Syrian Arab Republic
Ministry of Higher Education and Scientifics Research
Syrian Virtual University

الجمهورية العربية السورية
وزارة التعليم العالي والبحث العلمي
الجامعة الافتراضية السورية

In Silico Design of a Novel Multi-Epitope Vaccine for Personalized Immunotherapy in Nasopharyngeal carcinoma (NPC)

A Project submitted for the master's degree in Bioinformatics

Student: Sawsan Ali Olwan

Student ID: sawsan_205382

Supervisor

A.Prof. Majd Aljamali

2024-1446

Summary

Background: Cancer continues to pose one of the greatest health challenges globally, with conventional treatments such as chemotherapy and radiotherapy often carrying significant side effects and offering limited efficacy, especially in advanced stages of the disease. In recent years, immunotherapy has revolutionized the field of oncology by harnessing the body's natural immune defenses to identify and destroy cancer cells. By enhancing the immune system's capacity to recognize and target tumor-specific antigens, immunotherapy presents a promising alternative that is both more precise and potentially more enduring in its therapeutic impact.

Nasopharyngeal carcinoma (NPC) is a distinct malignancy within the head and neck cancer spectrum and is predominant in East and Southeast Asia. NPC is strongly associated with the Epstein–Barr virus (EBV), particularly in endemic regions. This viral association makes NPC distinct, as EBV latent proteins like Epstein-Barr nuclear antigen 1 (EBNA1), latent membrane protein 1 (LMP1), and latent membrane protein 2A (LMP2A) play central roles in tumor development and immune evasion. Additionally, survivin (BIRC5), a tumor-associated protein involved in inhibiting apoptosis and promoting tumor cell survival, is overexpressed in NPC, further driving disease progression. These antigens represent promising targets for developing immunotherapeutic strategies aimed at activating the immune system to attack NPC cells. Despite recent therapeutic advances, late-stage NPC remains challenging to treat effectively, highlighting the need for innovative strategies such as multi-epitope vaccines tailored to the immunogenic profile of individual patients.

Aim of the Study: This study aims to design and evaluate a novel multi-epitope vaccine targeting key antigens associated with NPC, specifically EBV latent proteins (EBNA1, LMP1, LMP2A) and the tumor-associated protein survivin. Utilizing advanced in silico approaches, the research seeks to identify highly immunogenic epitopes capable of eliciting strong cellular and humoral immune responses. The vaccine construct integrates these epitopes into a rationally designed framework, incorporating adjuvants, linkers, and structural elements to optimize immunogenicity, stability, and population coverage. Ultimately, the research aims to develop a targeted immunotherapeutic strategy that addresses the challenges of late-stage NPC, enhances treatment precision, and improves patient outcomes by offering a robust, safe, and effective vaccination platform.

Results: The study successfully designed a multi-epitope vaccine targeting EBV-associated antigens (EBNA1, LMP1, LMP2A) and survivin, critical for NPC progression. Computational analysis identified epitopes with strong immunogenicity, non-toxicity, non-allergenicity, and broad conservation across populations. The finalized vaccine construct demonstrated excellent solubility (SolPro score: 0.842; Protein-sol score: 0.629), good antigenicity (VaxiJen score: 0.550, ANTIGENpro score: 0.641), global coverage rate of 99.96% and a stable physicochemical profile, with a molecular weight of 56 kDa, theoretical isoelectric point (pI) of 9.83, an instability index of 31.86 classifying it as stable and a grand average of hydrophilicity (GRAVY) score of −0.376 suggesting hydrophilicity and enhanced solubility in aqueous environments. Codon optimization yielded a Codon Adaptation Index (CAI) of 1.0 which is ideal and a GC content of 49.09%, ensuring efficient expression in Escherichia coli (E.coli).

Structural validation confirmed a high-quality 3D structure with 96.6% of residues in favorable regions per Ramachandran plot analysis, supported by Z-scores of −9.78 indicating native-like protein stability. Disulfide engineering further improved the construct's stability by introducing strategically placed bonds that reinforced its structural integrity and robustness. Molecular docking showed strong and stable interactions with major histocompatibility complex class I (MHC-I) and major histocompatibility complex class II (MHC-II) receptors, with binding affinities of -15.4, and -13.3 kcal/mol, respectively. Molecular dynamics (MD) simulations further validated the structural stability and flexibility of the vaccine-receptor complexes, confirming their adaptability for efficient antigen binding. Immune simulations revealed robust primary and secondary immune responses, characterized by significant IgG production, activation of cytotoxic T cells, and cytokine secretion, ensuring a balanced and effective immune response. Additionally, the B cell simulation shows a robust response with significant memory B cell formation and sustained IgG production, supporting long-term immunity. These results highlight the vaccine’s structural integrity, broad population coverage, and potential efficacy in NPC immunotherapy.

Conclusion: This study demonstrates the potential of a computationally designed multi-epitope vaccine targeting EBV-associated antigens and survivin for NPC immunotherapy. The vaccine construct shows strong immunogenicity, safety, and broad population coverage, supported by immune simulations and structural analyses. By harnessing in silico tools and a novel antigenic combination, this approach offers a promising path toward more effective, targeted treatments for NPC, with the potential to improve patient outcomes and quality of life. However, further experimental validation and clinical studies are

essential to translate these findings into practical applications. This work not only advances the field of NPC treatment but also lays the foundation for the development of similar immunotherapeutic strategies against other EBV-associated cancers.

Table of Contents

Table of Tables

Table of Figures

Table of Abbreviations

NPC	Nasopharyngeal carcinoma
EBV	Epstein–Barr Virus
BIRC5	Baculoviral Inhibitor of Apoptosis Repeat-Containing 5
EBNA1	Epstein-Barr Nuclear Antigen 1
LMP1	Latent Membrane Protein 1
LMP2A	Latent Membrane Protein 2 **A**
DNA	Deoxyribonucleic Acid
TCF-4	T-cell Factor-4
SP	Streptavidin-Perosidase
CNP	Chronic Nasopharyngitis
CTL	Cytotoxic T Lymphocytes
MHC	Major Histocompatibility Complex
IEDB	Immune Epitope Database
CIITA	Class II transactivator
EMT	Epithelial-mesenchymal transition
IAP	Inhibitor of Apoptosis
TAP	Transporter Associated with Antigen Processing
HTL	Helper T Lymphocytes
IFN-γ	Interferon Gamma
IL-6	Interleukin-6
IL-2	Interleukin-2
RNN	Recurrent Neural Network
LBL	Linear B-cell
HLA-DR	Human Leukocyte Antigen-DR Isotype

PADRE	Pan-DR Epitope
pI	Isoelectric Point
GRAVY	Grand Average of Hydrophilicity
PSIPRED	Protein Structure Prediction
BLAST	Basic Local Alignment Search Tool
PSI-BLAST	Position-Specific Iterated BLAST
SOPMA	Self-Optimized Prediction Method with Alignment
ERRAT	Error Analysis of Tertiary Structures
ProSA	Protein Structure Analysis
PDBsum	Protein Data Bank Summary
AUC	Area Under the Curve
DbD2	Disulfide by Design 2
PDB	Protein Data Bank
MHC-I	Major Histocompatibility Complex Class I
MHC-II	Major Histocompatibility Complex Class II
RMSD	Root-Mean-Square Deviation
JCat	Java Codon Adaptation Tool
C-ImmSim	Computational Immune Simulation
IC50	Half-Maximal Inhibitory Concentration
NF-κB	Nuclear Factor Kappa B
JNK	c-Jun N-terminal Kinase
PI3K/Akt	Phosphatidylinositol-3-Kinase/Protein Kinase B Pathway
PD-1/PD-L1	Programmed Cell Death Protein 1/Programmed Death-Ligand 1
CD8	Cluster of Differentiation 8
CD4	Cluster of Differentiation 4
iMODS	Internal Coordinates Normal Mode Analysis Server
NMA	Normal Mode Analysis

SIB	Swiss Institute of Bioinformatics
PRODIGY	Protein Binding Energy Prediction
rpIL	50S Ribosomal Protein L7/L12
MD	Molecular Dynamics
SH3	Src Homology 3
TK	Tyrosine Kinase
SFKs	Src family Kinases
CNS	Central Nervous System
cDNA	Complementary Deoxyribonucleic Acid
E. coli	Escherichia Coli
Bp	Base Pairs
XhoI	Xanthomonas holcicola Restriction Enzyme I
BamHI	Bacillus amyloliquefaciens Restriction Enzyme H
PCR	Polymerase Chain Reaction
MCS	Multiple Cloning Site

Introduction

NPC is a distinct type of head and neck cancer with a global incidence exceeding 130,000 cases annually, more than 77% of the incidence rate of NPC occurs in East Asia and Southeast Asia, especially in southern China [1]. Histologically, NPC is classified into three types: keratinizing squamous cell carcinoma (Type I), non-keratinizing carcinoma (Type II), and undifferentiated carcinoma (Type III). The prevalence of EBV is 100% in Type II and Type III, which is predominant in endemic areas such as North Africa, Southeast Asia, and South China [2, 3]. This association not only distinguishes NPC from other head and neck cancers but also presents unique antigenic targets for therapeutic intervention.

Despite advancements in radiotherapy and chemotherapy, the prognosis for NPC remains poor, particularly for advanced-stage patients. This unfavorable outcome is largely due to late-stage diagnosis, high metastatic potential, and resistance to conventional treatments [3]. There has been a significant increase in research on NPC over the past two decades.

Immunotherapy has emerged as a promising avenue for cancer treatment, aiming to harness the body's immune system to combat tumor cells. In NPC, immunotherapeutic approaches have been limited by the lack of vaccines capable of eliciting potent and specific immune responses against tumor-associated antigens. Given the strong association of Type III and Type II NPC with EBV, targeting EBV-specific antigens provides a strategic approach for vaccine development. Since EBV is predominantly found in the latent phase in NPC, latent proteins such as EBNA1, LMP1, and LMP2A are consistently expressed and play crucial roles in oncogenesis and immune evasion [3].

In addition to EBV antigens, survivin—a member of the inhibitor of apoptosis protein (IAP) family—is overexpressed in many cancers, including NPC. It plays a key role in tumor cell survival by inhibiting apoptotic pathways and promoting cell proliferation, contributing significantly to NPC progression. In normal tissues, survivin is minimally expressed and primarily restricted to embryonic development and adult tissues with high cellular turnover, such as the placenta, hematopoietic stem cells, and the basal layer of the skin [4, 5]. This restricted expression highlights its relevance in cancer research, as it offers a specific avenue for therapeutic intervention.

This study aims to design a multi-epitope vaccine targeting NPC-specific antigens, including EBV latent proteins (EBNA1, LMP1, LMP2A) and survivin, to enhance immunotherapy for NPC. By employing in silico methods to predict and select highly immunogenic epitopes, we aim to construct a vaccine capable

of eliciting robust cellular and humoral immune responses. The integration of advanced immunoinformatics tools for epitope prediction, vaccine design, and molecular modeling provides a novel approach to addressing the therapeutic challenges associated with NPC.

Previous research has not extensively explored the use of computational vaccine design targeting both EBV-associated antigens and survivin in NPC. By developing an effective multi-epitope vaccine, this study has the potential to revolutionize NPC treatment by providing a personalized and targeted immunotherapeutic option. This innovation could significantly improve survival rates and quality of life for patients facing this challenging disease.

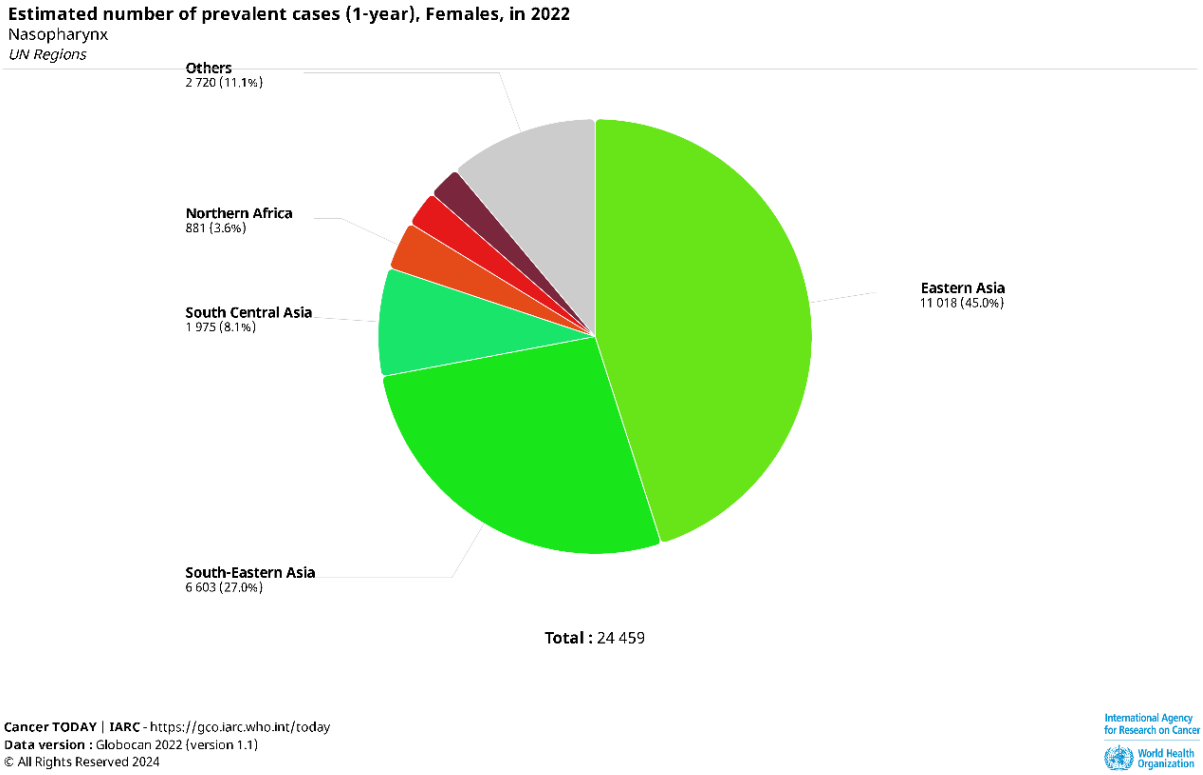

Figure 1:Global distribution of NPC estimated by GLOBOCAN in 2022 (incidences in Females) [2].

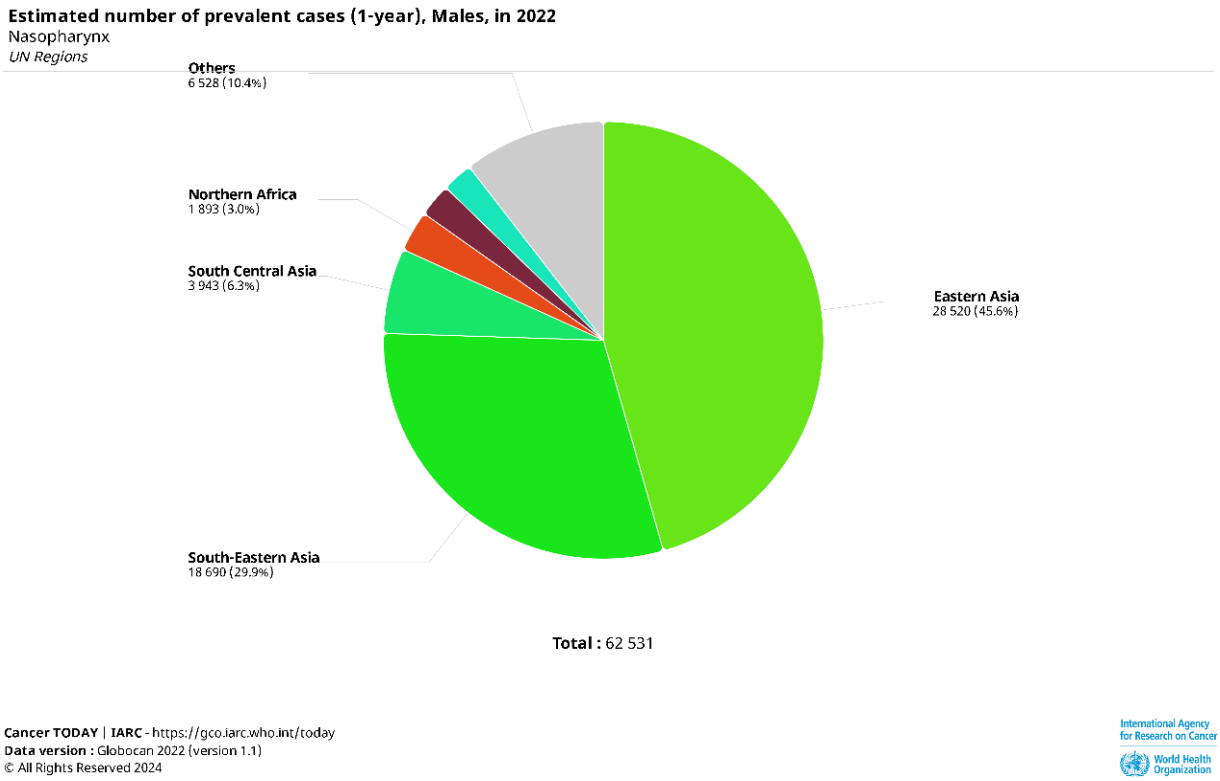

Figure 2: Global distribution of NPC estimated by GLOBOCAN in 2022 (incidences in males) [2].

Methodology

I. Protein selection

The Epstein-Barr virus (EBV), a member of the Herpesviridae family, is known for its ability to establish persistent infections within host cells, primarily through a state of latency. During latency, EBV expresses a subset of proteins that maintain the viral genome, evade immune responses, and drive oncogenic transformation, making them ideal targets for a vaccine. Among these are EBNA1, LMP1, and LMP2A, which play pivotal roles in promoting cancerous transformations within infected cells.

