Feldmann, H., & Jones, S. M. (2006). Cross-Protection against Marburg Virus Strains by Using a Live, Attenuated Recombinant Vaccine. *Journal of virology*, *80*(19). https://doi.org/10.1128/jvi.00959-06

Dar, H. A., Almajhdi, F. N., Aziz, S., & Waheed, Y. (2022). Immunoinformatics-Aided Analysis of RSV Fusion and Attachment Glycoproteins to Design a Potent Multi-Epitope Vaccine. *Vaccines*, *10*(9), 1381.

https://doi.org/10.3390/vaccines10091381

Darboe, F., Adetifa, J. U., Reynolds, J., Hossin, S., Plebanski, M., Netea, M. G., Rowland-Jones, S. L., Sutherland, J. S., & Flanagan, K. L. (2017). Frontiers | Minimal Sex-Differential Modulation of Reactivity to Pathogens and Toll-Like Receptor Ligands following Infant Bacillus Calmette–Guérin Russia Vaccination. *Frontiers in Immunology*, *8*. https://doi.org/10.3389/fimmu.2017.01092

David, A., Islam, S., Tankhilevich, E., & Sternberg, M. J. E. (2022). The AlphaFold Database of Protein Structures: A Biologist's Guide. *Journal of Molecular Biology*, *434*(2), 167336-167336. https://doi.org/10.1016/j.jmb.2021.167336

Della-Giustina, D., Duke, C., & Goldflam, K. (2021). Underrecognized Tickborne Illnesses: Borrelia Miyamotoi and Powassan Virus. *Wilderness & Environmental Medicine*, *32*(2), 240-246. https://doi.org/10.1016/j.wem.2021.01.005

Dey, J., Mahapatra, S. R., Singh, P. K., Prabhuswamimath, S. C., Misra, N., & Suar, M. (2023). Designing of multi-epitope peptide vaccine against Acinetobacter baumannii through combined immunoinformatics and protein interaction–based approaches. *Immunologic Research*. https://doi.org/10.1007/s12026-023-

09374-4

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013a). Prediction of IL4 Inducing Peptides. *Clinical and Developmental Immunology*, *2013*, 1-9. https://doi.org/10.1155/2013/263952

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013b). Prediction of IL4 inducing peptides. *Clinical & developmental immunology*, *2013*, 263952-263952. https://doi.org/10.1155/2013/263952

Dhanda, S. K., Vir, P., & Raghava, G. P. (2013a). Designing of interferon-gamma inducing MHC class-II binders. *Biology direct*, *8*(1), 30. https://doi.org/10.1186/1745-6150-8-30

Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013b). Designing of interferon-gamma inducing MHC class-II binders. *Biology direct*, *8*, 30-30. https://doi.org/10.1186/1745-6150-8-30

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2—a server for in silico prediction of allergens. *Journal of Molecular Modeling*, *20*(6). https://doi.org/10.1007/s00894-014-2278-5

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2013). AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*, *30*(6), 846-851.

https://doi.org/10.1093/bioinformatics/btt619

Dittmer, U., Peterson, K. E., Messer, R., Stromnes, I. M., Race, B., & Hasenkrug, K. J. (2001). Role of interleukin-4 (IL-4), IL-12, and gamma interferon in primary and vaccine-primed immune responses to Friend retrovirus infection. *Journal of virology*, *75*(2), 654-660. https://doi.org/10.1128/JVI.75.2.654-660.2001

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1). https://doi.org/10.1186/1471-2105-8-4

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*, 4-4. https://doi.org/10.1186/1471-2105-8-4

Falsey, A. R., Hennessey, P. A., Formica, M. A., Cox, C., & Walsh, E. E. (2005). Respiratory Syncytial Virus Infection in Elderly and High-Risk Adults. *New England Journal of Medicine*, *352*(17), 1749-1759. https://doi.org/10.1056/nejmoa043951

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis

Resource in Epitope Discovery and Synthetic Vaccine Design. *Frontiers in Immunology*, *8*, 278-278. https://doi.org/10.3389/fimmu.2017.00278

Flower, D. R. (2008). *Bioinformatics for Vaccinology*. John Wiley & Sons Inc. https://doi.org/DOI:10.1002/9780470699836

Forstnerič, V., Ivičak-Kocjan, K., Plaper, T., Jerala, R., & Benčina, M. (2017). The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. *PLoS pathogens*, *13*(8), e1006574-e1006574. https://doi.org/10.1371/journal.ppat.1006574

Fraser, C. C., Altreuter, D. H., Ilyinskii, P., Pittet, L., LaMothe, R. A., Keegan, M., Johnston, L., & Kishimoto, T. K. (2014). Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine*, *32*(24), 2896-2903.

Garegnani, L., Styrmisdóttir, L., Roson Rodriguez, P., Escobar Liquitay, C. M., Esteban, I., & Franco, J. V. (2021). Palivizumab for preventing severe respiratory syncytial virus (RSV) infection in children. *The Cochrane database of systematic reviews*, *11*(11), CD013757-CD013757. https://doi.org/10.1002/14651858.CD013757.pub2

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571-607): Humana Press.

Ghaffari-Nazari, H., Tavakkol-Afshari, J., Jaafari, M. R., Tahaghoghi-Hajghorbani, S., Masoumi, E., & Jalali, S. A. (2015). Improving Multi-Epitope Long Peptide Vaccine Potency by Using a Strategy that Enhances CD4+ T Help in BALB/c Mice. *PLoS ONE*, *10*(11), e0142563-e0142563. https://doi.org/10.1371/journal.pone.0142563

Ghandadi, M. (2022). An Immunoinformatic Strategy to Develop New Mycobacterium tuberculosis Multi-epitope Vaccine. *International Journal of Peptide Research and Therapeutics*, *28*(3), 99-99. https://doi.org/10.1007/s10989-022-10406-0

Gong, W., Liang, Y., Mi, J., Jia, Z., Xue, Y., Wang, J., Wang, L., Zhou, Y., Sun, S., & Wu, X. (2021). Peptides-Based Vaccine MP3RT Induced Protective Immunity Against Mycobacterium Tuberculosis Infection in a Humanized Mouse Model. *Front Immunol*, *12*, 666290. https://doi.org/10.3389/fimmu.2021.666290

Gonzalez-Galarza, F., Faviel, Mccabe, A., Santos, M. D., J, Eduardo,

Jones, J., Takeshita, L., Ortega-Rivera, D., Nestor, Cid-Pavon, M. D., Glenda, Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., Middleton, D., & Jones, R., Andrew. (2019). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1029

Guest, J. D., & Pierce, B. G. (2018). Frontiers | Computational Modeling of Hepatitis C Virus Envelope Glycoprotein Structure and Recognition. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.01117