EBNA1: EBNA1 is indispensable for maintaining the viral genome in host cells by tethering the episomal DNA to host chromosomes during cell division. It facilitates immune evasion by inhibiting antigen presentation through its Gly-Ala repeat domain, reducing recognition by cytotoxic T lymphocytes. Additionally, EBNA1 can modulate host cellular pathways to promote immune evasion and tumor progression. Its consistent expression in latency types I, II, and III, but absence in latency 0, makes it a prime target for therapeutic intervention. In NPC, which predominantly exhibits latency type II, EBNA1 plays a crucial role in sustaining viral persistence and driving oncogenesis through immune evasion and tumor-supportive mechanisms. This association underscores its significance as a therapeutic target in NPC treatment[3, 6, 7].

LMP1: Acting as an oncoprotein, LMP1 mimics CD40 signaling to activate key survival pathways such as nuclear factor kappa B (NF-κB), c-Jun N-terminal kinase (JNK), and phosphatidylinositol-3-kinase/protein kinase B pathway (PI3K/Akt). This activation results in the upregulation of pro-survival genes and immunosuppressive cytokines like IL-10, creating a tumor-friendly microenvironment. Furthermore, LMP1 plays a critical role in modulating immune checkpoints, such as PD-1/PD-L1, and promoting angiogenesis, underscoring its potential as a vaccine target [3, 6, 7].

LMP2A: LMP2A supports EBV latency by mimicking B-cell receptor signaling, thereby enhancing cell survival, and preventing apoptosis. It achieves this by activating the PI3K/Akt pathway, promoting cell survival signals. LMP2A also contributes to immune evasion by reducing MHC class II expression through downregulation of the class II transactivator (CIITA) pathway. Additionally, LMP2A's expression has been linked to increased cell motility and epithelial-mesenchymal transition (EMT), processes that contribute to NPC metastasis. These multifaceted roles make LMP2A an attractive target for immunotherapeutic strategies aimed at combating NPC. [3, 6, 7].

Survivin: Beyond EBV-associated proteins, survivin represents a significant tumor-associated target. As a member of the IAP protein family, survivin is highly overexpressed in NPC tissues compared to normal nasopharyngeal tissues (Figure 3). This overexpression is strongly correlated with advanced stages of NPC, increased metastasis, and poorer survival rates, making survivin a critical target for immunotherapy. Its minimal expression in normal cells compared to NPC tissues makes survivin a selective target for the immune system, enabling a vaccine to focus on cancer cells with minimal impact on healthy tissue.

In summary, a multi-epitope vaccine targeting EBNA1, LMP1, LMP2A, and survivin addresses key pathways of viral persistence, immune evasion, and tumor cell survival in NPC. This strategic approach holds promises for controlling NPC progression, managing latent EBV infections, and reducing the recurrence risk of this EBV-associated malignancy.

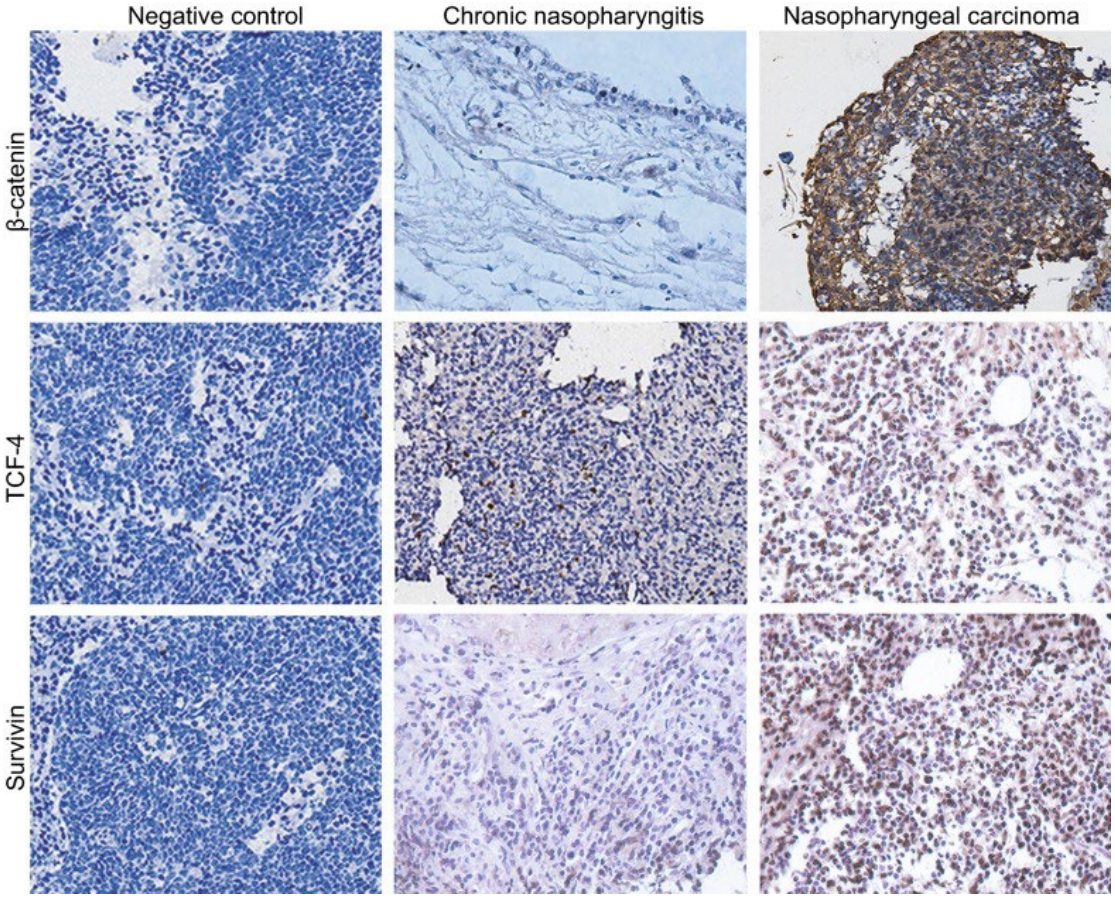

Figure 3: Immunohistochemical analysis on expression of β-catenin, T-cell Factor-4 (TCF-4) and survivin proteins in the NPC tissues and Chronic Nasopharyngitis (CNP) tissues (SP, ×200) [4].

II. Retrieval of Protein Sequences

The protein sequences for EBNA1, LMP1, LMP2A, and survivin were retrieved from the UniProt database. EBNA1 sequences were obtained from three different Epstein-Barr virus (EBV) strains with the following UniProt accession numbers: P03211 (strain B95-8), Q3KS54 (strain GD1), and Q1HVF7 (strain AG876), each consisting of 641 amino acids. LMP1 sequences were retrieved from two strains: P03230 (strain B95-8) and P13198 (strain Raji), both with a length of 386 amino acids. For LMP2A, three strain-specific sequences were included: P13285 (strain B95-8), P0C729 (strain GD1), and Q1HVJ2 (strain AG876), with sequence lengths ranging from 496 to 497 amino acids. Additionally, survivin (BIRC5) was retrieved from UniProt (Accession: O15392). Survivin has seven different isoforms, generated through alternative splicing, contributing to its structural diversity. All sequences were extracted in FASTA format and will be used for epitope prediction, structural modeling, and immunogenicity assessments in the vaccine development pipeline.

III. Prediction of T-cell epitopes: CTL

Major Histocompatibility Complex (MHC) molecules are essential in presenting peptides on the cell surface, facilitating the activation of T cells and playing a pivotal role in initiating T-cell-mediated immune responses. The interaction between MHC molecules and antigenic peptides represents the most selective and critical step in the antigen presentation pathway. To identify potential MHC class I-restricted cytotoxic T lymphocyte (CTL) epitopes, the immune epitope database (IEDB) online server was employed [8]. Predictions were conducted using the NetMHCpan 4.1 BA method, recognized as a top-performing binding predictor based on weekly automated benchmarks. The HLA allele reference set was utilized, with the inclusion of HLA-A0207 and HLA-B4601 due to their positive associations with NPC in previous studies [9, 10]. The identified CTL epitopes, each consisting of 9 amino acids (9-mer peptides), were selected based on their strong binding affinity with half-maximal inhibitory concentration (IC50) scores below 200 nM. IC_{50} values are defined as the concentration that inhibits 50% binding of a labeled reference peptide. The lower the IC50 value, the stronger the binding affinity of the peptide to the MHC molecule[11]. The antigenicity of the selected epitopes was assessed using VaxiJen 2.0, which evaluates the potential of epitopes to trigger immune responses specific to the target organism type [12]. A threshold of 0.4 was set to prioritize highly antigenic candidates. AllerTop 2.0 was used to predict allergenicity, categorizing epitopes as allergenic or non-allergenic [13], while ToxinPred was employed to filter out potentially toxic epitopes [14]. Immunogenicity scores for each CTL epitope were calculated using the IEDB Immunogenicity Prediction Tool to evaluate their

potential to activate T-cell-mediated responses [15]. Sequence conservation was analyzed using the IEDB Conservancy Analysis Tool, which identifies epitopes conserved across strains and isoforms of a protein [16]. Only epitopes with 100% conservation were considered.

Finally, T-cell epitope processing prediction was performed using the IEDB server, incorporating evaluations of proteasomal cleavage and transporter associated with antigen processing (TAP) transport efficiency [17]. Proteasomal cleavage prediction ensured that the selected epitopes could be naturally generated from their source proteins, while TAP transport efficiency assessed their likelihood of being effectively transported into the endoplasmic reticulum for MHC-I molecule presentation. These evaluations ensured that the selected epitopes were optimally processed and transported, enhancing their potential to elicit robust CD8+ T-cell responses. The Selected CTL epitopes were further validated using NetCTL 1.2 [18] and NetMHCcons [19] tools.

IV. Prediction of T-cell epitopes: HTL

Helper T lymphocyte (HTL) epitopes were predicted using the IEDB online server (NetMHCIIPan 4.1 BA method) [20]. The HLA allele reference set was used, and epitopes consisting of 15 amino acids (15-mer peptides), with IC50 scores below 200 nM were selected. The selected epitopes were further evaluated for antigenicity (using VaxiJen 2.0, with a threshold of 0.4) [12], allergenicity (AllerTop 2.0) [13], toxicity (ToxinPred) [14], and sequence conservation (Conservation Across Antigens IEDB) across strains and isoforms [16] to ensure broad applicability and safety.

To evaluate the functional potential of the selected HTL epitopes, their ability to induce critical cytokines was assessed using the IFNepitope [21], IL6-Pred [22], and il2pred [23] servers. These tools predicted the capacity of each epitope to stimulate key cytokines that modulate immune responses: interferon-gamma (IFN-γ), a hallmark of a robust Th1 (T-helper1 cells) response essential for activating cytotoxic T-cells and macrophages; interleukin-6 (IL-6), which supports T-cell differentiation and promotes an effective immune response; and interleukin-2 (IL-2), which is critical for T-cell proliferation, differentiation, and the development of long-lasting memory T cells. By prioritizing epitopes capable of eliciting these cytokines, the analysis ensured the selection of candidates that enhance anti-pathogen or anti-tumor immunity by driving potent immune activation and T-cell-mediated responses. Furthermore, the selected HTL epitopes were validated using NetMHCII version 2.3 [24] to confirm their strong binding affinity to MHC class II molecules, a key determinant of their potential immunogenicity.

V. Prediction of Linear B-cell Epitopes

Linear B-cell (LBL) epitopes were predicted using the ABCpred Prediction Server [25, 26], which employs a recurrent neural network (RNN) for precise identification of continuous B-cell epitopes within antigenic sequences. A window length of 16 amino acids was chosen and a threshold of 0.8 was applied to prioritize high-confidence epitopes. The selected epitopes were further evaluated for antigenicity using VaxiJen 2.0 (threshold > 0.4) [12], allergenicity using AllerTop 2.0 [13], and toxicity using ToxinPred [14]. Only epitopes fully conserved across protein strains and isoforms were shortlisted to ensure broad cross-reactivity and applicability in vaccine design.

VI. Population Coverage

The population coverage of the selected CTL and HTL epitopes was analyzed using the Population Coverage Tool from IEDB online server [27]. This tool assesses the potential efficacy of epitopes in providing immunological protection across different geographical regions and ethnic groups. The analysis focused on determining the percentage of the population covered by the proposed vaccine construct in various regions, calculating the average number of epitope hits per individual, and evaluating the HLA combinations recognized by 90% of the population (pc90). Standard deviation values were also computed to assess the consistency and reliability of the predictions. The study considered coverage data for regions including East Asia, North Africa, South Asia, and Southeast Asia, among others, and calculated an overall global coverage rate.

VII. Combining the final multi-epitope vaccine construct

The vaccine construct includes CTL, HTL, and LBL epitopes, along with IFN-γ, IL-2, and IL-6-inducing epitopes. To overcome HLA restriction and ensure a broad CD4+ T-cell response across diverse populations, the Pan HLA-DR reactive epitope (PADRE), a universal epitope, was incorporated [28].

The 50S ribosomal protein L7/L12 (rpIL), a key component of the prokaryotic ribosome, functions as a "danger signal" in vaccine development, alerting the immune system to the vaccine's presence. By enhancing antigen presentation and promoting immune activation, rpIL boosts overall immunogenicity, making it an ideal adjuvant for strengthening immune responses [28].

To complement this, a fynomer sequence was added to enhance stability and functionality by improving molecular interactions. Fynomers are small and stable globular protein derived from the Src homology 3 (SH3) domain of the human tyrosine-protein kinase Fyn. Fyn is a cytoplasmic, non-receptor tyrosine kinase (TK) of the Src family kinases (SFKs) consisting of 11 members in humans. It

plays a crucial role in various signal transduction pathways within the central nervous system (CNS). Their structure contains two antiparallel β-sheets and two flexible loops called RT and Src loops, essential in interactions with other proteins. It is highly stable (melting temperature ~70°C), monomeric, non-immunogenic, and lacks cysteine residues, minimizing misfolding risks. Easily expressed in E. coli, it serves as a versatile scaffold for enhancing structural stability and vaccine efficacy [28].

Linkers such as EAAAK and KK were used to maintain the spatial arrangement of epitopes, ensuring proper folding and interaction. Additionally, the H5E tag, a histidine-rich peptide with the sequence HEHEHEHEH, was incorporated to facilitate purification and detection. The overall design and organization of these components are illustrated in (Figure 6).

VIII. Assessment of Antigenicity, Allergenicity, Toxicity, Solubility, and Physicochemical Characteristics of the Vaccine

To evaluate the antigenicity, allergenicity, toxicity, solubility, and physicochemical characteristics of the designed vaccine construct, several computational tools were employed. Antigenicity was assessed using the VaxiJen v2.0 [12] and ANTIGENpro [29] servers. The allergenicity of the vaccine was analyzed with the AllerCatPro 2.0 server [30] to confirm its non-allergenic nature. The presence of toxic components was evaluated using the ToxinPred server [14], ensuring the vaccine construct, including the adjuvant sequence, epitopes, linkers, fynomer sequence, and H5E tag, contained no toxin elements. Solubility predictions were performed using SolPro [29] and Protein-sol [31] servers.

Physicochemical properties were evaluated using the ProtParam tool [32] from the ExPASy server. Parameters such as amino acid composition, molecular weight, theoretical isoelectric point (pI), extinction coefficient, instability index, aliphatic index, and GRAVY score were calculated to determine the vaccine's stability and suitability for physiological environments.

IX. Secondary structure prediction

The secondary structure of the designed vaccine construct was analyzed using protein structure prediction (PSIPRED), a highly reliable and widely utilized tool for secondary structure prediction. PSIPRED integrates position-specific Iterated BLAST (PSI-BLAST) to generate an evolutionary conservation profile of the protein sequence. This profile is processed by two advanced feed-forward neural networks, enabling precise predictions of alpha helices, beta strands, and random coils. This analysis provides crucial insights into the structural organization of the

vaccine construct, supporting its design and functionality [33]. To complement this, the self-optimized prediction method with alignment (SOPMA) tool was used for an independent prediction of the secondary structure elements. SOPMA predicts the structural composition based on multiple sequence alignments and provides a detailed breakdown of the proportions of alpha helices, beta strands, and random coils within the construct [34]. These methods together ensured a reliable and accurate prediction of the vaccine's secondary structure, essential for assessing its stability and flexibility.

X. Modeling, Refinement, and Validation of the 3D Structure

The 3D structure of the multi-epitope vaccine was modeled using the Robetta server [35], utilizing its ab initio structure prediction capabilities. The generated model was refined using the GalaxyRefine server, which applies molecular dynamics simulations to optimize side-chain conformations and improve the overall structural quality through iterative relaxation [36]. Structural validation was performed on the refined vaccine using tools like Error Analysis of Tertiary Structures (ERRAT) [37] and Protein Structure Analysis (ProSA) [38, 39], which assess the global and local quality of the predicted model by evaluating geometry, residue conformation, and stereochemistry. The Ramachandran plot, generated using PROCHECK analysis [40] in the PDBsum server [41], provided an additional layer of validation by analyzing the stereochemical properties of the refined model. Finally, the vaccine structure was visualized and rendered for further analysis using UCSF ChimeraX software [42].

XI. Prediction of Conformational B-Cell Epitopes

Following the validation of the vaccine's 3D structure, conformational B-cell epitopes were predicted using the ElliPro server [43]. This tool identifies discontinuous epitopes by analyzing protein geometry, solvent accessibility, and flexibility. Prediction parameters were set to a minimum protrusion score of 0.8 and a maximum residue distance of 6 Å, ensuring a balance between sensitivity and specificity. An area under the curve (AUC) score of 0.732 confirmed the reliability of ElliPro in identifying key epitopes with a high likelihood of effectively engaging the immune system.

XII. Disulfide engineering of the vaccine construct

Disulfide engineering was conducted using disulfide by design 2 (DbD2), a specialized tool designed for rational disulfide bond formation in proteins. DbD2 evaluates residue pairs in a protein structure for proximity and geometric compatibility, assuming mutation to cysteine residues, to identify potential sites for disulfide bond introduction. The input model, the refined 3D vaccine

structure, was provided as a protein data bank (PDB) file, and DbD2 analyzed the structure to output residue pairs that met the criteria for disulfide bond engineering [44]. Key selection parameters included optimal Chi3 angles (torsion angles) between -87° and +97° and interaction energy values below 2.2 kcal/mol [45], indicating favorable bond stability. This approach ensures that the engineered disulfide bonds enhance the structural stability of the vaccine construct, contributing to its robustness and suitability for downstream applications.

XIII. Molecular docking of the vaccine construct with MHC-I, and MHC-II receptors and molecular dynamics simulation

In this study, the vaccine construct was docked against MHC-I and MHC-II receptors to assess its binding affinity and interactions. The receptor structures (PDB IDs: 1KG0 for MHC-I and 1ZIW for MHC-II) were pre-processed using ChimeraX to remove ligands and water molecules, ensuring a clean binding interface. Only the alpha chain was retained for docking to focus on the biologically relevant binding region responsible for immune recognition while minimizing computational complexity. This approach allowed for a precise evaluation of the vaccine's potential to interact with MHC molecules effectively. Molecular docking was performed using the ClusPro 2.0 server [46], followed by detailed analysis of binding interactions through the PDBsum server. PDBsum provided comprehensive insights into the interface areas, interacting residues, and the types of interactions (e.g., hydrogen bonds, salt bridges, and non-bonded contacts) for each receptor-vaccine complex [41]. Additionally, binding affinity and dissociation constants (Kd) were computed using the PRODIGY server, which quantified the thermodynamic stability of the complexes by calculating the binding free energy (ΔG) values [47]. This multi-step approach ensured a thorough assessment of the structural and energetic properties of the vaccine-receptor interactions, supporting the evaluation of its immunogenic potential.

To unravel the intricate dynamics of vaccine-receptor interactions, the best-scoring complexes were further analyzed using the internal coordinates normal mode analysis (iMODS) server to assess their stability and flexibility. This tool employs normal mode analysis (NMA) to evaluate molecular interactions without the need for time-dependent simulations, providing a realistic depiction of structural behavior [48]. The vaccine-receptor complexes were evaluated using key parameters such as B-factors, covariance, deformability, and elastic network analysis. The B factor quantifies the thermal motion of atoms in a crystal structure, reflecting their displacement from their mean positions. It is essential for understanding the flexibility and stability of various regions within a

macromolecule. An elevated B factor often indicates greater atomic motion, suggesting that those areas might be more flexible or less rigid. In the covariance map, highly correlated regions represent residues moving together dynamically, which is crucial for maintaining structural stability or facilitating specific functions. Conversely, anti-correlated regions reflect residues contributing to flexibility or large-scale movements, which are critical for molecular functions such as binding or conformational transitions. This comprehensive analysis highlights the stability and dynamic behavior of the vaccine-receptor complexes.

XIV. Codon Optimization and In Silico Vaccine Cloning

Codon optimization of the final vaccine sequence was performed using the java codon adaptation tool (JCat) to enhance expression efficiency in E.coli (strain K12). This tool adapts the sequence to align with the host organism's codon usage preferences, thereby improving translational efficiency and maximizing protein production [49].

For in silico cloning, SnapGene software was used to design the integration of the optimized vaccine sequence into the pET-28a(+) expression vector [50]. This vector was selected for its widespread application in high-expression protein systems in E. coli. The cloning strategy incorporated XhoI at the 5' end and BamHI at the 3' end of the vaccine sequence to facilitate directional cloning. A restriction site analysis was conducted to ensure that these enzymes do not cut within the vaccine sequence, preserving the integrity of the insert.

The cloning workflow involved simulating the digestion of the pET-28a(+) vector with XhoI and BamHI, followed by ligation of the optimized vaccine sequence into the vector. To verify the construct, a simulated agarose gel electrophoresis was performed, confirming the expected fragment sizes. This computational approach ensured the correct design of the recombinant plasmid, enabling subsequent experimental validation and protein expression studies in E. coli.

XV. Immune simulation

The immune response simulations were carried out using the computational immune simulation C-ImmSim server, a computational tool that models immune system dynamics and predicts the effects of vaccine constructs [51]. To replicate a realistic vaccination schedule, three doses of the vaccine were administered at intervals corresponding to days 1, 28, and 56. These intervals were strategically chosen to simulate the progression from an initial immune activation to subsequent booster responses, allowing the evaluation of primary, secondary, and tertiary immune reactions.

The simulation was configured with default settings, excluding the addition of lipopolysaccharides, ensuring the focus remained on the vaccine construct's intrinsic immunogenic properties. This approach provided a comprehensive overview of the vaccine's ability to elicit antibody production, cytokine responses, and the development of long-lasting immunological memory. The use of C-ImmSim enabled a detailed and systematic exploration of the vaccine's potential efficacy, making it a valuable step in the design process.

Methods steps summary

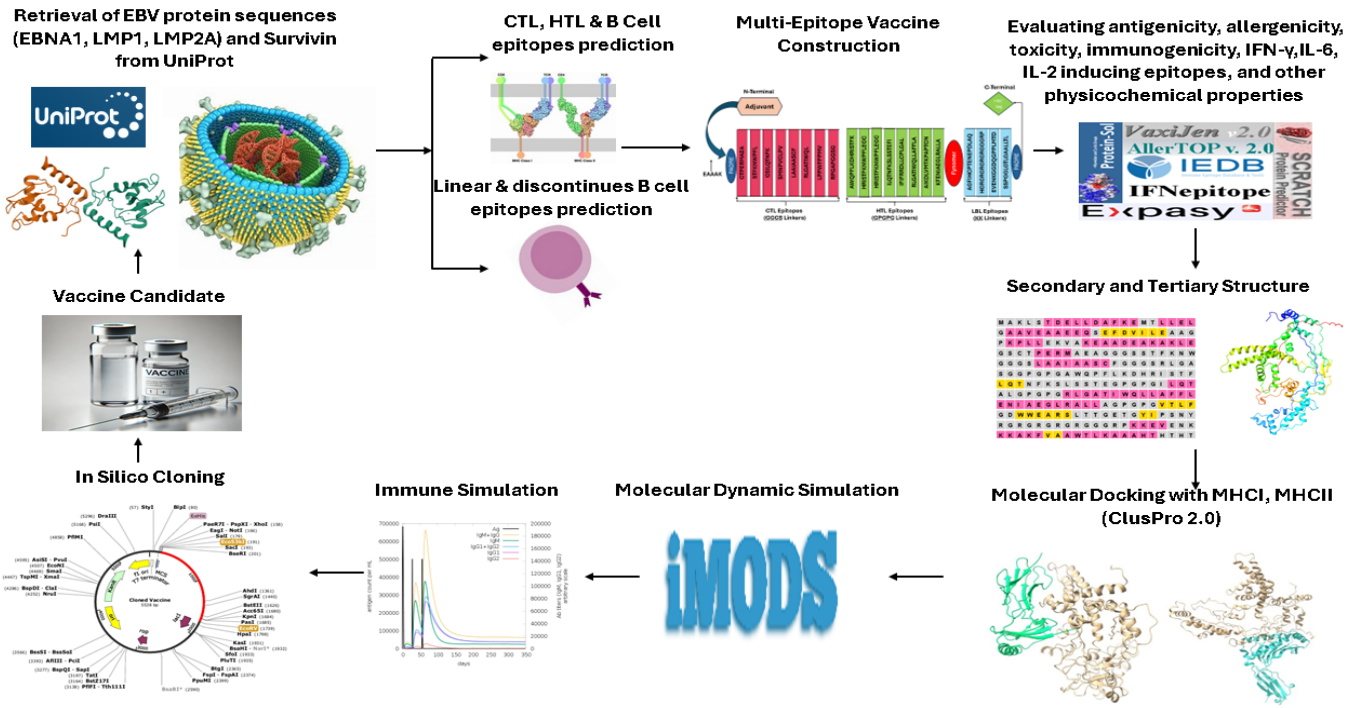

Figure 4: Schematic workflow of in silico multi-epitope vaccine design process.

Computational Tools and Bioinformatics Resources Utilized in the Study

Tool Name	URL
UniProt database	https://www.uniprot.org/
PDB database	https://www.rcsb.org
VaxiJen 2.0	https://ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
ANTIGENpro	http://scratch.proteomics.ics.uci.edu/
AllerTOP 2.0	https://ddg-pharmfac.net/allertop_test/
AllerCatPro 2.0	https://allercatpro.bii.a-star.edu.sg/
ToxinPred	https://webs.iiitd.edu.in/raghava/toxinpred/
NetCTL 1.2	https://services.healthtech.dtu.dk/services/NetCTL-1.2/
NetMHCcons-1.1	https://services.healthtech.dtu.dk/services/NetMHCcons-1.1/
NetMHCII version 2.3	https://services.healthtech.dtu.dk/services/NetMHCII-2.3/
IEDB	https://www.iedb.org/
MHC-I Binding (IEDB)	http://tools.iedb.org/mhci/
MHC-II Binding (IEDB)	http://tools.iedb.org/mhcii/
MHC-I Processing (IEDB)	http://tools.iedb.org/processing/
Epitope Conservancy Analysis (IEDB)	http://tools.iedb.org/conservancy/
Population Coverage (IEDB)	http://tools.iedb.org/population/
Class I Immunogenicity (IEDB)	http://tools.iedb.org/immunogenicity/
ElliPro	http://tools.iedb.org/ellipro/
IFNepitopes	https://webs.iiitd.edu.in/raghava/ifnepitope/
il2pred	https://webs.iiitd.edu.in/raghava/il2pred/
IL6-Pred	https://webs.iiitd.edu.in/raghava/il6pred/
ABCpred	https://webs.iiitd.edu.in/raghava/abcpred/
Disulfide by Design 2.0	http://cptweb.cpt.wayne.edu/DbD2/
ProtParam	https://web.expasy.org/protparam/
ProSA-web	https://prosa.services.came.sbg.ac.at/prosa.php/
Protein-Sol	https://protein-sol.manchester.ac.uk/
SolPro	http://scratch.proteomics.ics.uci.edu/
ERRAT	https://www.doe-mbi.ucla.edu/errat/
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/
Robetta server	https://robetta.bakerlab.org/

PDBsum server	https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/
GalaxyRefine	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
PROCHECK	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
SOPMA	https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
ClusPro 2.0	https://cluspro.org/
PRODIGY	https://rascar.science.uu.nl/prodigy/
iMODS	https://imods.iqf.csic.es/
JCat	https://www.jcat.de/
C-ImmSim	https://kraken.iac.rm.cnr.it/C-IMMSIM/
SnapGene software	https://www.snapgene.com/
UCSF ChimeraX software	https://www.cgl.ucsf.edu/chimera/

Results

I. Prediction of T-cell epitopes: CTL

Using the IEDB, potential CTL epitopes were identified for the four target proteins. The selection process involved rigorous criteria to ensure the suitability of the candidates. First, epitopes with strong binding affinity (IC50 < 200 nM) were filtered. Subsequent evaluations focused on ensuring non-allergenicity (AllerTop 2.0), non-toxicity (ToxinPred), and a high immunogenicity score (IEDB Immunogenicity Prediction Tool). Antigenicity was also assessed using VaxiJen 2.0 (threshold > 0.4).To guarantee reliability, conservation analysis confirmed that the selected epitopes were 100% conserved across all strains and isoforms. Additionally, processing predictions using IEDB's MHC-I Processing (Proteasome, TAP) tool validated their compatibility with MHC class I presentation pathways. After completing all analyses, the final selected epitopes, listed in (Table 1), were identified as the most promising candidates for vaccine design.

The identified CTL epitopes were further validated using NetCTL 1.2 and NetMHCcons tools. NetCTL 1.2 integrates predictions for MHC binding, proteasomal cleavage, and TAP transport, enabling comprehensive analysis of T-cell epitope processing and presentation. NetMHCcons combines multiple MHC-binding prediction algorithms to improve prediction accuracy through a consensus-based approach. Both tools confirmed that the selected epitopes showed strong to moderate binding affinity to at least one MHC supertype using

NetCTL 1.2 and strong to moderate binding affinity to nearly the same alleles using NetMHCcons. These results reinforce the reliability of the selected epitopes for vaccine design.

Table 1: The selected CTL epitopes for the final vaccine construct using IEDB server.

Protein	Epitope	Allele	Processing Score	Antigenicity	Allergenicity/Toxicity	Immunogenicity	Conservancy
Survivin (BIRC5)	[5]TLPPAWQPF	HLA-A*23:01 HLA-A*24:02 HLA-B*15:01	1.80	0.9674	NON-ALLERGEN / Non-Toxin	0.12761	100%
Survivin (BIRC5)	[33]CTPERMAEA	HLA-A*02:03 HLA-A*02:06 HLA-A*68:02	0.83	0.8084	NON-ALLERGEN / Non-Toxin	0.07377	100%
EBNA1	[389]SSSGSPPRR	HLA-A*68:01	1.36	0.8168	NON-ALLERGEN / Non-Toxin	-0.17026	100%
EBNA1	[53]RPGAPGGSG	HLA-B*07:02	0.06	1.1793	NON-ALLERGEN / Non-Toxin	0.33271	100%
LMP1	[132]RLGATIWQL	HLA-A*02:01 HLA-A*02:06 HLA-A*32:01 HLA-A*02:03 HLA-B*13:01	1.82	0.7761	NON-ALLERGEN / Non-Toxin	0.05043	100%
LMP2A	[136]LAAIAASCF	HLA-B*35:01 HLA-B*15:01	2.42	0.7048	NON-ALLERGEN / Non-Toxin	0.01872	100%
LMP2A	[373]GSILQTNFK	HLA-A*11:01 HLA-A*30:01 HLA-A*03:01 HLA-A*68:01	1.19	1.0565	NON-ALLERGEN / Non-Toxin	-0.03965	100%
LMP2A	[119]SMNPVCLPV	HLA-A*02:03 HLA-A*02:01 HLA-A*02:06	0.78	1.5822	NON-ALLERGEN / Non-Toxin	-0.03965	100%

II. Prediction of T-cell epitopes: HTL

The identification of potential HTL epitopes for the target proteins utilized the IEDB platform, applying a series of stringent selection criteria. Epitopes demonstrating strong binding affinity (IC50 < 200 nM) were prioritized, alongside assessments for antigenicity (VaxiJen 2.0, threshold > 0.4), non-allergenicity (AllerTop 2.0), non-toxicity (ToxinPred), and complete sequence conservation across all strains and isoforms. A unique aspect of this process was the prediction of cytokine induction. Epitopes were evaluated for their ability to stimulate key immune mediators, including IFN-γ, IL-6, and IL-2. This ensured the selection of epitopes capable of orchestrating a robust and balanced immune response. After completing all analyses, the final selected epitopes, listed in (Table 2), were identified as the most promising candidates for vaccine design.

Furthermore, the selected HTL epitopes were validated using NetMHCII version 2.3 to add an additional layer of confirmation. All the selected epitopes demonstrated moderate to strong binding affinities to at least one of the selected HLA class II alleles using this tool, further supporting their potential as robust vaccine candidates.

Table 2: The selected HTL epitopes for the final vaccine construct using IEDB server.

Protein	Epitope	Allele	Antigenicity	Allergenicity/ Toxicity	IFN-γ	IL-2/IL-6	Conservancy
Survivin (BIRC5)	[9]AWQPFLKDHRISTFK	HLA-DRB1*11:01 HLA-DRB1*03:01	0.9005	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Non-Inducer	100%
Survivin (BIRC5)	[17]HRISTFKNWPFLEGC	HLA-DRB1*15:01 HLA-DQA1*04:01/DQB1*04:02 HLA-DRB1*07:01 HLA-DPA1*01:03/DPB1*02:01 HLA-DRB1*01:01	1.3323	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Non-Inducer	100%
EBNA1	[476]PKFENIAEGLRALLA	HLA-DRB1*01:01 HLA-DRB5*01:01 HLA-DQA1*04:01/DQB1*04:02 HLA-DQA1*01:02/DQB1*06:02	0.4822	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Non-Inducer	100%
EBNA1	[530]QCRLTPLSRLPFGMA	HLA-DRB1*11:01 HLA-DRB1*12:01 HLA-DRB1*01:01 HLA-DRB4*01:01 HLA-DRB1*15:01	1.1843	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Non-Inducer	100%
LMP1	[69]IFIFRRDLLCPLGAL	HLA-DPA1*03:01/DPB1*04:02 HLA-DRB1*01:01 HLA-DRB1*03:01 HLA-DRB1*13:02 HLA-DRB3*01:01	0.7629	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Inducer	100%
LMP1	[132]RLGATIWQLLAFFLA	HLA-DPA1*01:03/DPB1*02:01 HLA-DPA1*01:03/DPB1*04:01 HLA-DPA1*03:01/DPB1*04:02	0.504	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Inducer	100%
LMP2A	[373]GSILQTNFKSLSSTE	HLA-DPA1*01:03/DPB1*02:01 HLA-DPA1*01:03/DPB1*04:01 HLA-DPA1*02:01/DPB1*01:01 HLA-DPA1*03:01/DPB1*04:02 HLA-DQA1*04:01/DQB1*04:02 HLA-DRB1*01:01	0.8826	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer , Non-Inducer	100%
LMP2A	[375]ILQTNFKSLSSTEFI	HLA-DQA1*04:01/DQB1*04:02 HLA-DRB1*01:01 HLA-DRB1*04:01 HLA-DRB1*04:05 HLA-DRB1*07:01 HLA-DRB1*09:01 HLA-DRB5*01:01	0.764	NON-ALLERGEN, Non-Toxin	POSITIVE	Inducer, Non-Inducer	100%

III. Prediction of Linear B-cell Epitopes

Using the ABCpred Prediction Server, linear B-cell epitopes were identified based on a high-confidence threshold of 0.8. These epitopes were evaluated for their antigenicity (VaxiJen 2.0, threshold > 0.4), non-allergenicity (AllerTop 2.0)and non-toxicity (ToxinPred), ensuring their safety and immunogenic

potential. Conservation analysis confirmed that the selected epitopes were completely conserved across protein isoforms and virus strains, making them suitable candidates for vaccine development. After completing all analyses, the final selected epitopes, listed in (Table 3), were identified as the most promising candidates for further experimental validation and inclusion in vaccine formulations.