Guo, W. H., Zhu, Y. J., Haimiti, G., Xie, X. R., Niu, C., Li, M., Shi, J., Yin, Z. W., Yu, M. K., Ding, J. B., & Zhang, F. B. (2023). Bioinformatics-based design of a fusion vaccine with CTLA-4 variable region to combat Brucella. *Brazilian Journal of Medical and Biological Research*, *56*. https://doi.org/10.1590/1414-431X2023e12938

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery, C., & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE*, *8*(9), e73957-e73957. https://doi.org/10.1371/journal.pone.0073957

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE*, *8*(9), e73957. https://doi.org/10.1371/journal.pone.0073957

Gupta, S. K., Bajwa, P., Deb, R., Chellappa, M. M., & Dey, S. (2014). Flagellin a toll-like receptor 5 agonist as an adjuvant in chicken vaccines. *Clinical and vaccine immunology : CVI*, *21*(3), 261-270. https://doi.org/10.1128/CVI.00669-13

Hajighahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi, S. S., & Ghasemi, Y. (2017). Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. *Infection, Genetics and Evolution*, *48*, 83-94. https://doi.org/https://doi.org/10.1016/j.meegid.2016.12.010

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. In: Cold Spring Harbor Laboratory.

Harris, E. (2023). CDC: RSV Vaccine Recommended for Older People. *JAMA*, *330*(5), 401. https://doi.org/10.1001/jama.2023.12537

Henderson, D. A. (1980). Henderson, D. A. (1980). The eradication of

smallpox. Scientific American, 243(4), 41-47. *Scientific American*, *243(4), 41-47*.

Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, *41*(Web Server issue), W384-W388. https://doi.org/10.1093/nar/gkt458

Hermance, M. E., & Thangamani, S. (2017). Powassan Virus: An Emerging Arbovirus of Public Health Concern in North America. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, *17*(7), 453-462. https://doi.org/10.1089/vbz.2017.2110

Hutchinson, E. G., & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Science*, *5*(2), 212-220.

Ilinskaya, A. N., & Dobrovolskaia, M. A. (2016). Understanding the immunogenicity and antigenicity of nanomaterials: Past, present and future. *Toxicology and applied pharmacology*, *299*, 70-77. https://doi.org/10.1016/j.taap.2016.01.005

Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, *154*(3), 394-406.

https://doi.org/10.1111/imm.12889

Jiang, F., Peng, C., Cheng, P., Wang, J., Lian, J., & Gong, W. (2023). PP19128R, a Multiepitope Vaccine Designed to Prevent Latent Tuberculosis Infection, Induced Immune Responses In Silico and In Vitro Assays. *Vaccines*, *11*(4), 856. https://www.mdpi.com/2076-393X/11/4/856

Jiang, F., Wang, L., Wang, J., Cheng, P., Shen, J., & Gong, W. (2023). Design and development of a multi-epitope vaccine for the prevention of latent tuberculosis infection. *Medicine Advances*, *1*(4), 361-382. https://doi.org/10.1002/med4.40

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server issue), W5-W9. https://doi.org/10.1093/nar/gkn201

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *Journal of immunology (Baltimore, Md. : 1950)*, *199*(9), 3360-3368. https://doi.org/10.4049/jimmunol.1700893

Kapadia, S. U., Rose, J. K., Lamirande, E., Vogel, L., Subbarao, K., & Roberts, A. (2005). Long-term protection from SARS coronavirus infection conferred by a single immunization with an attenuated VSV-based vaccine. *Virology*, *340*(2). https://doi.org/10.1016/j.virol.2005.06.016

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., Srivastava, A. P., Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports 2020 10:1*, *10*(1). https://doi.org/10.1038/s41598-020-67749-1

Katelaris, A. L., Jackson, C., Southern, J., Gupta, R. K., Drobniewski, F., Lalvani, A., Lipman, M., Mangtani, P., & Abubakar, I. (2020). Effectiveness of BCG Vaccination Against Mycobacterium tuberculosis Infection in Adults: A Cross-sectional Analysis of

a UK-Based Cohort. *The Journal of Infectious Diseases*, *221*(1). https://doi.org/10.1093/infdis/jiz430

Kaushik, V., G, S. K., Gupta, L. R., Kalra, U., Shaikh, A. R., Cavallo, L., & Chawla, M. (2022). Immunoinformatics Aided Design and In-Vivo Validation of a Cross-Reactive Peptide Based Multi-Epitope Vaccine Targeting Multiple Serotypes of Dengue Virus. *Frontiers in Immunology*, *13*, 865180-865180. https://doi.org/10.3389/fimmu.2022.865180

Kaushik, V., Jain, P., Akhtar, N., Joshi, A., Gupta, L. R., Grewal, R. K., Oliva, R., Shaikh, A. R., Cavallo, L., & Chawla, M. (2022). Immunoinformatics-Aided Design and In Vivo Validation of a Peptide-Based Multiepitope Vaccine Targeting Canine Circovirus. *ACS Pharmacology & Translational Science*, *5*(8), 679-691. https://doi.org/10.1021/acsptsci.2c00130

Kemenesi, G., & Bányai, K. (2018). Tick-Borne Flaviviruses, with a Focus on Powassan Virus. *Clinical microbiology reviews*, *32*(1), e00106-00117. https://doi.org/10.1128/CMR.00106-17

Khairkhah, N., Bolhassani, A., Agi, E., Namvar, A., & Nikyar, A. (2022). Immunological investigation of a multiepitope peptide vaccine candidate based on main proteins of SARS-CoV-2 pathogen. *PLoS ONE*, *17*(6), e0268251.

https://doi.org/10.1371/journal.pone.0268251

Khan, M., Beckham, J. D., Piquet, A. L., Tyler, K. L., & Pastula, D. M. (2019). An Overview of Powassan Virus Disease. *The Neurohospitalist*, *9*(4), 181-182. https://doi.org/10.1177/1941874419844888

Khan, M. T., Mahmud, A., Islam, M. M., Sumaia, M. S. N., Rahim, Z., Islam, K., & Iqbal, A. (2023a). Multi-epitope vaccine against drug-resistant strains of Mycobacterium tuberculosis: a proteome-wide subtraction and immunoinformatics approach. *Genomics & Informatics*, *21*(3), e42. https://doi.org/10.5808/gi.23021

Khan, M. T., Mahmud, A., Islam, M. M., Sumaia, M. S. N., Rahim, Z., Islam, K., & Iqbal, A. (2023b). Multi-epitope vaccine against drug-resistant strains of Mycobacterium tuberculosis: a proteome-wide subtraction and immunoinformatics approach. *Genomics & Informatics*, *21*(3), e42-e42. https://doi.org/10.5808/gi.23021