Table 3: The selected LBL epitopes for the final vaccine construct using by ABCpred.

Protein	Epitope	ABCpred Score	Antigenicity	Allergenicity/Toxicity	Conservancy
Survivin (BIRC5)	[41]AGFIHCPTENEPDLAQ	0.89	0.4732	NON-ALLERGEN, Non-Toxin	100%
EBNA1	[39]HGRGRGRGRGRGGGRP	0.94	1.2239	NON-ALLERGEN, Non-Toxin	100%
LMP1	[326]EVENKGGDQGPPLMTD	0.85	0.8124	NON-ALLERGEN, Non-Toxin	100%
LMP2A	[289]SSPGGLGTLGAALLTL	0.85	0.8574	NON-ALLERGEN, Non-Toxin	100%

IV. Population Coverage

The selected CTL and HTL epitopes demonstrated strong global immunological protection potential, achieving a remarkable global coverage rate of 99.96% (Figure 5) and excellent regional coverage in endemic areas: East Asia (99.64%), North Africa (99.62%), South Asia (99.98%), and Southeast Asia (97.82%). Average epitope hits in these regions were 7.25 in East Asia, 6.82 in North Africa, 9.62 in South Asia, and 4.26 in Southeast Asia, reflecting robust immunological protection. Consistent performance was evident in the HLA combinations recognized by 90% of the population (pc90), which averaged 4.64 with minimal variability (standard deviation: 4.11% for coverage) (Table 4). These findings underscore the vaccine's robust design and its potential to deliver broad and effective immunological protection, particularly in regions with the highest prevalence of cases, as well as across diverse global populations.

MHC class	Coverage	Average hit	PC90
combined	99.96%	10.12	6.15

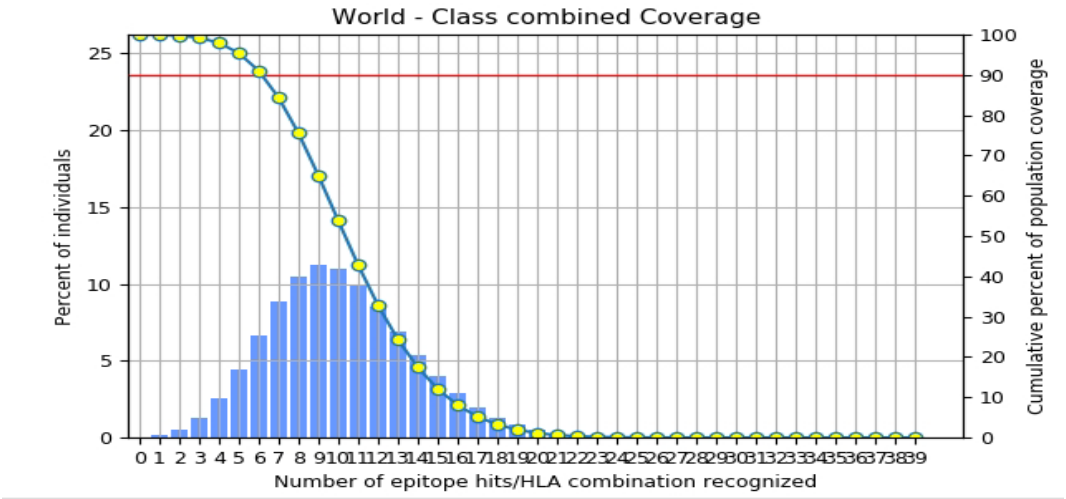

Figure 5: **Global coverage for the selected epitopes for the vaccine construct.** The chart represents the number of epitope-HLA combinations recognized by individuals across different populations. The x-axis shows the number of epitope-HLA combinations recognized, while the left y-axis indicates the percentage of individuals recognizing a given number of combinations. The blue bars depict the distribution of individuals, with most recognizing between 5 and 12 epitope-HLA combinations, peaking around 10 combinations. The right y-axis represents the cumulative percentage of population coverage, shown by the green curve with yellow markers. The red horizontal line at 90% cumulative population coverage (PC90) indicates that at least 6 epitope-HLA combinations are recognized in 90% of individuals. Similarly, approximately 80% of the population recognizes 8 or more combinations, while 50% of the population recognizes 11 or more. The total population coverage for Class I and Class II epitopes combined is 99.96%, demonstrating the broad immune response potential of the vaccine construct.

Table 4: Population Coverage by Region.

Population/Area	Class Combined		
	Coverage	Average Hit	pc90
Central Africa	99.99%	10.4	6.88
Central America	97.67%	6.85	3.15
East Africa	99.99%	10.65	6.57
East Asia	**99.64%**	**7.25**	**3.55**
Europe	100.0%	11.63	8.11
North Africa	**99.62%**	**6.82**	**3.87**
North America	100.0%	12.2	8.36

Northeast Asia	99.0%	7.47	4.35
Oceania	99.81%	7.71	4.41
South Africa	82.75%	6.5	3.1
South America	99.92%	10.2	5.57
South Asia	**99.98%**	**9.62**	**6.26**
Southeast Asia	**97.82%**	**4.26**	**1.85**
Southwest Asia	99.05%	5.91	3.21
West Africa	99.42%	8.5	4.66
West Indies	98.85%	5.8	2.49
Average	98.42%	7.82	4.64
Standard Deviation	4.11	2.8	2.16

V. Combining the final multi-epitope vaccine construct

The finalized multi-epitope vaccine construct demonstrated a rational and well-structured linear design incorporating essential immunogenic and stabilizing components. The construct consists of 514 amino acids, including 8 CTL epitopes, 8 HTL epitopes, and 4 linear B-cell epitopes, strategically selected to ensure broad and robust immune responses. The inclusion of PADRE enhanced immunogenicity by promoting a broad CD4+ T-cell response, ensuring the vaccine's efficacy across diverse populations. The rpIL adjuvant effectively boosted immune activation, while the fynomer sequence contributed to improved molecular interactions and stability of the construct. Linkers (EAAAK and KK) successfully ensured proper spatial arrangement of the epitopes, and the H5E tag facilitated efficient purification and detection. The integration of these components resulted in a stable and functional vaccine construct, as represented in the schematic shown in (Figure 6).

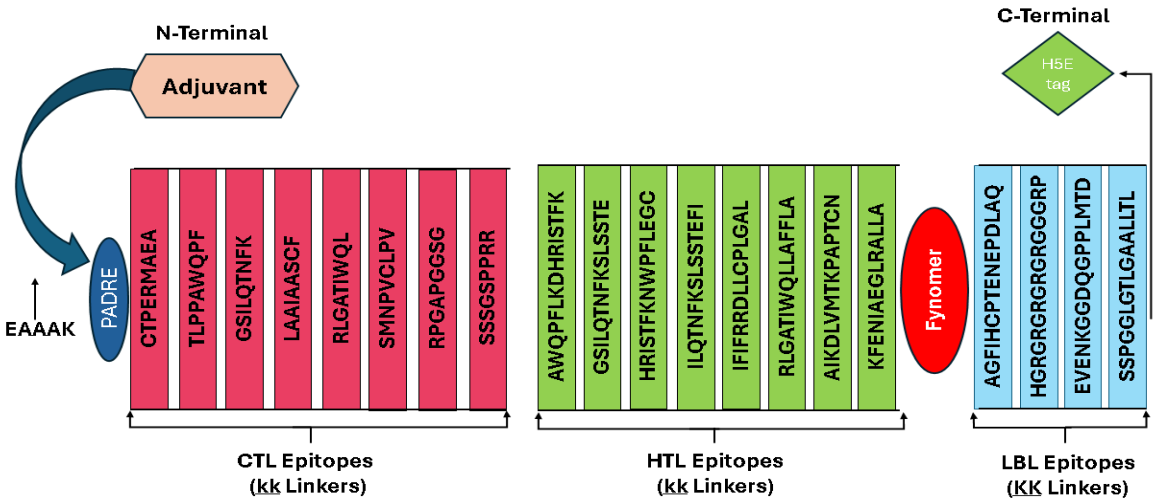

Figure 6: **Graphical representation of the combined vaccine construct**.

VI. Assessment of Antigenicity, Allergenicity, Toxicity, Solubility, and Physicochemical Characteristics of the Vaccine

The vaccine construct exhibited a good antigenicity score of 0.550 using VaxiJen v2.0 and 0.641 using ANTIGENpro, confirming its ability to elicit an immune response. AllerCatPro 2.0 analysis confirmed the vaccine's non-allergenic nature, while ToxinPred analysis ensured no toxic components were present in the final construct. Solubility scores of 0.842 (SolPro) and 0.629 (Protein-sol) indicate excellent solubility.

The physicochemical properties of the vaccine construct were analyzed using the ProtParam tool from the Expasy web server, developed by the Swiss institute of bioinformatics (SIB). The construct consists of 514 amino acids with a molecular weight of 56.00 kDa and a theoretical pI of 9.83, indicating its essential nature. The construct contains 92 positively charged residues and 54 negatively charged residues, yielding a net positive charge favorable for stability in physiological environments. This pI and positive net charge at physiological pH enhances its solubility, interaction with negatively charged molecules, and potential effectiveness as part of a vaccine. The extinction coefficient was determined as 50,335 M^{-1} cm^{-1} assuming all pairs of cysteines forming cystines, and 49,960 M^{-1} cm^{-1} assuming reduced cysteines. The instability index of 31.86 classified the construct as stable, and the aliphatic index of 79.11 indicated thermostability. The GRAVY score of −0.376 suggested hydrophilicity and enhanced solubility in aqueous environments.

Collectively, these results support the vaccine construct's stability, solubility, and suitability for inducing an effective immune response, meeting the criteria for vaccine formulation.

VII. Secondary structure prediction

The secondary structure of the vaccine was predicted using PSIPRED. The results of this prediction are shown in (Figure 7(A,B)), where alpha helices are represented in pink, beta strands in yellow, and coils in gray. Additionally, the SOPMA tool was utilized to further evaluate the vaccine's secondary structure. The SOPMA analysis revealed that 40.86% of the residues form alpha helices, 10.70% form extended strands, and 48.44% are random coils (Figure 7C). These findings indicate that the vaccine construct contains a balanced proportion of structural elements, which supports its stability and flexibility for effective immune interactions. Furthermore, the presence of random coils is advantageous

for epitope presentation, as these regions are typically more accessible for immune recognition, facilitating the efficient display of epitopes to immune cells.

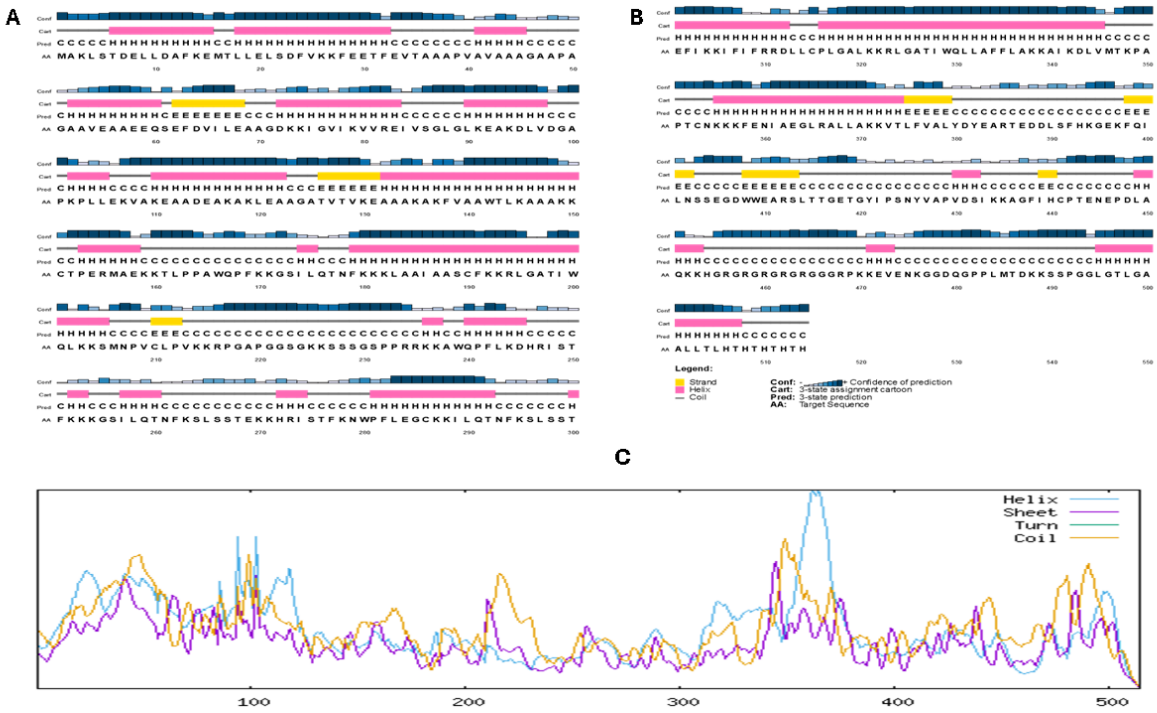

Figure 7: Predicted Secondary Structure Analysis of the Vaccine Construct. (A, B) Graphical representation of the secondary structure of the vaccine construct predicted by the PSIPRED server, where yellow represents strands, pink represents helices, and gray represents coils. (C) Distribution of alpha helices, beta strands, random coils, and turns across the vaccine construct using SOPMA tool.

VIII. Modeling, Refinement, and Validation of the 3D Structure

After generating the 3D vaccine model using the Robetta server, the first model was selected as the best prediction. Following refinement of the 3D structural model using the GalaxyRefine server, Model 3 was chosen as the optimal final vaccine model based on key parameters, including GDT-HA (0.9728), root-mean-square deviation (RMSD) (0.358), Mol Probity (1.472), clash score (8.8), and poor rotamers (0.2), among other refined models (Figure 9B and Table 5). The Z-score for the vaccine structure in the ProSA web server was −6.63 indicates the structure is within the range of native-like proteins, which are naturally occurring and correctly folded proteins (Figure 9C, 9D).

The overall quality of the refined model, as predicted by the ERRAT server, was determined to be 94.35. Stereochemical quality analysis of the refined protein structure was conducted using the Ramachandran plot via the PROCHECK tool

available on the PDBsum server. The analysis revealed that 96.6% of the residues were in the most favorable regions, 3.4% in the additionally allowed regions, no residues in the generously allowed regions or disallowed regions (Figure 9A) which confirms the reliability of this model.

Table 5: Validation metrics for refined 3D structure models using GalaxyRefne server.

Model	GDT-HA	RMSD	Mol Probity	Clash score	Poor rotamers	Rama favored
MODEL 1	0.9674	0.354	1.540	10.5	0.2	99.2
MODEL 2	0.9742	0.340	1.472	8.8	0.5	99.0
MODEL 3	**0.9728**	**0.358**	**1.472**	**8.8**	**0.2**	**99.2**
MODEL 4	0.9781	0.344	1.540	10.5	1.0	98.8
MODEL 5	0.9757	0.343	1.492	9.3	0.7	99.4

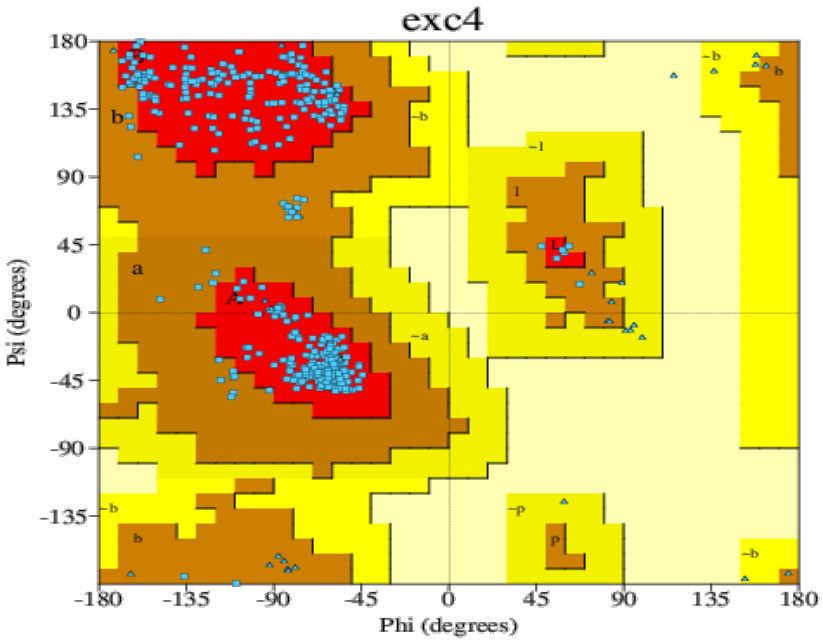

Figure 8: Ramachandran Plot Analysis of the Vaccine Construct Before Refienment. The results indicate that 93.5% of residues fall within the most favored regions, while 6.5% are in additional allowed regions. Notably, no residues are found in the generously allowed or disallowed regions.

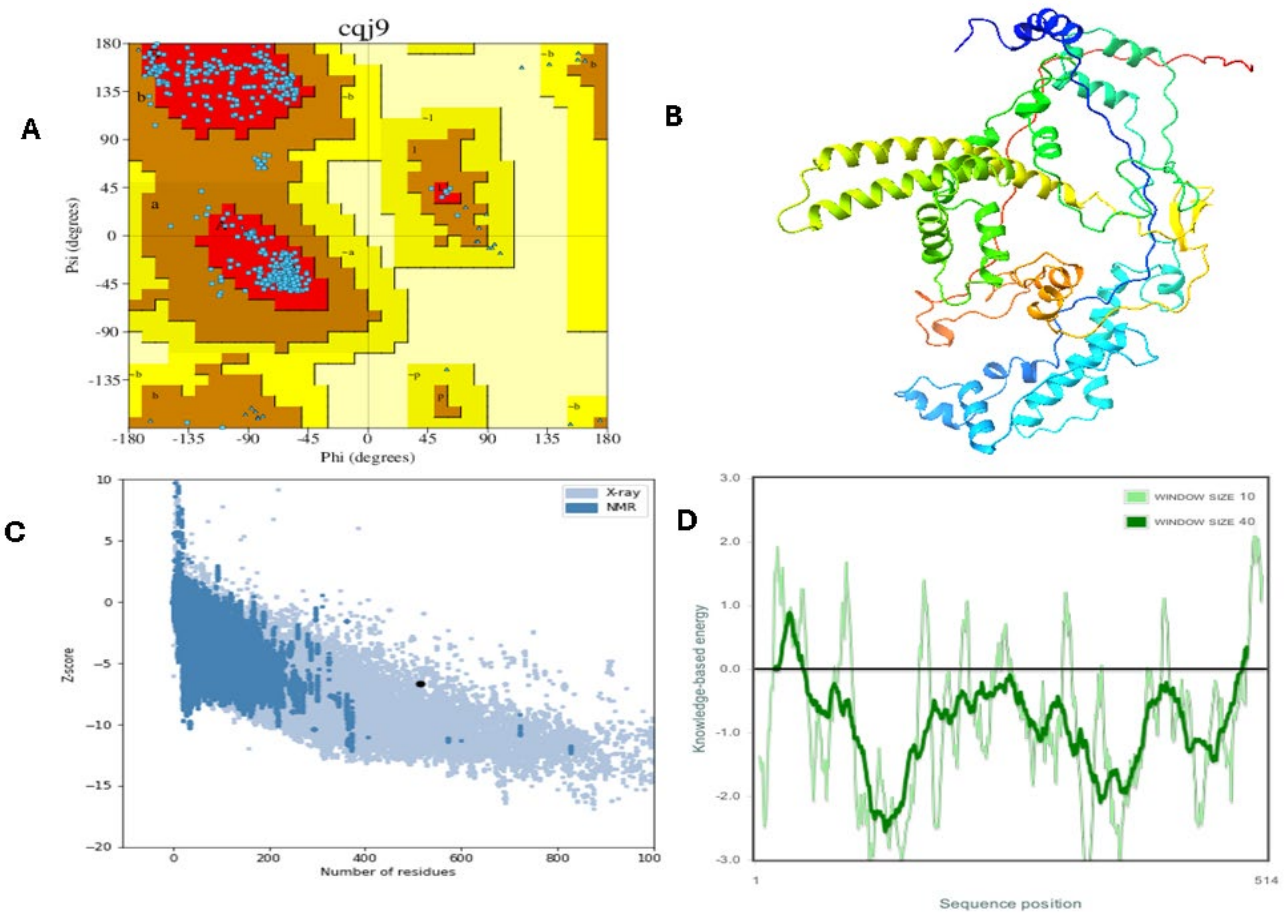

Figure 9: Structural Validation and Quality Assessment of the Vaccine Construct After Refienment. (A) Analysis of the Ramachandran plot after refienment utilizing the PROCHECK server showed 96.6%, and 3.4% residues laying in favored, additional allowed, respectively. (B) The three-dimensional refined vaccine. (C) ProSA validation of predicted structure with Z-score of −6.63 and (D) plots the residues scores to check the local model quality.

IX. Prediction of Conformational B-Cell Epitopes

Based on the results obtained from the ElliPro server, 5 conformational B-cell epitopes were identified from the refined three-dimensional (3D) vaccine model. The epitopes' amino acid residues, number of residues, and their respective scores are detailed in (Table 6). Additionally, a graphical representation of these epitopes mapped onto the 3D vaccine structure is presented (Figure 10), highlighting their spatial distribution and potential accessibility for immune interactions.

Table 6: List of conformational/discontinuous B-cell epitopes predicted over final vaccine construct.

NO	Residue	Number of residues	Score
1	A:A340, A:I341, A:K342, A:D343, A:L344, A:V345, A:M346, A:T347, A:K348, A:P349, A:A350, A:P351, A:T352, A:C353, A:N354, A:K355, A:K356, A:K357, A:F358, A:E359, A:N360, A:I361	22	0.943
2	A:L195, A:G196, A:I199, A:W200, A:K203, A:K204, A:S205, A:M206, A:N207, A:T497, A:L498, A:G499, A:A500, A:A501, A:L502, A:L503, A:T504, A:L505, A:H506, A:T507, A:H508, A:T509, A:H510, A:T511, A:H512, A:T513, A:H514	27	0.903
3	A:A56, A:Q60, A:S61, A:E62, A:F63, A:D64, A:V65, A:I66, A:L67, A:E68, A:A69, A:A70, A:G71, A:D72, A:K73, A:I75, A:G76, A:V80, A:E83, A:I84, A:A123, A:G124, A:A125, A:T126, A:V127, A:T128, A:K130	27	0.875
4	A:M1, A:A2, A:K3, A:L4, A:S5, A:E8	6	0.874
5	A:S86, A:G87, A:L88, A:G89, A:L90	5	0.861

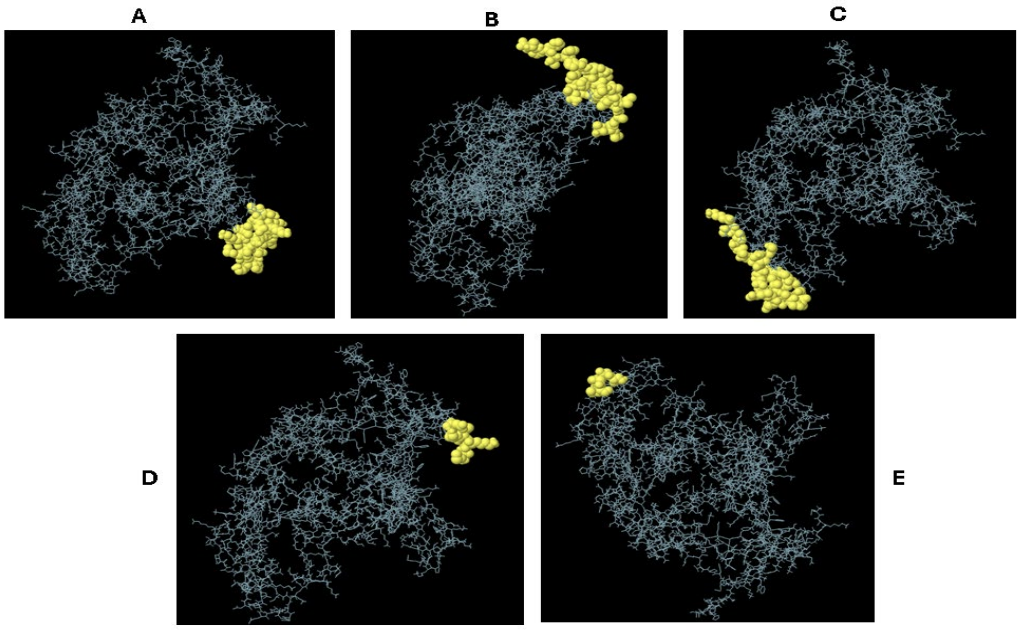

Figure 10: Graphical representation of the five discontinuous B-cell epitopes mapped onto the 3D vaccine structure.

X. Disulfide engineering of the vaccine construct

Using the Disulfide by Design 2 (DbD2) server, a total of 43 residue pairs in the vaccine construct were identified as potential candidates for disulfide bond formation (Table 7). These candidates were selected based on proximity, geometric constraints, and energy favorability, which are key parameters for predicting viable disulfide bonds. After applying Chi3 torsion angle constraints and energy thresholds, only three residue pairs (220-PRO-229-SER, 156-MET-162-LEU, and 218-GLY-222-GLY) met the criteria for stable disulfide bond formation. These residues were identified as optimal sites for potential disulfide

engineering if mutated to cysteine, which could enhance structural stability and robustness for downstream applications. However, for this study, the vaccine construct will proceed with its original sequence without introducing these mutations.

Table 8: List of residue pairs in the vaccine construct that have the ability to create disulfide bonds.

Res1 Chain	Res1 Seq #	Res1 AA	Res2 Chain	Res2 Seq #	Res2 AA	Chi3	Energy	Sum B-Fact
A	**220**	**PRO**	**A**	**229**	**SER**	**-83.33**	**0.34**	**0**
A	**156**	**MET**	**A**	**162**	**LEU**	**86.16**	**1.74**	**0**
A	**218**	**GLY**	**A**	**222**	**GLY**	**94.59**	**2.08**	**0**
A	337	ALA	A	362	ALA	-93.06	0.54	0
A	330	GLN	A	369	LEU	-102.71	1.47	0
A	43	ALA	A	147	ALA	-93.63	1.62	0
A	45	ALA	A	142	TRP	104.13	1.67	0
A	297	LEU	A	301	GLU	101.79	1.84	0
A	334	PHE	A	365	LEU	-109.12	1.92	0
A	172	GLY	A	214	LYS	71.47	2.35	0
A	324	LEU	A	328	ILE	122.4	2.4	0
A	113	ALA	A	136	ALA	80.02	2.43	0
A	271	LYS	A	274	ILE	119.74	2.5	0
A	107	LYS	A	423	PRO	94.2	2.67	0
A	17	THR	A	255	GLY	94.71	2.97	0
A	33	GLU	A	217	PRO	81.83	3.29	0
A	94	LYS	A	115	GLU	117.54	3.54	0
A	96	LEU	A	111	GLU	-91.84	3.62	0
A	316	PRO	A	319	ALA	119.31	3.76	0
A	43	ALA	A	146	ALA	-74.68	3.78	0
A	178	ASN	A	181	LYS	116.23	4.05	0
A	153	PRO	A	165	ALA	95.63	4.14	0
A	47	ALA	A	139	VAL	124.93	4.19	0
A	263	LYS	A	492	PRO	114.41	4.3	0
A	117	LYS	A	132	ALA	-100.28	4.7	0
A	264	SER	A	491	SER	99.99	4.72	0
A	4	LEU	A	8	GLU	90.78	4.8	0
A	32	PHE	A	226	LYS	-115.99	5.2	0
A	109	ALA	A	139	VAL	122.37	5.21	0
A	224	GLY	A	227	SER	111.22	5.23	0
A	212	PRO	A	406	GLY	-85.67	5.29	0
A	242	PHE	A	379	LEU	104.51	5.39	0
A	49	PRO	A	116	ALA	-80.03	5.44	0
A	109	ALA	A	140	ALA	-82.14	5.48	0
A	112	ALA	A	136	ALA	-113.29	5.52	0
A	50	ALA	A	54	VAL	79.46	5.57	0
A	321	LYS	A	329	TRP	88.33	5.62	0
A	469	PRO	A	480	GLN	-103.79	5.66	0
A	228	SER	A	231	SER	117.81	5.78	0
A	189	SER	A	259	GLN	-88.29	6.5	0
A	120	LEU	A	125	ALA	125.04	8.03	0
A	399	GLN	A	413	SER	103.27	8.09	0
A	277	PHE	A	373	VAL	121.59	8.21	0

XI. Molecular Docking and Binding Affinity Analysis of the Vaccine Construct with MHC-I, and MHC-II Receptors

The molecular docking analysis, performed using ClusPro 2.0 server, demonstrated stable and detailed binding interactions between the vaccine construct and the target receptors: MHC-I and MHC-II. The PDBsum results provided a comprehensive overview of the interface areas, number of interacting residues, and interaction types, further validating the strength and specificity of the docking outcomes (Figures 11, 12).

In the MHC I-vaccine complex, 34 vaccine residues interacted with 36 MHC-I residues, forming interface areas of 1805 Å² for the vaccine and 1755Å² for MHC-I. This interaction was supported by 3 salt bridges, 23 hydrogen bonds, and 247 non-bonded contacts, demonstrating significant binding stability.

For the MHC II-vaccine complex, 36 vaccine residues interacted with 31 MHC-II residues, creating interface areas of 1825 Å² for the vaccine and 1876Å² for MHC-II. The binding involved 9 salt bridges, 24 hydrogen bonds, and 252 non-bonded contacts, suggesting a highly stable and energetically favorable interaction.

The binding affinities were assessed using the PRODIGY server (Table 8), which yielded ΔG values of −15.4 kcal/mol for MHC-I, and −13.3 kcal/mol for MHC-II. The corresponding dissociation constants (Kd) were 1.40E-11 M, and 3.90E-10 M, respectively, indicating strong and stable interactions for the two complexes.

These results underscore the vaccine construct's potential to form highly stable and energetically favorable interactions with immune receptors, supporting its efficacy in stimulating an immune response against EBV.

Table 9: The predicted binding affinities of the docked complexes of the vaccine with MHC-I and MHC-II, using PRODIGY server.

Protein-protein complex	Gibbs free energy ΔG (kcal mol-1)	Kd (M) at °C
Vaccine-MHC class I receptor	-15.4	1.40E-11
Vaccine-MHC class II receptor	-13.3	3.90E-10

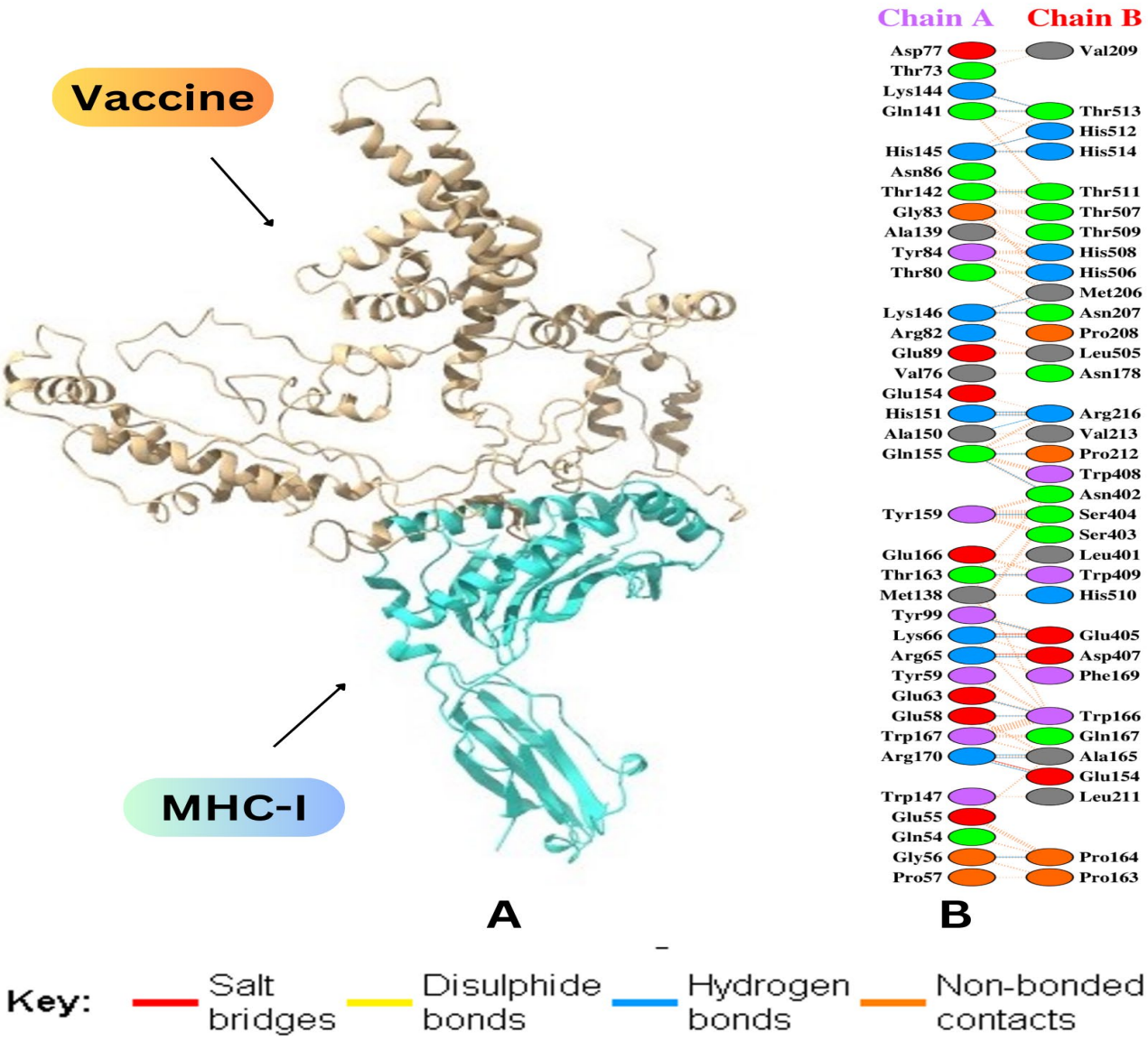

Figure 11: **Molecular Docking Analysis of the Vaccine-MHC-I Complex.** (A) Visualization of docking results for the vaccine-MHC-I complex. The vaccine construct is shown in gold, while MHC-I is depicted in Turquoise. (B) Map of total interacting residues and bonds between the vaccine (Chain B) and MHC-I (Chain A) proteins.