Khan, Z., Ualiyeva, D., Amissah, O. B., Sapkota, S., Hameed, H. M. A., & Zhang, T. (2022). Insight Into Novel Anti-tuberculosis Vaccines by Using Immunoinformatics Approaches. *Frontiers in microbiology*, *13*, 866873-866873.

https://doi.org/10.3389/fmicb.2022.866873

Khemasuwan, D., & Colt, H. G. (2021). Applications and challenges of AI-based algorithms in the COVID-19 pandemic. *BMJ Innovations*, *7*(2). https://doi.org/10.1136/bmjinnov-2020-000648

Kiss, G., Holl, J. M., Williams, G. M., Alonas, E., Vanover, D., Lifland, A. W., Gudheti, M., Guerrero-Ferreira, R. C., Nair, V., Yi, H., Graham, B. S., Santangelo, P. J., & Wright, E. R. (2014). Structural analysis of respiratory syncytial virus reveals the position of M2-1 between the matrix protein and the ribonucleoprotein complex. *Journal of virology*, *88*(13), 7602-7617. https://doi.org/10.1128/JVI.00256-14

Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, *40*(W1), W294-W297.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017a). The ClusPro web server for protein-protein docking. *Nature protocols*, *12*(2), 255-278. https://doi.org/10.1038/nprot.2016.169

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017b). The ClusPro web server for

protein–protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Kraemer, M. U. G., Reiner, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D., Perkins, T. A., Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., . . . Golding, N. (2019). Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nature microbiology*, *4*(5), 854–863. https://doi.org/10.1038/s41564-019-0376-y

Krow-Lucal, E. R., Lindsey, N. P., Fischer, M., & Hills, S. L. (2018). Powassan Virus Disease in the United States, 2006–2016. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, *18*(6), 286–290. https://doi.org/10.1089/vbz.2017.2239

Laham, F. R., Mansbach, J. M., Piedra, P. A., Hasegawa, K., Sullivan, A. F., Espinola, J. A., & Camargo, C. A., Jr. (2017). Clinical Profiles of Respiratory Syncytial Virus Subtypes A AND B Among Children Hospitalized with Bronchiolitis. *The Pediatric infectious disease journal*, *36*(8), 808–810. https://doi.org/10.1097/INF.0000000000001596

Larsen, J. E. P., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome research*, *2*, 2–2.

https://doi.org/10.1186/1745-7580-2-2

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129-134. https://doi.org/10.1002/pro.3289

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, *26*(2), 283-291.

Lee, S. J., Shin, S. J., Lee, M. H., Lee, M.-G., Kang, T. H., Park, W. S., Soh, B. Y., Park, J. H., Shin, Y. K., Kim, H. W., Yun, C.-H., Jung, I. D., & Park, Y.-M. (2014). A Potential Protein Adjuvant Derived from Mycobacterium tuberculosis Rv0652 Enhances Dendritic Cells-Based Tumor Immunotherapy. *PLoS ONE*, *9*(8), e104351. https://doi.org/10.1371/journal.pone.0104351

Li, J. M., & Zhu, D. Y. (2006). Therapeutic DNA vaccines against tuberculosis: a promising but arduous task. *Chin Med J (Engl)*, *119*(13), 1103-1107.

Li, X., Zhang, S., Lei, J., Zhu, Y., Zhou, X., Xiao, J., & Xiang, T. (2018). Prophylactic herpes simplex virus type 2 vaccine adjuvanted with a universal CD4 T cell helper peptide induces long-term

protective immunity against lethal challenge in mice. *International Immunopharmacology*, *61*, 100-108. https://doi.org/10.1016/j.intimp.2018.05.024

Li, Y., Jain, N., Limpanawat, S., To, J., Quistgaard, E. M., Nordlund, P., Thanabalu, T., & Torres, J. (2015). Interaction between human BAP31 and respiratory syncytial virus small hydrophobic (SH) protein. *Virology*, *482*, 105-110. https://doi.org/10.1016/j.virol.2015.03.034

Ma, Y., Tang, K., Zhang, Y., Zhang, C., Cheng, L., Zhang, F., Zhuang, R., Jin, B., & Zhang, Y. (2020). Protective CD8+ T-cell response against Hantaan virus infection induced by immunization with designed linear multi-epitope peptides in HLA-A2.1/Kb transgenic mice. *Virology Journal*, *17*(1). https://doi.org/10.1186/s12985-020-01421-y

Madan, R., Pandit, K., Bhati, L., Kumar, H., Kumari, N., & Singh, S. (2021). Mining the Mycobacterium tuberculosis proteome for identification of potential T-cell epitope based vaccine candidates. *Microbial Pathogenesis*, *157*, 104996. https://doi.org/10.1016/j.micpath.2021.104996

Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., & Baldi, P. (2010). High-throughput prediction of protein

antigenicity using protein microarray data. *Bioinformatics (Oxford, England)*, *26*(23), 2936-2943. https://doi.org/10.1093/bioinformatics/btq551

Maleki, A., Russo, G., Parasiliti Palumbo, G. A., Pappalardo, F., Maleki, A., Russo, G., Parasiliti Palumbo, G. A., & Pappalardo, F. (2022). In silico design of recombinant multi-epitope vaccine against influenza A virus. *BMC Bioinformatics 2022 22:14*, *22*(14). https://doi.org/10.1186/s12859-022-04581-6

Malonis, R. J., Georgiev, G. I., Haslwanter, D., VanBlargan, L. A., Fallon, G., Vergnolle, O., Cahill, S. M., Harris, R., Cowburn, D., Chandran, K., Diamond, M. S., & Lai, J. R. (2022). A Powassan virus domain III nanoparticle immunogen elicits neutralizing and protective antibodies in mice. *PLoS pathogens*, *18*(6), e1010573-e1010573. https://doi.org/10.1371/journal.ppat.1010573

Mao, Y., Xiao, X., Zhang, J., Mou, X., & Zhao, W. (2023). Designing a multi-epitope vaccine against Peptostreptococcus anaerobius based on an immunoinformatics approach. *Synthetic and Systems Biotechnology*, *8*(4), 757-770. https://doi.org/https://doi.org/10.1016/j.synbio.2023.11.004

Marzec, J., Cho, H.-Y., High, M., Mccaw, Z. R., Polack, F., &

Kleeberger, S. R. (2019). Toll-like receptor 4-mediated respiratory syncytial virus disease and lung transcriptomics in differentially susceptible inbred mouse strains. *Physiological Genomics*, *51*(12), 630-643. https://doi.org/10.1152/physiolgenomics.00101.2019

Massengo, N. R. B., Tinto, B., & Simonin, Y. (2023). One Health Approach to Arbovirus Control in Africa: Interests, Challenges, and Difficulties. *Microorganisms*, *11*(6), 1496. https://www.mdpi.com/2076-2607/11/6/1496

McLean Dm Fau - Donohue, W. L., & Donohue, W. L. Powassan virus: isolation of virus from a fatal case of encephalitis. (0008-4409 (Print)). https://doi.org/D - CLML: 5936:25966:182 OTO - NLM

McShane, H. (2011). Tuberculosis vaccines: beyond bacille Calmette–Guérin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1579). https://doi.org/10.1098/rstb.2011.0097

Mejias, A., Rodriguez-Fernandez, R., Peeples, M. E., & Ramilo, O. (2019). Respiratory Syncytial Virus Vaccines: Are We Making Progress? *The Pediatric infectious disease journal*, *38*(10), e266-e269. https://doi.org/10.1097/INF.0000000000002404

Messina, A., Germano, C., Avellis, V., Tavella, E., Dodaro, V., Massaro,

A., Vitale, R., Masturzo, B., & Manzoni, P. (2022). New strategies for the prevention of respiratory syncytial virus (RSV). *Early Human Development*, *174*, 105666. https://doi.org/10.1016/j.earlhumdev.2022.105666

Mir, W. A. Y., Shrestha, D. B., Rana, W., Yelma Reddy, S. R., Paudel, A., & Verda, L. (2021). Successful Treatment of Respiratory Syncytial Virus Infection in an Immunocompromised Patient With Ribavirin. *Cureus*, *13*(8), e16930-e16930. https://doi.org/10.7759/cureus.16930

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2021). *ColabFold - Making protein folding accessible to all*. Cold Spring Harbor Laboratory.

Mishra, S., Kumar, R., Tiwari, S. K., & Ranjan, P. (2022). Machine learning approaches in the diagnosis of infectious diseases: a review. *Bulletin of Electrical Engineering and Informatics*, *11*(6). https://doi.org/10.11591/eei.v11i6.4225

Moin, A. T., Ullah, M. A., Patil, R. B., Faruqui, N. A., Araf, Y., Das, S., Uddin, K. M. K., Hossain, M. S., Miah, M. F., Moni, M. A., Chowdhury, D. U. S., & Islam, S. (2023). A computational approach to design a polyvalent vaccine against human respiratory syncytial virus. *Scientific Reports*, *13*(1), 9702-

9702. https://doi.org/10.1038/s41598-023-35309-y

Monick, M. M., Yarovinsky, T. O., Powers, L. S., Butler, N. S., Carter, A. B., Gudmundsson, G., & Hunninghake, G. W. (2003). Respiratory Syncytial Virus Up-regulates TLR4 and Sensitizes Airway Epithelial Cells to Endotoxin. *Journal of Biological Chemistry*, *278*(52), 53035-53044. https://doi.org/10.1074/jbc.m308093200

Moodley, A., Fatoba, A., Okpeku, M., Emmanuel Chiliza, T., Blessing Cedric Simelane, M., & Pooe, O. J. (2022). Reverse vaccinology approach to design a multi-epitope vaccine construct based on the Mycobacterium tuberculosis biomarker PE_PGRS17. *Immunologic Research*, *70*(4), 501-517. https://doi.org/10.1007/s12026-022-09284-x

Morris, J. A., Blount, R. E., & Savage, R. E. (1956). Recovery of Cytopathogenic Agent from Chimpanzees with Goryza. *Experimental Biology and Medicine*, *92*(3), 544-549. https://doi.org/10.3181/00379727-92-22538

Moss, W. J. (2017). Measles. *The Lancet*, *390*(10111). https://doi.org/10.1016/S0140-6736(17)31463-0

Mugunthan, S. P., & Harish, M. C. (2021). Multi-epitope-Based Vaccine Designed by Targeting Cytoadherence Proteins of

Mycoplasma gallisepticum. *ACS Omega*, *6*(21), 13742-13755. https://doi.org/10.1021/acsomega.1c01032

Mukherjee, S., Dmitry Tworowski, Rajesh Detroja, Sunanda Biswas Mukherjee, and Milana Frenkel-Morgenstern. (2020). Immunoinformatics and Structural Analysis for Identification of Immunodominant Epitopes in SARS-CoV-2 as Potential Vaccine Targets. *Vaccines 2020, Vol. 8, Page 290*, *8*(2). https://doi.org/10.3390/vaccines8020290

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017a). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*, 42851-42851. https://doi.org/10.1038/srep42851

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017b). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*(1), 42851. https://doi.org/10.1038/srep42851

Nam, H. H., & Ison, M. G. (2019). Respiratory syncytial virus infection in adults. *BMJ*, l5021. https://doi.org/10.1136/bmj.l5021

Naqvi, S. T. Q., Yasmeen, M., Ismail, M., Muhammad, S. A., Nawazish-

I-Husain, S., Ali, A., Munir, F., & Zhang, Q. (2021). Designing of Potential Polyvalent Vaccine Model for Respiratory Syncytial Virus by System Level Immunoinformatics Approaches. *BioMed Research International*, *2021*, 9940010-9940010. https://doi.org/10.1155/2021/9940010

Nayak, S. S., Sethi, G., & Ramadas, K. (2023). Design of multi-epitope based vaccine against Mycobacterium tuberculosis: a subtractive proteomics and reverse vaccinology based immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, *41*(23), 14116-14134. https://doi.org/10.1080/07391102.2023.2178511

Naz, S., Aroosh, A., Caner, A., Şahar, E. A., Toz, S., Ozbel, Y., Abbasi, S. W., Naz, S., Aroosh, A., Caner, A., Şahar, E. A., Toz, S., Ozbel, Y., & Abbasi, S. W. (2023). Immunoinformatics Approach to Design a Multi-Epitope Vaccine against Cutaneous Leishmaniasis. *Vaccines 2023, Vol. 11, Page 339*, *11*(2). https://doi.org/10.3390/vaccines11020339

Nguyen, T. L., & Kim, H. (2024a). Designing a Multiepitope Vaccine against Eastern Equine Encephalitis Virus: Immunoinformatics and Computational Approaches. *ACS Omega*, *9*(1), 1092-1105. https://doi.org/10.1021/acsomega.3c07322

Nguyen, T. L., & Kim, H. (2024b). Discovering peptides and computational investigations of a multiepitope vaccine target Mycobacterium tuberculosis. *Synthetic and Systems Biotechnology*, *9*(3), 391-405. https://doi.org/10.1016/j.synbio.2024.03.010

Nguyen, T. L., & Kim, H. (2024). Immunoinformatics and computational approaches driven designing a novel vaccine candidate against Powassan virus. *Sci Rep*, *14*(1), 5999. https://doi.org/10.1038/s41598-024-56554-9

Nguyen, T. L., Lee, Y., & Kim, H. (2022). Immunogenic Epitope-Based Vaccine Prediction from Surface Glycoprotein of MERS-CoV by Deploying Immunoinformatics Approach. *International Journal of Peptide Research and Therapeutics*, *28*(3). https://doi.org/10.1007/s10989-022-10382-5

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. In *Methods in Molecular Biology* (pp. 59-73): Springer New York.