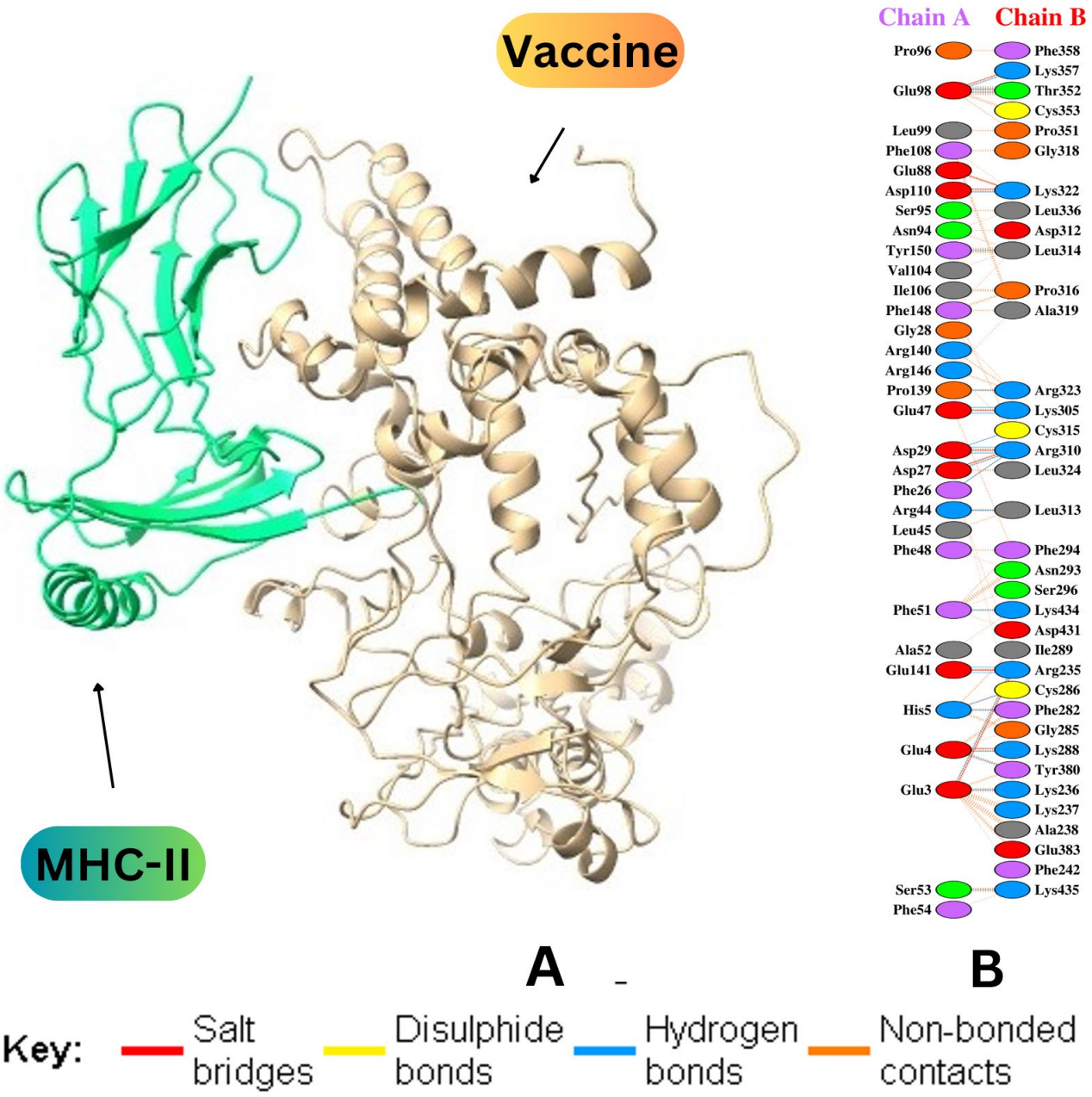

Figure 12: **Molecular Docking Analysis of the Vaccine-MHC-II Complex.** (A) Visualization of docking results for the vaccine-MHC-II complex. The vaccine construct is shown in gold, while MHC-II is depicted in green. (B) Map of total interacting residues and bonds between the vaccine (Chain B) and MHC-II (Chain A) proteins.

XII. Molecular dynamics simulation of the vaccine-receptor complexes

MD simulation, performed using the iMODS server, provided comprehensive insights into the structural and functional dynamics of the vaccine-receptor complexes involving MHC-I and MHC-II receptors. Deformability analysis (Figures 13B, 14B) identified key flexible regions with hinge points critical for

structural transitions and binding adaptability. The B-factor profiles (Figures 13D, 14D) highlighted mobile and rigid regions within the complexes, correlating well with experimental data. Eigenvalue analyses (Figures 13E, 14E), low eigenvalue for the first mode reflects the structure's adaptability, while the increasing values for higher modes indicate progressively rigid motions. Variance analyses (Figures 13C, 14C) demonstrated that the first few modes captured most of the structural motion, emphasizing the dominance of global dynamics in the complexes.

Covariance mapping (Figures 13F, 14F) illustrates dynamic correlations between residues, with red regions indicating positively correlated motions, blue regions showing anti-correlated movements, and white areas representing uncorrelated motions. These coordinated dynamics are essential for maintaining stable binding. Additionally, elastic network models (Figures 13G, 14G) revealed densely connected regions, representing tightly interconnected parts of the molecule critical for maintaining structural stability and integrity. Conversely, sparsely connected regions were identified in flexible areas, such as loops, terminal ends, or regions undergoing functional conformational changes, highlighting their role in facilitating efficient receptor binding and dynamic molecular interactions.

Together, these analyses confirm the vaccine-receptor complexes' structural stability, flexibility, and adaptability, providing a robust framework for optimizing vaccine design to enhance immune responses.

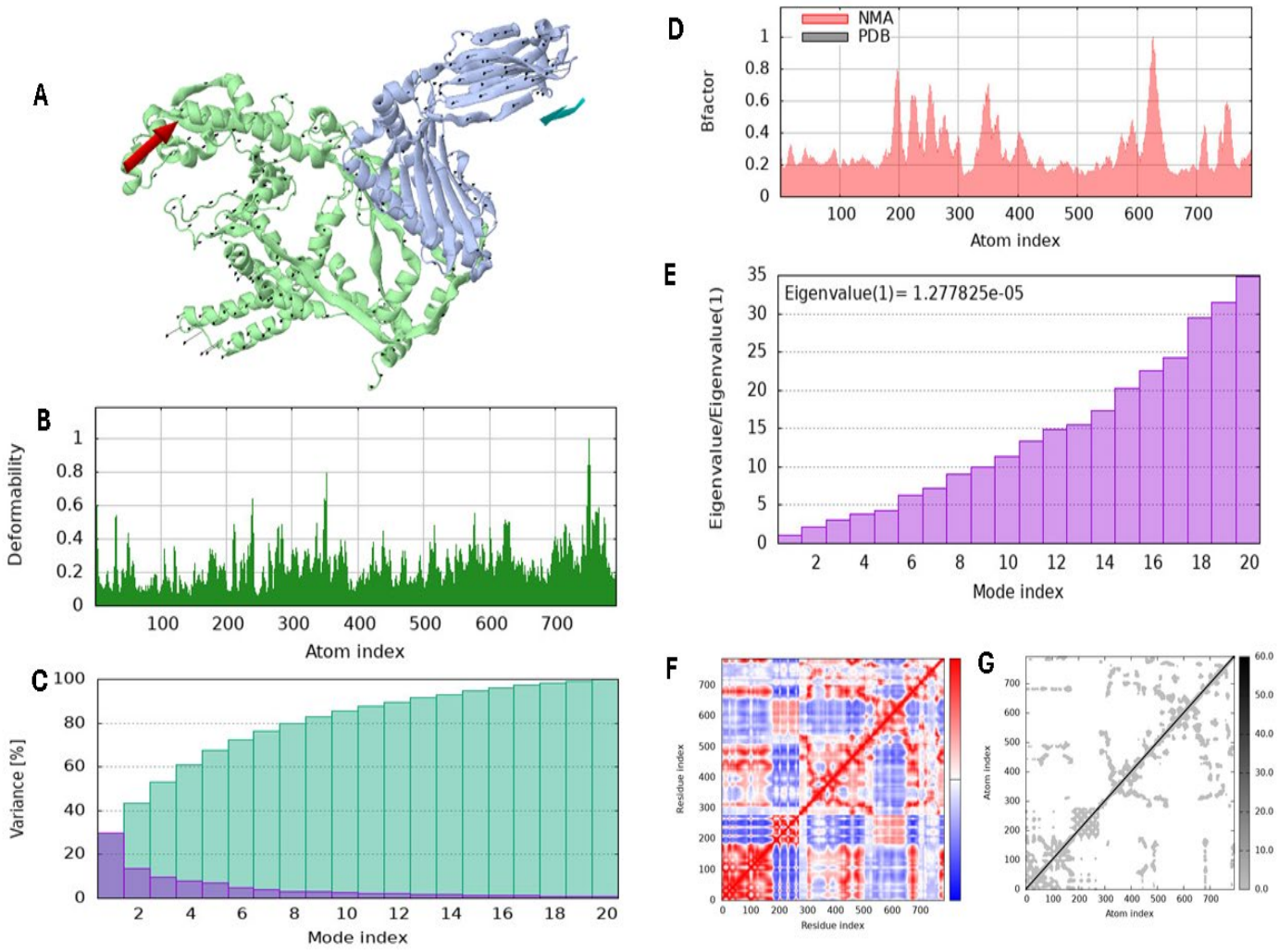

Figure 13: **Molecular dynamics simulation of vaccine construct and MHC-I complex by iMODS server.** (A) vaccine construct and MHC-I docking complex. (B) Main-chain deformability. (C) B-factor values. (D) The eigenvalue. (E) Variance. (F) Co-variance map. (G) Elastic network of model.

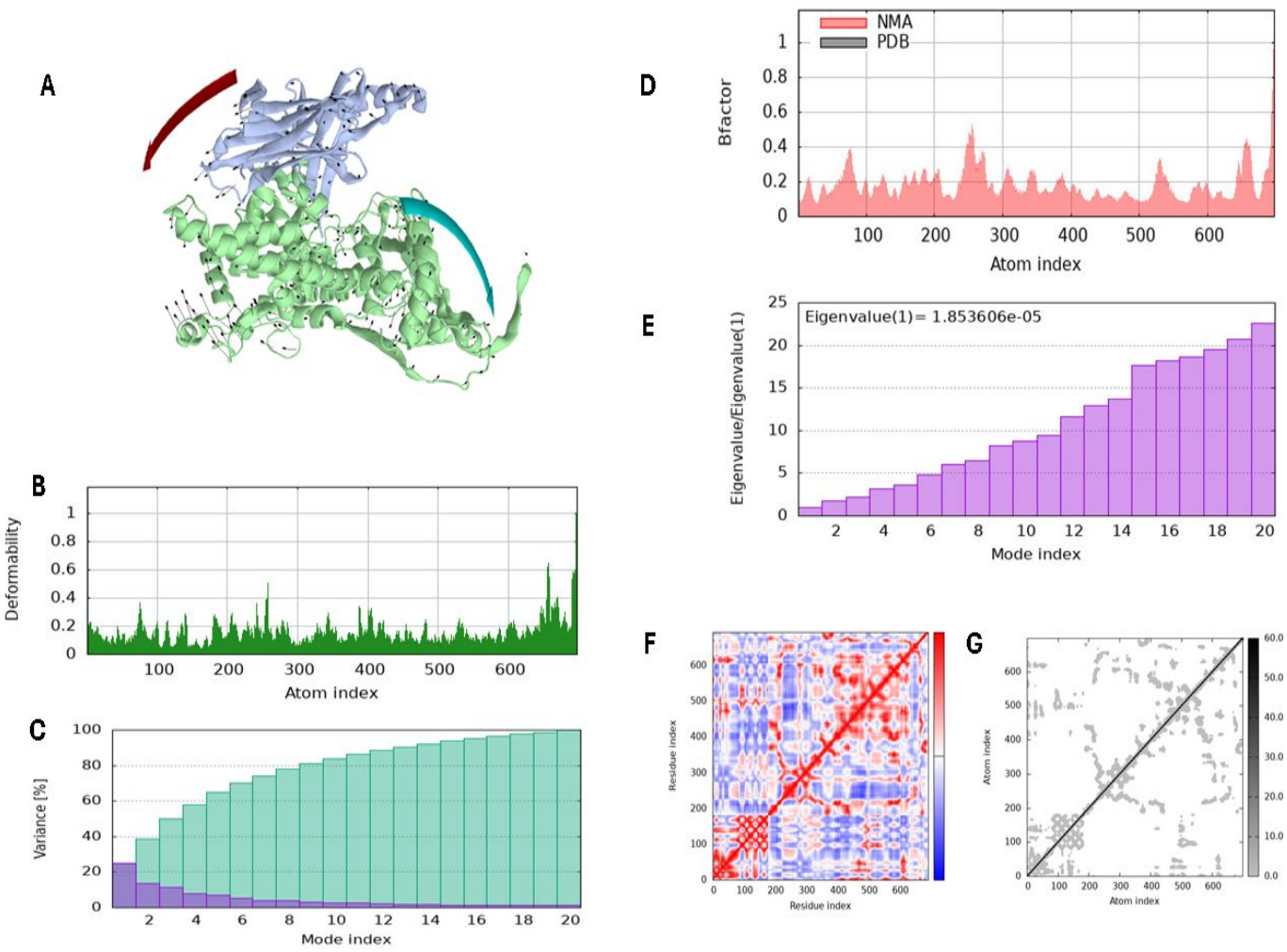

Figure 14: **Molecular dynamics simulation of vaccine construct and MHC-II complex by iMODS server.** (A) vaccine construct and MHC-II docking complex. (B) Main-chain deformability. (C) B-factor values. (D) The eigenvalue. (E) Variance. (F) Co-variance map. (G) Elastic network of model.

XIII. Codon optimization and in-silico vaccine cloning in pET-28a(+) expression vector

Codon optimization of the final vaccine sequence using JCat generated a 1542 bp cDNA sequence, achieving a Codon Adaptation Index (CAI) of 1.0, indicating optimal adaptation for E. coli expression. The GC content was 49.09%, which is within the optimal range for E. coli expression. These results suggest that the codon-optimized vaccine sequence is well-suited for high expression in E. coli, enhancing the efficiency of protein production for vaccine development. To facilitate cloning, XhoI and BamHI restriction sites were introduced at the 5' and 3' ends of the sequence.

Using SnapGene, in silico cloning successfully integrated the optimized vaccine sequence into the pET-28a(+) vector, generating a recombinant plasmid construct for expression. Virtual restriction digestion and gel electrophoresis simulation confirmed the expected fragment sizes: 5329 bp vector backbone and 1548 bp vaccine insert, validating that XhoI and BamHI did not cut within the insert.

These results confirm the accuracy of the in silico cloning strategy and the proper design of the recombinant plasmid construct, ensuring its suitability for downstream applications in vaccine development and high-level protein expression in E. coli.

The output from SnapGene software shown in (Figure 15), illustrating the cloning of the vaccine sequence into the pET-28a(+) vector.