Nilsson, J. B., Kaabinejadian, S., Yari, H., Kester, M. G. D., van Balen, P., Hildebrand, W. H., & Nielsen, M. (2023). Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. *Science advances*, *9*(47), eadj6367-eadj6367.

https://doi.org/10.1126/sciadv.adj6367

Obando-Pacheco, P., Justicia-Grande, A. J., Rivero-Calle, I., Rodríguez-Tenreiro, C., Sly, P., Ramilo, O., Mejías, A., Baraldi, E., Papadopoulos, N. G., Nair, H., Nunes, M. C., Kragten-Tabatabaie, L., Heikkinen, T., Greenough, A., Stein, R. T., Manzoni, P., Bont, L., & Martinón-Torres, F. (2018). Respiratory Syncytial Virus Seasonality: A Global Overview. *The Journal of Infectious Diseases*, *217*(9), 1356-1364. https://doi.org/10.1093/infdis/jiy056

Pal, R., Ghosh, S., & Mukhopadhyay, S. (2021). Moonlighting by PPE2 Protein: Focus on Mycobacterial Virulence. *The Journal of Immunology*, *207*(10), 2393-2397. https://doi.org/10.4049/jimmunol.2100212

Pal, R., & Mukhopadhyay, S. (2021). PPE2 protein of Mycobacterium tuberculosis affects myeloid hematopoiesis in mice. *Immunobiology*, *226*(1), 152051. https://doi.org/10.1016/j.imbio.2020.152051

Peng, C., Tang, F., Wang, J., Cheng, P., Wang, L., & Gong, W. (2023). Immunoinformatic-Based Multi-Epitope Vaccine Design for Co-Infection of Mycobacterium tuberculosis and SARS-CoV-2. *Journal of personalized medicine*, *13*(1), 116.

https://doi.org/10.3390/jpm13010116

Piantadosi, A., Rubin, D. B., McQuillen, D. P., Hsu, L., Lederer, P. A., Ashbaugh, C. D., Duffalo, C., Duncan, R., Thon, J., Bhattacharyya, S., Basgoz, N., Feske, S. K., & Lyons, J. L. (2016). Emerging Cases of Powassan Virus Encephalitis in New England: Clinical Presentation, Imaging, and Review of the Literature. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *62*(6), 707-713. https://doi.org/10.1093/cid/civ1005

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Marc, G. P., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Robert W. Frenck, J., Hammitt, L. L., . . . Gruber, W. C. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*, *383*(27). https://doi.org/10.1056/NEJMoa2034577

Poland, G. A., Ovsyannikova, I. G., Crooke, S. N., & Kennedy, R. B. (2020). SARS-CoV-2 Vaccine Development: Current Status. *Mayo Clinic Proceedings*, *95*(10). https://doi.org/10.1016/j.mayocp.2020.07.021

Pormohammad, A., Zarei, M., Ghorbani, S., Mohammadi, M., Razizadeh,

M. H., Turner, D. L., Turner, R. J., Pormohammad, A., Zarei, M., Ghorbani, S., Mohammadi, M., Razizadeh, M. H., Turner, D. L., & Turner, R. J. (2021). Efficacy and Safety of COVID-19 Vaccines: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *Vaccines 2021, Vol. 9, Page 467*, *9*(5). https://doi.org/10.3390/vaccines9050467

Pyasi, S., Sharma, V., Dipti, K., Jonniya, N. A., & Nayak, D. (2021). Immunoinformatics Approach to Design Multi-Epitope- Subunit Vaccine against Bovine Ephemeral Fever Disease. *Vaccines*, *9*(8), 925. https://doi.org/10.3390/vaccines9080925

Rahmat Ullah, S., Majid, M., Rashid, M. I., Mehmood, K., & Andleeb, S. (2021). Immunoinformatics Driven Prediction of Multiepitopic Vaccine Against Klebsiella pneumoniae and Mycobacterium tuberculosis Coinfection and Its Validation via In Silico Expression. *International Journal of Peptide Research and Therapeutics*, *27*(2), 987-999. https://doi.org/10.1007/s10989-020-10144-1

Ramirez, A., Co, M., & Mathew, A. (2016). CpG Improves Influenza Vaccine Efficacy in Young Adult but Not Aged Mice. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0150425

Rando, H. M., Lordan, R., Lee, A. J., Naik, A., Wellhausen, N., Sell, E.,

Kolla, L., Consortium, C.-R., Gitter, A., & Greene, C. S. (2023). Application of Traditional Vaccine Development Strategies to SARS-CoV-2. *mSystems*, *8*(2). https://doi.org/10.1128/msystems.00927-22

Rapin, N., Lund, O., & Castiglione, F. (2011). Immune system simulation online. *Bioinformatics*, *27*(14), 2013-2014. https://doi.org/10.1093/bioinformatics/btr335

Rapin, N., Lund, O., & Castiglione, F. (2012). C-Immsim 10.1 server. *PLoS pathogens*, *8*.

Rashidi, S., Faraji, S. N., Mamaghani, A. J., Hatam, S., Kazemi, B., Bemani, P., Tabaei, S. J. S., Hatam, G., Rashidi, S., Faraji, S. N., Mamaghani, A. J., Hatam, S., Kazemi, B., Bemani, P., Tabaei, S. J. S., & Hatam, G. (2022). Bioinformatics analysis for the purpose of designing a novel multi-epitope DNA vaccine against Leishmania major. *Scientific Reports 2022 12:1*, *12*(1). https://doi.org/10.1038/s41598-022-22646-7

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). 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 Research*, *48*(W1), W449-W454.

https://doi.org/10.1093/nar/gkaa379

Robert, M. A., Stewart-Ibarra, A. M., & Estallo, E. L. (2020). Climate change and viral emergence: evidence from Aedes-borne arboviruses. *Current Opinion in Virology*, *40*, 41-47. https://doi.org/10.1016/j.coviro.2020.05.001

Robinson, C. L., Romero, J. R., Kempe, A., Pellegrini, C., & Advisory Committee on Immunization Practices Child/Adolescent Immunization Work, G. (2017). Advisory Committee on Immunization Practices Recommended Immunization Schedule for Children and Adolescents Aged 18 Years or Younger - United States, 2017. *MMWR. Morbidity and mortality weekly report*, *66*(5), 134-135. https://doi.org/10.15585/mmwr.mm6605e1

Ruaro-Moreno, M., Monterrubio-López, G. P., Reyes-Gastellou, A., Castelán-Vega, J. A., Jiménez-Alberto, A., Aparicio-Ozores, G., Delgadillo-Gutiérrez, K., González-Y-Merchand, J. A., & Ribas-Aparicio, R. M. (2023). Design of a Multi-Epitope Vaccine against Tuberculosis from Mycobacterium tuberculosis PE_PGRS49 and PE_PGRS56 Proteins by Reverse Vaccinology. *Microorganisms*, *11*(7), 1647. https://doi.org/10.3390/microorganisms11071647

Ruckwardt, T. J. (2023). The road to approved vaccines for respiratory syncytial virus. *npj Vaccines*, *8*(1), 138-138. https://doi.org/10.1038/s41541-023-00734-7

Ruckwardt, T. J., Morabito, K. M., & Graham, B. S. (2019). Immunological Lessons from Respiratory Syncytial Virus Vaccine Development. *Immunity*, *51*(3), 429-442. https://doi.org/10.1016/j.immuni.2019.08.007

Saha, S., & Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 40-48. https://doi.org/10.1002/prot.21078

Salk, J. E. (1955). Considerations in the Preparation and Use of Poliomyelitis Virus Vaccine. *Journal of the American Medical Association*, *158*(14). https://doi.org/10.1001/jama.1955.02960140001001

Sami, S. A., Marma, K. K. S., Mahmud, S., Khan, M. A. N., Albogami, S., El-Shehawi, A. M., Rakib, A., Chakraborty, A., Mohiuddin, M., Dhama, K., Uddin, M. M. N., Hossain, M. K., Tallei, T. E., & Emran, T. B. (2021). Designing of a Multi-epitope Vaccine against the Structural Proteins of Marburg Virus Exploiting the Immunoinformatics Approach. *ACS Omega*, *6*(47), 32043-32071.

https://doi.org/10.1021/acsomega.1c04817

Sasai, M., & Yamamoto, M. (2013). Pathogen recognition receptors: ligands and signaling pathways by Toll-like receptors. *Int Rev Immunol*, *32*(2), 116-133. https://doi.org/10.3109/08830185.2013.774391

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(Web Server issue), W363-W367. https://doi.org/10.1093/nar/gki481

Scriba, T. J., Kaufmann, S. H. E., Henri Lambert, P., Sanicas, M., Martin, C., & Neyrolles, O. (2016). Vaccination Against Tuberculosis With Whole-Cell Mycobacterial Vaccines. *The Journal of Infectious Diseases*, *214*(5), 659-664. https://doi.org/10.1093/infdis/jiw228

Shanmugam, A., Rajoria, S., George, A. L., Mittelman, A., Suriano, R., & Tiwari, R. K. (2012). Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PLoS ONE*, *7*(2), e30839-e30839. https://doi.org/10.1371/journal.pone.0030839

Sharma, A. D., Grewal, R. K., Gorle, S., Cuspoca, A. F., Kaushik, V., Rajjak Shaikh, A., Cavallo, L., & Chawla, M. T cell epitope based vaccine design while targeting outer capsid proteins of

rotavirus strains infecting neonates: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1-19. https://doi.org/10.1080/07391102.2023.2226721

Sharma, A. D., Grewal, R. K., Gorle, S., Cuspoca, A. F., Kaushik, V., Rajjak Shaikh, A., Cavallo, L., & Chawla, M. (2023). T cell epitope based vaccine design while targeting outer capsid proteins of rotavirus strains infecting neonates: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1-19. https://doi.org/10.1080/07391102.2023.2226721

Sharma, R., Rajput, V. S., Jamal, S., Grover, A., & Grover, S. (2021a). An immunoinformatics approach to design a multi-epitope vaccine against Mycobacterium tuberculosis exploiting secreted exosome proteins. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-93266-w

Sharma, R., Rajput, V. S., Jamal, S., Grover, A., & Grover, S. (2021b). An immunoinformatics approach to design a multi-epitope vaccine against Mycobacterium tuberculosis exploiting secreted exosome proteins. *Scientific Reports*, *11*(1), 13836-13836. https://doi.org/10.1038/s41598-021-93266-w

Shen, Z., Zhang, Y., Li, H., & Du, L. (2022). Rapid typing diagnosis and

clinical analysis of subtypes A and B of human respiratory syncytial virus in children. *Virology Journal*, *19*(1), 15-15. https://doi.org/10.1186/s12985-022-01744-y

Shiraz, M., Lata, S., Kumar, P., Shankar, U. N., & Akif, M. (2021). Immunoinformatics analysis of antigenic epitopes and designing of a multi-epitope peptide vaccine from putative nitro-reductases of Mycobacterium tuberculosis DosR. *Infection, Genetics and Evolution*, *94*, 105017. https://doi.org/10.1016/j.meegid.2021.105017

Srivastava, S., Battu, M. B., Khan, M. Z., Nandicoori, V. K., & Mukhopadhyay, S. (2019). Mycobacterium tuberculosis PPE2 Protein Interacts with p67phox and Inhibits Reactive Oxygen Species Production. *The Journal of Immunology*, *203*(5), 1218-1229. https://doi.org/10.4049/jimmunol.1801143

Stone, E. T., Hassert, M., Geerling, E., Wagner, C., Brien, J. D., Ebel, G. D., Hirsch, A. J., German, C., Smith, J. L., & Pinto, A. K. (2022). Balanced T and B cell responses are required for immune protection against Powassan virus in virus-like particle vaccination. *Cell reports*, *38*(7), 110388-110388. https://doi.org/10.1016/j.celrep.2022.110388

Tahir Ul Qamar, M., Shokat, Z., Muneer, I., Ashfaq, U. A., Javed, H.,

Anwar, F., Bari, A., Zahid, B., & Saari, N. (2020). Multiepitope-Based Subunit Vaccine Design and Evaluation against Respiratory Syncytial Virus Using Reverse Vaccinology Approach. *Vaccines*, *8*(2), 288. https://doi.org/10.3390/vaccines8020288

Telford, S. R., & Piantadosi, A. L. (2023). Powassan virus persistence after acute infection. *mBio*, *14*(4), e0071223-e0071223. https://doi.org/10.1128/mbio.00712-23

Tidman, R., Abela-Ridder, B., & de Castañeda, R. R. (2021). The impact of climate change on neglected tropical diseases: a systematic review. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *115*(2). https://doi.org/10.1093/trstmh/traa192

Topalidou, X., Kalergis, A. M., & Papazisis, G. (2023). Respiratory Syncytial Virus Vaccines: A Review of the Candidates and the Approved Vaccines. *Pathogens (Basel, Switzerland)*, *12*(10), 1259. https://doi.org/10.3390/pathogens12101259

Uddin, M. N., & Roni, M. A. (2021). Challenges of Storage and Stability of mRNA-Based COVID-19 Vaccines. *Vaccines*, *9*(9). https://doi.org/10.3390/vaccines9091033