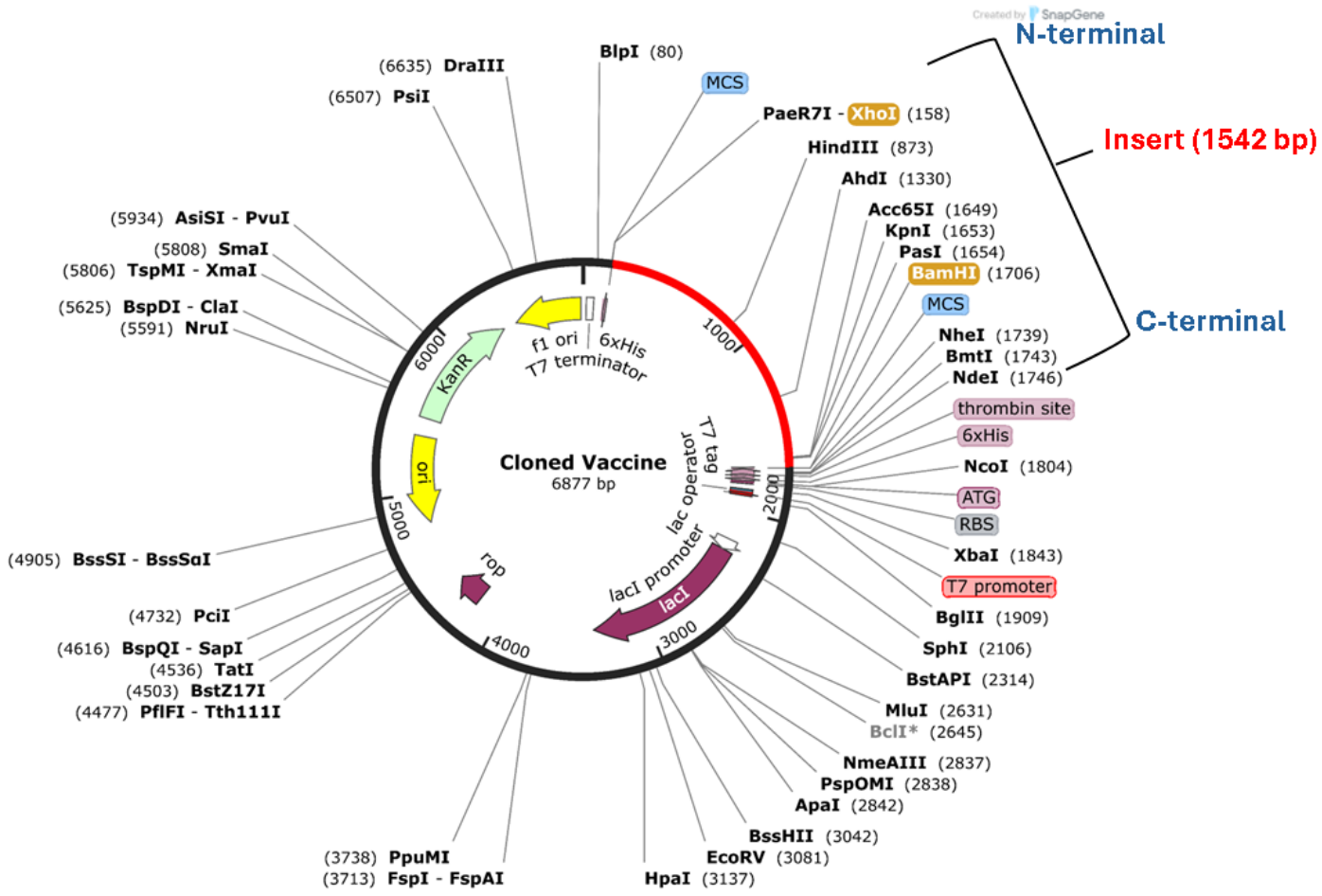

Figure 15: **In silico restriction cloning of the vaccine construct into the pET-28a(+) vector by SnapGene tool.** Red bar represents the codon-optimized vaccine gene, while the black circle depicts the vector backbone**.**

XIV. Immune simulation

Based on the results from the C-IMMSIM immune simulation analysis, administration of the vaccine demonstrated a robust induction of immune responses through several key markers. The initial injection elicited a noticeable increase in IgM levels, indicative of a primary immune response. Subsequent doses, as evidenced in the simulations, enhanced IgG1, IgG2, and IgG + IgM concentrations, particularly after the second and third injections. This pattern reflects the typical secondary immune response, characterized by higher affinity antibodies due to class switching. Over time, a decline in these immunoglobulins was observed, aligning with the contraction phase of the immune response; however, this does not imply diminished protection, as memory T and B cells were effectively generated.

In addition to humoral responses, a significant activation of CTLs, HTLs, and memory T cells was observed, showcasing the vaccine's capability to invoke cellular immunity. Dendritic cells, NK cells, and macrophages exhibited increased activity post-exposure, contributing to antigen processing and the orchestration of adaptive immunity. Cytokine profiling revealed elevated levels of IFN-γ, IL-12, and IL-10, markers of a balanced and robust immune response, essential for pathogen clearance and long-term immunity. Overall, the simulations confirm that three doses generated sufficient immunogenic response. However, if the vaccine is administered through 12 consecutive injections, the immune system will respond more strongly. essential for enhancing anti-tumor responses in cancer immunotherapy.

The following figures (Figures 16 to 20) represent the in silico simulation of the immune response triggered by the designed vaccine after three subsequent injections, illustrating key aspects of humoral, cellular, and innate immune activation.

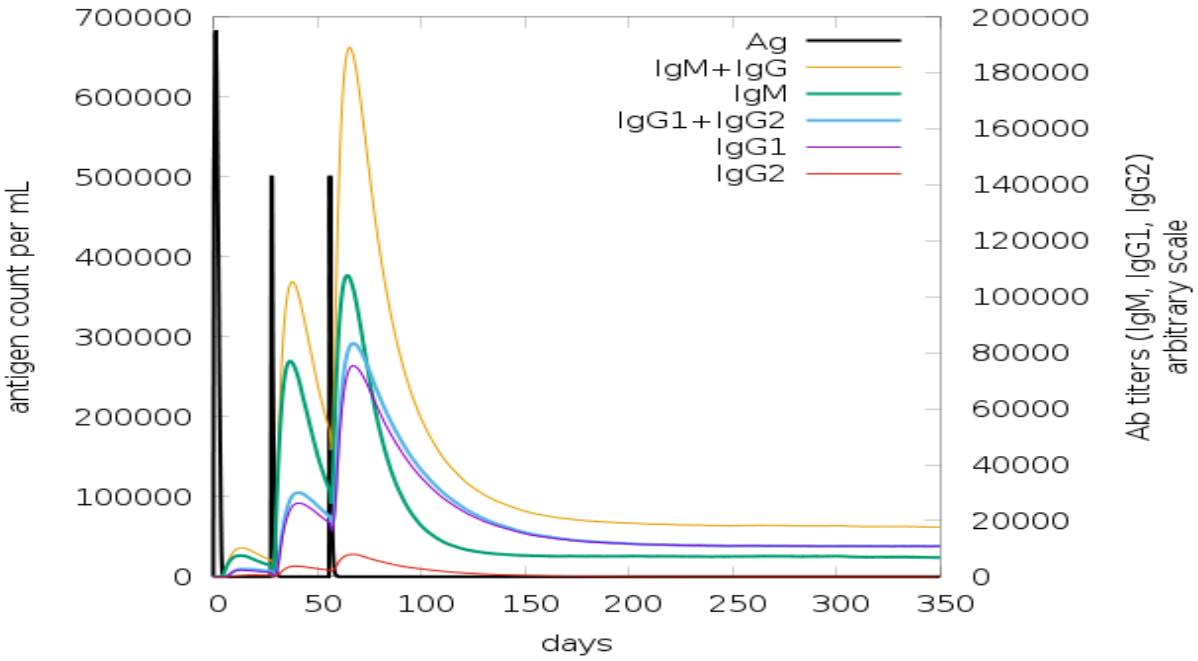

Figure 16: Humoral Immune Response Simulation.
Antigen count and antibody titers, including IgM, IgG, and IgG subtypes, indicating the activation of primary, secondary, and tertiary immune responses after vaccination.

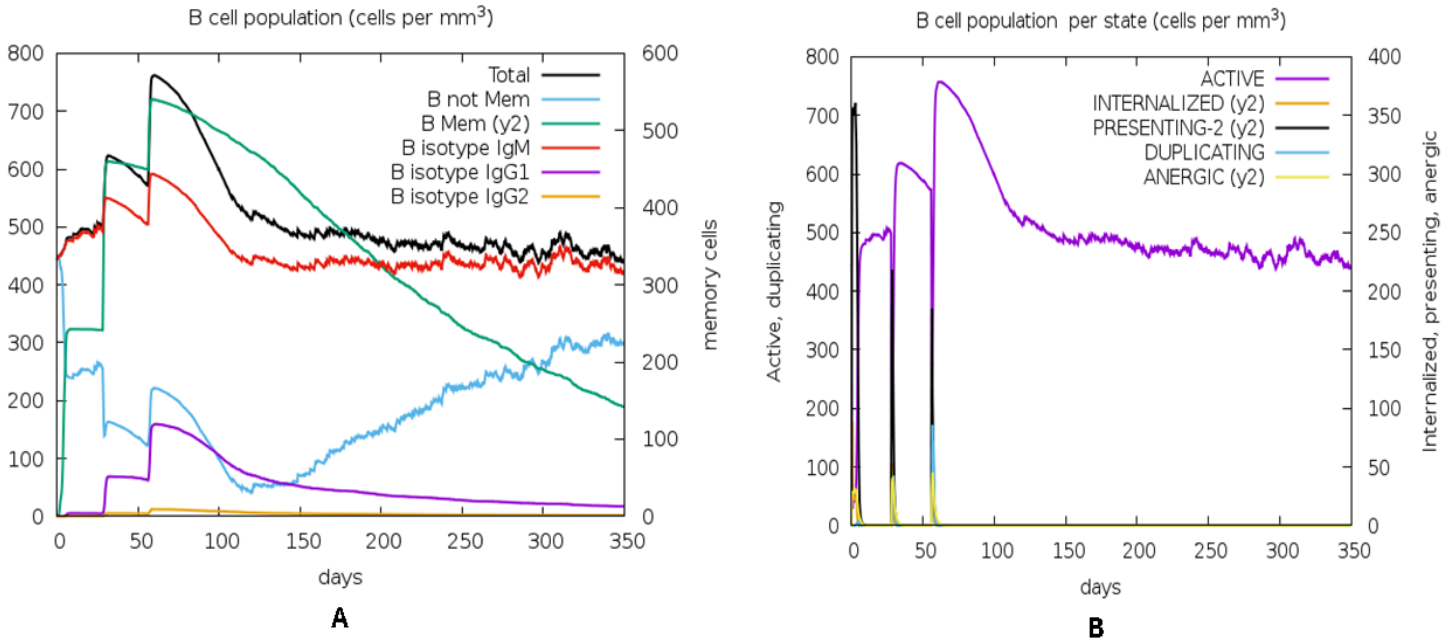

Figure 17: B-Cell Immune Response Dynamics.
(A) Total B-cell population and isotype distribution. (B) B-cell population per state, including active, duplicating, internalized, presenting, and anergic cells.

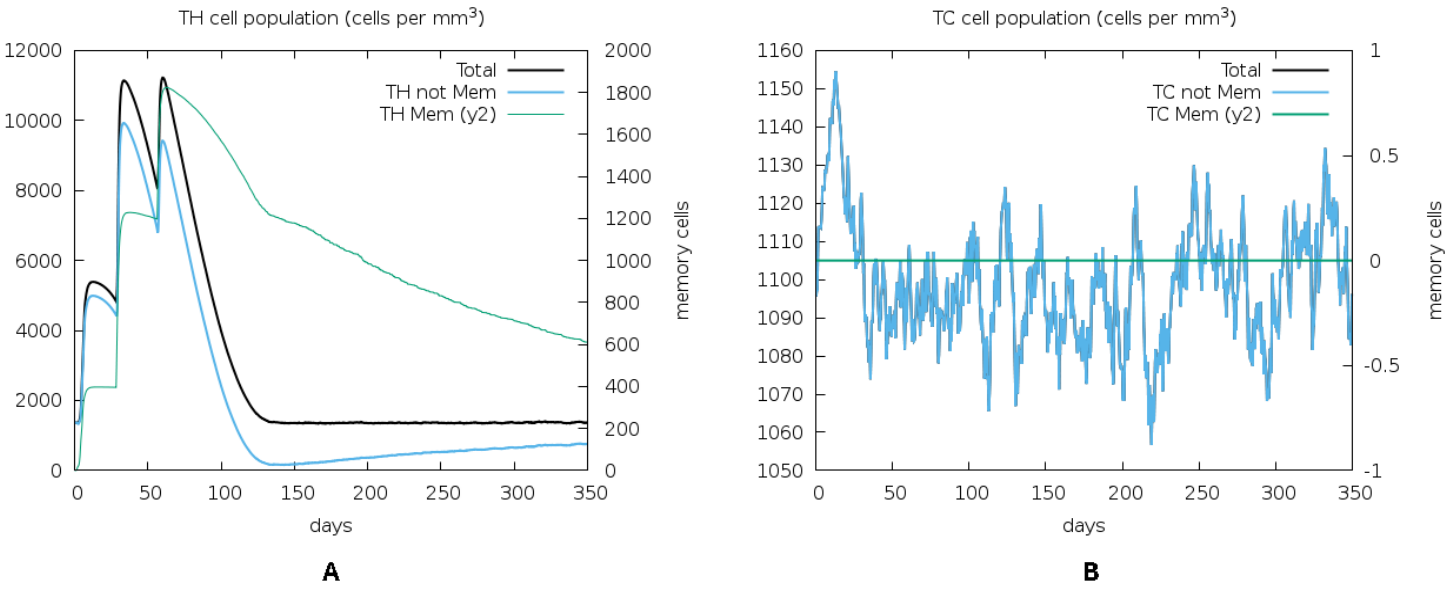

Figure 18: T-Cell Mediated Immune Response.
(A) Helper T-cell population. (B) Cytotoxic T-cell population.

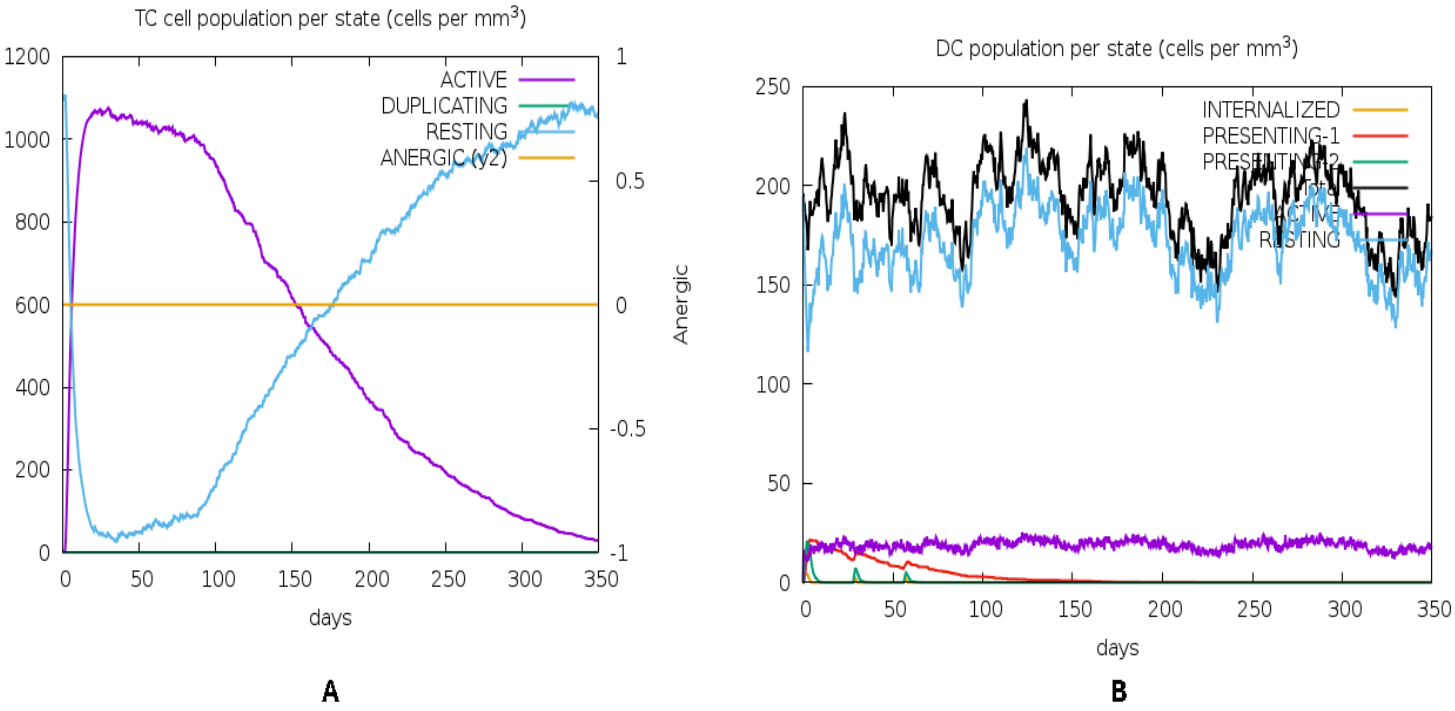

Figure 19: Cytotoxic T-Cell States and Innate Immune Activation.
(A) Cytotoxic T-cell population per state. (B) Dendritic cell population per state.

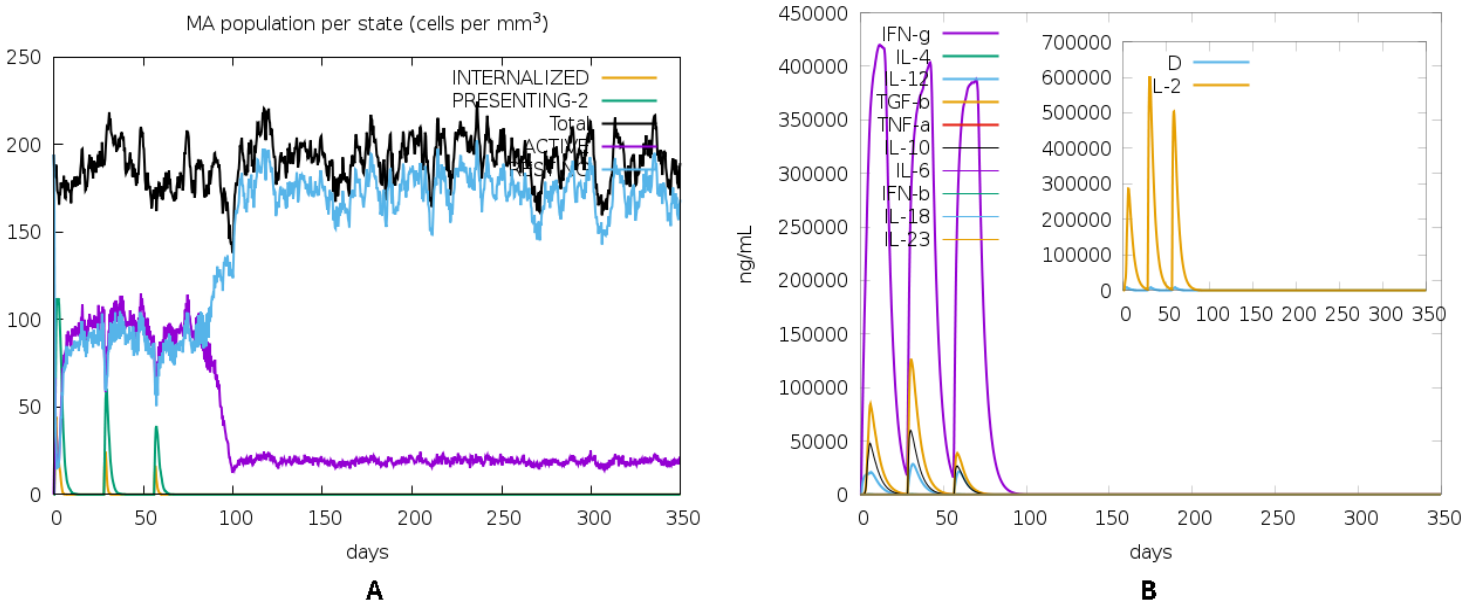

Figure 20: Macrophage Activation and Cytokine Response.
(A) Macrophages population per state. (B) Cytokine and interleukin concentration levels during immune simulation.