VanBlargan, L. A., Himansu, S., Foreman, B. M., Ebel, G. D., Pierson,

T. C., & Diamond, M. S. (2018). An mRNA Vaccine Protects Mice against Multiple Tick-Transmitted Flavivirus Infections. *Cell reports*, *25*(12), 3382-3392.e3383. https://doi.org/10.1016/j.celrep.2018.11.082

Vandini, S., Biagi, C., & Lanari, M. (2017). Respiratory Syncytial Virus: The Influence of Serotype and Genotype Variability on Clinical Course of Infection. *International Journal of Molecular Sciences*, *18*(8), 1717. https://doi.org/10.3390/ijms18081717

Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., & Oliva, R. (2011). COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, *27*(20), 2915-2916. https://doi.org/10.1093/bioinformatics/btr484

Varadi, M., & Velankar, S. (2023). The impact of AlphaFold Protein Structure Database on the fields of life sciences. *PROTEOMICS*, *23*(17). https://doi.org/10.1002/pmic.202200128

Vilches, T. N., Zhang, K., Exan, R. V., Langley, J. M., & Moghadas, S. M. (2021). Projecting the impact of a two-dose COVID-19 vaccination campaign in Ontario, Canada. *medRxiv : the preprint server for health sciences*. https://doi.org/10.1101/2020.12.10.20246827

Villanueva, P., Crawford, N. W., Croda, M. G., Collopy, S., Jardim, B. A., Jardim, T. d. A. P., Marshall, H., Prat-Aymerich, C., Sawka, A., Sharma, K., Troeman, D., Wadia, U., Warris, A., Wood, N., Messina, N. L., Curtis, N., & Pittet, L. F. (2023). Safety of BCG vaccination and revaccination in healthcare workers. *Human Vaccines & Immunotherapeutics*, *19*(2). https://doi.org/10.1080/21645515.2023.2239088

Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., Angus, B., Baillie, V. L., Barnabas, S. L., Bhorat, Q. E., Bibi, S., Briner, C., Cicconi, P., Collins, A. M., Colin-Jones, R., Cutland, C. L., Darton, T. C., Dheda, K., Duncan, C. J. A., . . . Zuidewind, P. (2021). Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet*, *397*(10269). https://doi.org/10.1016/S0140-6736(20)32661-1

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, D., P, Tjaart A, Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296-W303.

https://doi.org/10.1093/nar/gky427

WHO. (2023). *Global tuberculosis report 2023*. https://iris.who.int/bitstream/handle/10665/373828/9789240083851-eng.pdf?sequence=1

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(Web Server issue), W407-W410. https://doi.org/10.1093/nar/gkm290

Wilson, A. L., Courtenay, O., Kelly-Hope, L. A., Scott, T. W., Takken, W., Torr, S. J., & Lindsay, S. W. (2020). The importance of vector control for the control and elimination of vector-borne diseases. *PLoS neglected tropical diseases*, *14*(1). https://doi.org/10.1371/journal.pntd.0007831

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*(1), 7-8.

Yin, Z., Chen, Y.-L., Schul, W., Wang, Q.-Y., Gu, F., Duraiswamy, J., Kondreddi, R. R., Niyomrattanakit, P., Lakshminarayana, S. B., Goh, A., Xu, H. Y., Liu, W., Liu, B., Lim, J. Y. H., Ng, C. Y., Qing, M., Lim, C. C., Yip, A., Wang, G., . . . Shi, P.-Y. (2009). An adenosine nucleoside inhibitor of dengue virus. *Proceedings of*

the National Academy of Sciences of the United States of America, *106*(48), 20435-20439. https://doi.org/10.1073/pnas.0907010106

Zhang, L. (2018). Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cellular & molecular immunology*, *15*(2), 182-184. https://doi.org/10.1038/cmi.2017.92

Zhang, W., Xu, L., Park, H.-B., Hwang, J., Kwak, M., Lee, P. C. W., Liang, G., Zhang, X., Xu, J., & Jin, J.-O. (2020). Escherichia coli adhesion portion FimH functions as an adjuvant for cancer immunotherapy. *Nature Communications*, *11*(1), 1187-1187. https://doi.org/10.1038/s41467-020-15030-4

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40-40. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., Huo, F., Cao, Q., Jia, R., Huang, Q., Wang, Z. A., Theodorescu, D., Lv, Q., Li, P., & Yan, C. (2022). FimH confers mannose-targeting ability to Bacillus Calmette-Guerin for improved immunotherapy in bladder cancer. *Journal for immunotherapy of cancer*, *10*(3), e003939. https://doi.org/10.1136/jitc-2021-003939

Zhou, T., Chen, D., Chen, Q., Jin, X., Su, M., Zhang, H., Tian, L., Wen,

S., Zhong, L., Ma, Y., Ma, D., Liang, L., Lu, X., Ni, Q., Yang, N., Pi, G., Zhu, Y., Chen, X., Ma, J., . . . Liu, E. (2024). The impact of the COVID-19 pandemic on RSV outbreaks in children: A multicenter study from China. *Respiratory Medicine*, *234*. https://doi.org/10.1016/j.rmed.2024.107828

Zimmer, A. J., Klinton, J. S., Oga-Omenka, C., Heitkamp, P., Nawina Nyirenda, C., Furin, J., & Pai, M. (2022). Tuberculosis in times of COVID-19. *Journal of epidemiology and community health*, *76*(3), 310-316. https://doi.org/10.1136/jech-2021-217529

Abstract in Korean (국문초록)

생물정보학적 접근을 통한

다중 항원결정기 백신 후보물질의 설계 및 평가

Nguyen Thi Truc Ly

농생명공학부 바이오모듈레이션 전공

서울대학교 대학원

백신은 감염성 질병 예방 및 통제에 중요한 기초 역할을 해왔으며, 역사상 가장 성공적인 공중 보건 개입 중 하나로 간주된다. 천연두의 박멸에서 COVID-19 의 통제에 이르기까지, 백신은 전 세계적으로 수백만 명의 생명을 구한 것으로 평가된다. 그러나 이러한 주목할 만한 성공에도 불구하고, 세균 및 바이러스 감염을 포함한 여러 병원체는 여전히 글로벌 건강에 심각한 위협이 되고 있다. 새로운 감염병의 출현 및 재유행이 주요 사망 원인으로 자리 잡으면서, 이러한 변화하는 문제를 해결할 수 있는 혁신적인 백신 전략의 개발이 절실하다.

전통적인 백신 개발은 약독화 또는 불활화 병원체에 의존하며 여러 한계를 가진다. 이러한 백신은 면역이 약화된 개인에서 원하는 면역 반응을 항상 유도하지 못할 수 있으며, 생백신의 경우 병원성으로 다시 변환될

위험이 존재한다. 또한 SARS-CoV-2 와 같은 빠르게 변이하는 바이러스의 경우 전통적인 접근 방식이 항원 다양성을 따라잡기 어렵다. 따라서 이러한 한계를 극복하고 감염성 질병에 대해 지속적인 보호를 제공할 수 있는 혁신적인 백신 설계 접근법이 필요하다.

이러한 상황에서 생물정보학은 백신 개발을 촉진하는 강력한 도구로 부상하였다. 면역정보학(Immunoinformatics)은 생물정보학의 하위 분야로, 면역 반응을 분석하고 세포성 및 체액성 면역을 자극할 수 있는 항원을 예측하는 데 중점을 둔다. 이 분야는 면역계에 의해 인식되는 특정 단백질 단위인 항원결정기를 식별하여 백신의 전략적 설계를 가능하게 하였다. 이러한 항원결정기는 여러 면역원성 항원결정기를 결합하여 강력하고 표적화된 면역 반응을 유도하는 다중 항원결정기 백신을 만드는 데 사용될 수 있다. 전산학적 방법의 사용은 백신 개발에 소요되는 시간과 비용을 크게 줄이며, 안전하고 효과적인 백신을 설계할 수 있는 플랫폼을 제공한다.

다중 항원결정기 백신은 백신 설계의 새로운 영역이다. 기존 백신이 전체 병원체 또는 큰 서브유닛에 의존하는 반면, 다중 항원결정기 백신은 항원의 가장 면역원이 높은 부분만을 포함하도록 설계된다. 이러한 항원결정기는 세포독성 T 림프구(CTL), 보조 T 림프구(HTL), B 세포를 자극하여 보다 표적화된 면역 반응을 생성한다. 또한, 여러 항원에서 항원결정기를 선택함으로써 다중 항원결정기 백신은 바이러스에서 자주 나타나는 항원 변이를 해결할 수 있어, 더 광범위한 보호 효과를 제공한다.

본 논문은 생물정보학 접근법을 사용하여 세 가지 중요한 감염병인

Powassan 바이러스(POWV), 결핵(TB), 호흡기 세포융합 바이러스(RSV)에 대한 다중 항원결정기 백신 후보의 설계 및 평가에 중점을 둔다. 이러한 질병들은 중대한 글로벌 보건 문제를 야기하며, 효과적인 백신의 부재로 인해 높은 사망률을 기록하고 있다. 본 연구에서는 면역정보학과 계산 백신학 도구를 활용하여 잠재적인 백신 후보를 개발하고, 철저한 전산학적 평가를 통해 이들의 면역원성 잠재력을 탐구하였다. 연구 결과는 새로운 건강 위협에 신속히 적응할 수 있는 차세대 백신 설계의 가능성을 보여주며, 감염성 질병에 대한 효과적이고 적응 가능한 백신 개발에 기여하고자 한다.

주요어: 다중 항원결정기 백신, 면역정보학, 생물정보학, 감염병, 전산학적 백신학

학번: 2020-34548

Acknowledgments

I have reached this significant milestone with the support of many individuals whose help was invaluable. I would like to take this opportunity to express my deep gratitude, though words alone cannot capture the full extent of my appreciation.

First and foremost, I extend my deepest thanks to my supervisor, Prof. Heebal Kim, who has given me the opportunity to become a member of BIOPOP and then a graduate student at Seoul National University. I attribute my research accomplishments and the completion of this dissertation to his unwavering belief, support, guidance, and encouragement throughout my doctoral program.

I am also extremely grateful to the rest of my dissertation committee: Prof. Cheol-Heui Yun, Prof. Sungho Won, Prof. Woori Kwak, and Prof. Minseok Seo for taking the precious time to review my dissertation. Their motivational insights and advice have been instrumental in my dissertation and future research direction.

As an international student, I am profoundly grateful to Seoul National University for awarding me a full scholarship. This generous support allowed me to concentrate fully on my studies and research. I am also deeply grateful to all professors of SNU, especially those in the world-class university program in Biomodulation Major, for their mentorship, assistance, and dedication to imparting knowledge. SNU

and all the professors with whom I had the opportunity to study have greatly motivated me to complete this dissertation and keep pursuing the dream of contributing to education and research in Vietnam.

Further, I extend my deep gratitude to all members of the laboratory of Bioinformatics and Population Genetics (BIOPOP), the laboratory of Microbiome (MICROPOP), and eGnome. A special thanks to Youngho Lee, Dr. Byung June Ko, SangJung Lee, Vladimir Li, Dr. Jaehoon Jung, Dr. Sohyoung Won, Heesu Jeong, Jungwon Yoon, Jeongwoo Seo, Hanshin Shin, SoYun Jhang, Jinchul Jo, Jisol Lee, Jina Kim, Jisu Kim, JinSoo Lim, Jooyong Hong, Dr. DongAhn Yoo, Dr. Sojin Ahn, Dr. Kwondo Kim, Dr. Jisung Jang, Lina Lee, Dr. Clémentine Charton, Youngseok Choi, Dong-Jae Youm, etc. You all have greatly assisted me over the past five years, not just academically and in research but also in life. Certainly, words are not enough to express my gratitude. I will cherish your assistance and kindness forever and hope to welcome you to Vietnam someday!

As a Vietnamese student, I am grateful to the Embassy of the Socialist Republic of Viet Nam in the Republic of Korea, the Science & Technology Office of Vietnam in Korea, the Association of Vietnamese Communities in Korea, the Vietnamese Students Association in Korea, the Vietnamese Students Association at Seoul National University, the Vietnamese Innovation Network in Korea, along all my friends in Korea.

Your continuous support and encouragement have given me a sense of home and the motivation to reach this significant milestone.

I am heartfelt thanks to my parents, siblings, brother-in-law, and two lovely nieces for their endless love and unwavering faith, especially my older brother's sacrifice. Their love and faith are constant, regardless of who I am, where I go, or what I do, giving me compassion in life, guiding and strengthening me each step of the way, and providing me with passion and wisdom in every word of this dissertation. I also express my gratitude to my relatives, neighbors, and friends in Vietnam, who have consistently supported and encouraged me throughout my journey.

Last but certainly not least, I am grateful to all the teachers who have educated me throughout my academic path. A special thanks to Prof. Vo Vien, Prof. Nguyen Phi Hung, Dr. Truong Thi Cam Mai, and the faculty of Quy Nhon University, Sookmyung Women's University, Son My High School, Vo Nguyen Giap High School, etc., for fostering my dreams. Thank you, Prof. Sihyun Ham, for your kindness and your constant encouragement, always reminding me, "You can do it." I believe you are now watching over me from heaven, smiling down. Hope to see you again someday when my soul has matured!

Seoul, December 2024