Discussion

The development of an effective cancer vaccine remains a cornerstone in the advancement of immunotherapy, particularly for complex malignancies like NPC. This research is driven by the limitations of conventional treatments, which often fail to address key mechanisms underlying NPC progression, including viral persistence and immune evasion [52]. Addressing these challenges is critical to improving patient outcomes.

In this context, immunotherapy has emerged as a promising avenue, leveraging the body's natural defenses to target cancer cells more precisely. Within immunotherapy, the application of immunoinformatics tools has transformed the design and development of vaccines. This computational approach enables the identification of potent epitopes capable of eliciting robust immune responses, streamlining vaccine development while reducing costs and time [53]. The use of in silico methods allows researchers to simulate and predict immunological interactions, offering a level of precision that was previously unattainable.

In this study, a multi-epitope vaccine was designed to target EBV-associated antigens (EBNA1, LMP1, and LMP2A) and survivin, a tumor-associated protein critical to NPC progression. The significance of these proteins in NPC has been well-documented [3, 4, 7, 54, 55], supporting their role as therapeutic targets.

Multi-epitope vaccines offer several advantages over conventional vaccine approaches, including the ability to stimulate both cellular and humoral immune responses, improved immunogenicity, and broader population coverage. This strategy has been effectively applied in vaccine development for viral infections and cancer, showing strong immunogenic potential and tumor protection [28, 45, 56].

The results highlighted the vaccine's strong immunogenic potential, with selected CTL, HTL, and LBL epitopes demonstrating high antigenicity, non-allergenicity, non-toxicity and comprehensive coverage in human populations. Conservation analysis confirmed that the epitopes were 100% conserved across protein isoforms and virus strains, ensuring broad applicability. Additionally, the inclusion of PADRE sequences ensured a broad CD4+ T-cell response across diverse HLA types, and the adjuvant sequence effectively enhanced immune activation. PADRE sequences have been shown to enhance vaccine efficacy by sustaining T-helper cell activation, improving immunogenicity, and increasing population coverage [38, 57].

The vaccine construct was enhanced with a fynomer sequence, a high-affinity scaffold that strengthens stability and functional efficacy. Previous studies, such as Sarvmeili et al. have demonstrated the effectiveness of fynomer-based vaccine designs, particularly in a SARS-CoV-2 multi-epitope vaccine, where it improved structural stability, antigen-receptor interactions, and solubility [28]. Their findings highlight fynomer's potential in optimizing vaccine architecture, supporting its incorporation into this NPC vaccine construct to enhance immunogenic performance.

The final vaccine construct designed in this study consists of 514 amino acids with an estimated molecular weight of 56 kDa. Using appropriate computational tools, its structure was evaluated for stability, solubility, thermostability, and other physicochemical properties, confirming its suitability for expression and immune system engagement [56].

Structural analysis plays a pivotal role in vaccine formulation, as secondary and tertiary structures influence epitope presentation and immune response induction. Secondary structure analysis revealed a well-balanced composition of alpha helices, beta strands, and random coils, ensuring proper antigen folding and accessibility. The predominance of random coil regions is particularly beneficial, as these flexible structures enhance epitope exposure for immune system recognition [56].

The 3D structure was modeled and refined using advanced computational tools, demonstrating high structural quality. Model validation through Ramachandran plot analysis and ProSA scoring (Figure 9) confirmed its native-like conformation, reinforcing its potential for proper folding and functionality [38, 39, 40]. Refinement and validation of the vaccine model are critical for ensuring structural integrity and functionality. The achieved scores across different validation parameters confirm that the construct is well-folded and possesses a native-like conformation, supporting its potential as a viable vaccine candidate.

Molecular docking analysis was performed to evaluate the interaction and binding affinity of the vaccine construct with MHC-I and MHC-II receptors, which play a crucial role in antigen presentation and immune activation. The docking results demonstrated strong and stable binding, supported by multiple salt bridges, hydrogen bonds, and non-bonded contacts, indicating a high degree of interaction stability (Figures 11, 12) [45]. Binding affinity analysis revealed negative ΔG values, confirming the energetic favorability of the interactions. The dissociation constants further validated the strong affinity between the vaccine and MHC molecules, reinforcing its potential to enhance antigen presentation and stimulate an immune response [45, 47].

Structural dynamics and stability of the vaccine-receptor complexes were assessed through MD simulations using the iMODS server (Figure 13, 14). The results demonstrated system balance, stability, and high flexibility, as indicated by low eigenvalues, covariance mapping, and elastic network models, which highlighted stable residue correlations and flexible interaction regions. These findings confirm that the vaccine construct maintains its structural integrity and adaptability, reinforcing its potential to optimize antigen presentation and enhance immune responses. A similar approach was employed by Kumar et al. [45], who utilized iMODS to assess structural stability and flexibility, further validating the effectiveness of this method.

In order to imitate the usual immune responses, an immune simulation was performed, demonstrating the vaccine's ability to activate both humoral and cellular immunity [51]. Following three vaccine doses, a significant increase in IgG and IgM titers was observed, indicating the formation of memory B-cells and the potential for long-term immunity. Additionally, elevated levels of CTL, HTL, and IFN-γ confirmed the vaccine's ability to induce a robust Th1-dominant immune response, which is critical for counteracting the immune-evasive mechanisms of NPC and promoting anti-tumor immunity. The immune simulation results aligned with expected immune responses, showing a progressive enhancement of immune activation with subsequent doses. While

three injections successfully generated a strong immunogenic response (Figures 16 to 20), additional doses could further strengthen and prolong the immune response, potentially leading to greater protection and sustained immunity [28, 45].

The successful expression of the vaccine construct requires efficient adaptation to a suitable host system for recombinant protein production [58]. In this study, the sequence was reverse transcribed and optimized for expression in E. coli strain K12 before being cloned into the pET-28a(+) vector, a widely used system for high-yield protein expression. Codon optimization using JCat resulted in a 1542 bp cDNA sequence with a Codon Adaptation Index (CAI) of 1.0, indicating optimal adaptation for E. coli translation. Additionally, the GC content of 49.09% fell within the ideal range for bacterial transcription and translation, further suggesting high expression efficiency.

Scientifically, this vaccine construct holds significant promise. By combining EBV-associated antigens with survivin, the design targets both viral and tumor-specific pathways, potentially reducing tumor progression and recurrence rates. This strategy aligns with the broader goal of developing personalized and precise cancer treatments.

However, despite these promising results, the study has limitations. The findings are primarily based on in silico analyses, which, while robust, require experimental validation. Laboratory studies are needed to test the vaccine's efficacy in vitro and in vivo. Additionally, potential long-term safety and immune response durability must be assessed through clinical trials. The reliance on computational tools also introduces potential biases based on the algorithms and databases used, which should be addressed in future research.

Looking ahead, the next steps involve initiating preclinical trials to thoroughly assess the vaccine's safety and efficacy. Further optimization of delivery systems and dosing schedules could enhance its performance. Additionally, evaluating the vaccine's effectiveness across diverse patient populations and various stages of NPC will be critical to ensure its broader applicability, ultimately providing a transformative treatment option for NPC patients worldwide.

Conclusion

This study presents a comprehensive approach for designing a novel multi-epitope vaccine targeting NPC, a cancer strongly associated with EBV infection. Utilizing advanced in silico methods, the research identified highly immunogenic and conserved epitopes from EBV latent proteins (EBNA1, LMP1, LMP2A) and

survivin, incorporating them into a rationally structured vaccine construct. The vaccine demonstrated strong antigenicity, non-toxicity, acceptable physicochemical properties, and stable structures, alongside strong binding interactions with immune receptors (MHC-I, and MHC-II). Immune simulations confirmed the vaccine's ability to activate robust cellular and humoral immune responses, ensuring both immediate and long-term immunity. Additionally, codon optimization ensured efficient expression in E.coli, making the vaccine viable for experimental production. This study underscores the potential of computational vaccine design as a cost-effective and efficient strategy for developing targeted immunotherapies, paving the way for a new era in personalized cancer immunotherapy.

References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

2. Global Cancer Observatory

3. Su, Z. Y., Siak, P. Y., Leong, C. O., & Cheah, S. C. (2023). The role of Epstein-Barr virus in nasopharyngeal carcinoma. *Frontiers in microbiology*, *14*, 1116143. https://doi.org/10.3389/fmicb.2023.1116143

4. Jin, P. Y., Zheng, Z. H., Lu, H. J., Yan, J., Zheng, G. H., Zheng, Y. L., Wu, D. M., & Lu, J. (2019). Roles of β-catenin, TCF-4, and survivin in nasopharyngeal carcinoma: correlation with clinicopathological features and prognostic significance. Cancer cell international, 19, 48. https://doi.org/10.1186/s12935-019-0764-7

5. Garg, H., Suri, P., Gupta, J. C., Talwar, G. P., & Dubey, S. (2016). Survivin: a unique target for tumor therapy. *Cancer cell international*, *16*, 49. https://doi.org/10.1186/s12935-016-0326-1

6. Silva, J. M., Alves, C. E. C., & Pontes, G. S. (2024). Epstein-Barr virus: the mastermind of immune chaos. Frontiers in immunology, 15, 1297994. https://doi.org/10.3389/fimmu.2024.1297994

7. Tsao, S. W., Tsang, C. M., & Lo, K. W. (2017). Epstein-Barr virus infection and nasopharyngeal carcinoma. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1732), 20160270. https://doi.org/10.1098/rstb.2016.0270

8. 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

9. Li, X., Fasano, R., Wang, E., Yao, K. T., & Marincola, F. M. (2009). HLA associations with nasopharyngeal carcinoma. Current molecular medicine, 9(6), 751–765. https://doi.org/10.2174/156652409788970698

10. Hildesheim, A., Apple, R. J., Chen, C. J., Wang, S. S., Cheng, Y. J., Klitz, W., Mack, S. J., Chen, I. H., Hsu, M. M., Yang, C. S., Brinton, L. A., Levine, P. H., & Erlich, H. A. (2002). Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. Journal of the National Cancer Institute, 94(23), 1780–1789. https://doi.org/10.1093/jnci/94.23.1780

11. Reardon, B., Koşaloğlu-Yalçın, Z., Paul, S., Peters, B., & Sette, A. (2021). Allele-Specific Thresholds of Eluted Ligands for T-Cell Epitope Prediction. *Molecular & cellular proteomics : MCP*, *20*, 100122. https://doi.org/10.1016/j.mcpro.2021.100122

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

13. 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), 2278. https://doi.org/10.1007/s00894-014-2278-5

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

15. Calis, J. J., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology, 9(10), e1003266. https://doi.org/10.1371/journal.pcbi.1003266

16. 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. https://doi.org/10.1186/1471-2105-8-361

17. Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., Kloetzel, P. M., Rammensee, H. G., Schild, H., & Holzhütter, H. G. (2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cellular and molecular life sciences : CMLS, 62(9), 1025–1037. https://doi.org/10.1007/s00018-005-4528-2

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

19. Karosiene, E., Lundegaard, C., Lund, O., & Nielsen, M. (2012). NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*, *64*(3), 177–186. https://doi.org/10.1007/s00251-011-0579-8

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

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

22. Dhall, A., Patiyal, S., Sharma, N., Usmani, S. S., & Raghava, G. P. S. (2023). A Web-Based Method for the Identification of IL6-Based Immunotoxicity in Vaccine Candidates. *Methods in molecular biology (Clifton, N.J.)*, *2673*, 317–327. https://doi.org/10.1007/978-1-0716-3239-0_22

23. Lathwal, A., Kumar, R., Kaur, D.P., & Raghava, G.P. (2021). In silico model for predicting IL-2 inducing peptides in human. *bioRxiv*. https://doi.org/10.1101/2021.06.20.449146

24. 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

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

26. Saha, S., & Raghava, G. P. (2007). Prediction methods for B-cell epitopes. Methods in molecular biology (Clifton, N.J.), 409, 387–394. https://doi.org/10.1007/978-1-60327-118-9_29

27. 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. https://doi.org/10.1186/1471-2105-7-153

28. Sarvmeili, J., Baghban Kohnehrouz, B., Gholizadeh, A., Shanehbandi, D., & Ofoghi, H. (2024). Immunoinformatics design of a structural proteins driven multi-epitope candidate vaccine against different SARS-CoV-2 variants based on fynomer. Scientific reports, 14(1), 10297. https://doi.org/10.1038/s41598-024-61025-2

29. Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. Nucleic acids research, 33(Web Server issue), W72–W76. https://doi.org/10.1093/nar/gki396

30. Nguyen, M. N., Krutz, N. L., Limviphuvadh, V., Lopata, A. L., Gerberick, G. F., & Maurer-Stroh, S. (2022). AllerCatPro 2.0: a web server for predicting protein allergenicity potential. Nucleic acids research, 50(W1), W36–W43. https://doi.org/10.1093/nar/gkac446

31. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics (Oxford, England)*, *33*(19), 3098–3100. https://doi.org/10.1093/bioinformatics/btx345

32. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy

server. Methods in molecular biology (Clifton, N.J.), 112, 531–552. https://doi.org/10.1385/1-59259-584-7:531

33. McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics (Oxford, England), 16(4), 404–405. https://doi.org/10.1093/bioinformatics/16.4.404

34. Geourjon, C., & Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer applications in the biosciences : CABIOS, 11(6), 681–684. https://doi.org/10.1093/bioinformatics/11.6.681

35. Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic acids research, 32(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gkh468

36. Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research, 40(Web Server issue), W294–W297. https://doi.org/10.1093/nar/gks493

37. Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein science : a publication of the Protein Society, 2(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

38. 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

39. Sippl M. J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins, 17(4), 355–362. https://doi.org/10.1002/prot.340170404

40. 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. https://doi.org/10.1107/S0021889892009944

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

42. Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein science : a publication of the Protein Society, 32(11), e4792. https://doi.org/10.1002/pro.4792

43. Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics, 9, 514. https://doi.org/10.1186/1471-2105-9-514

44. Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC bioinformatics, 14, 346. https://doi.org/10.1186/1471-2105-14-346

45. Kumar, A., Misra, G., Mohandas, S., & Yadav, P. D. (2024). Multi-epitope vaccine design using in silico analysis of glycoprotein and nucleocapsid of NIPAH virus. *PloS one*, *19*(5), e0300507. https://doi.org/10.1371/journal.pone.0300507

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

47. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. Bioinformatics (Oxford, England), 32(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

48. López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, P. (2014). iMODS: internal coordinates normal mode analysis server. Nucleic acids research, 42(Web Server issue), W271–W276. https://doi.org/10.1093/nar/gku339

49. Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research, 33(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gki376

50. SnapGene software (www.snapgene.com)

51. Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PloS one, 5(4), e9862. https://doi.org/10.1371/journal.pone.0009862

52. Cai, M., Wang, Y., Ma, H., Yang, L., & Xu, Z. (2024). Advances and challenges in immunotherapy for locally advanced nasopharyngeal carcinoma. *Cancer treatment reviews*, *131*, 102840. https://doi.org/10.1016/j.ctrv.2024.102840

53. Oli, A. N., Obialor, W. O., Ifeanyichukwu, M. O., Odimegwu, D. C., Okoyeh, J. N., Emechebe, G. O., Adejumo, S. A., & Ibeanu, G. C. (2020). Immunoinformatics and Vaccine Development: An Overview. *ImmunoTargets and therapy*, *9*, 13–30. https://doi.org/10.2147/ITT.S241064

54. Straathof, K. C., Leen, A. M., Buza, E. L., Taylor, G., Huls, M. H., Heslop, H. E., Rooney, C. M., & Bollard, C. M. (2005). Characterization of latent membrane protein 2 specificity

in CTL lines from patients with EBV-positive nasopharyngeal carcinoma and lymphoma. *Journal of immunology (Baltimore, Md. : 1950)*, *175*(6), 4137–4147. https://doi.org/10.4049/jimmunol.175.6.4137

55. Münz, C., Bickham, K. L., Subklewe, M., Tsang, M. L., Chahroudi, A., Kurilla, M. G., Zhang, D., O'Donnell, M., & Steinman, R. M. (2000). Human CD4(+) T lymphocytes consistently respond to the latent Epstein-Barr virus nuclear antigen EBNA1. *The Journal of experimental medicine*, *191*(10), 1649–1660. https://doi.org/10.1084/jem.191.10.1649

56. Larijani, A., Kia-Karimi, A., & Roostaei, D. (2023). Design of a multi-epitopic vaccine against Epstein-Barr virus via computer-based methods. Frontiers in immunology, 14, 1115345. https://doi.org/10.3389/fimmu.2023.1115345

57. Ma, S., Zhu, F., Zhang, P., Xu, Y., Zhou, Z., Yang, H., Tan, C., Chen, J., & Pan, P. (2025). Development of a novel multi-epitope subunit mRNA vaccine candidate to combat Acinetobacter baumannii. *Scientific reports*, *15*(1), 1410. https://doi.org/10.1038/s41598-024-84823-0

58. Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology*, *5*, 172. https://doi.org/10.3389/fmicb.2014.00172