

Immunoinformatic and Structural Design of an Influenza A DNA Vaccine Based on Conserved Nucleoprotein Epitopes and Heat Shock Protein 70 Adjuvant

Sanaz Sadeh¹, PhD; Amir Ghaemi², PhD;
Nazila Arbab Soleimani¹, PhD; Majid Moghbeli¹, PhD; Golnaz BahramAli³, PhD

¹Department of Microbiology, Damghan Branch, Islamic Azad University, Damghan, Iran;

²Department of Influenza and other Respiratory Viruses, Pasteur Institute of Iran, Tehran, Iran;

³Hepatitis and AIDS Department, Pasteur Institute of Iran, Tehran, Iran

Correspondence:

Amir Ghaemi, PhD;

Department of Influenza and Respiratory Viruses, Pasteur Institute of Iran, No. 69, 12th Farwardin Ave., Postal code: 13169-43551, Tehran, Iran

Tel: +98 21 64112187

Email: ghaem_amir@yahoo.com
a_ghaemi@pasteur.ac.ir

Received: 17 November 2025

Revised: 09 February 2026

Accepted: 28 March 2026

What's Known

- Influenza A mutates rapidly, limiting the breadth and durability of current vaccines.
- Multi-epitope and DNA-based platforms can enhance cross-strain immunity.
- Nucleoprotein (NP) is highly conserved and immunogenic, making it an attractive target for universal vaccine design.

What's New

- We designed a structurally validated DNA vaccine integrating conserved NP epitopes with cell-penetrating peptide (CPP)-enhanced delivery, a tetanus toxoid (TT) helper epitope, and a tandem heat shock protein 70 (Hsp70) adjuvant.
- The construct uniquely combines epitope prediction, antigen-processing analysis, molecular dynamics (MD)-refined modeling, and toll-like receptor (TLR) docking.
- This study proposes a stable, strongly immunogenic NP-based vaccine candidate ready for experimental evaluation.

Abstract

Background: Influenza A is a health concern, as it constantly mutates and changes antigens. The conventional vaccine is not very effective in providing immunity. We have designed and characterized a DNA multi-epitope vaccine based on conserved regions of the influenza A nucleoprotein (NP) and have assessed the immune response.

Methods: This is a bioinformatics study carried out in 2025 at the Pasteur Institute of Iran in Tehran. Epitopes were predicted using the immune epitope database and were screened for conservancy, antigenicity, toxicity, and allergenicity using the alignment-independent antigenicity prediction server (VaxiJen), toxicity prediction tool (ToxinPred), and allergenicity prediction tool (AllergenFP).

The vaccine contains a cell-penetrating peptide (CPP) derived from the hemagglutinin protein, a vascular endothelial growth factor 183 (VEGF-183) signal peptide, a tetanus toxoid (TT) helper peptide (830-844), NP peptides, and two repeats of heat shock protein 70 (Hsp70)(407-426). The peptides are arranged using flexible and rigid linkers. The structure was predicted using the self-optimized prediction method with alignment (SOPMA) and AlphaFold2. The structure was further analyzed using high-ambiguity-driven protein-protein docking (HADDOCK) with human leukocyte antigen (HLA) class I and II and toll-like receptor (TLR) 2 and 4, followed by a 200-ns molecular dynamics (MD) simulation.

Results: The docking study showed proper interaction of the selected peptides with HLA and murine BALB/c (H-2d) major histocompatibility complex (MHC) class I and II and the whole vaccine with innate immune system receptors. *In silico* immune simulation indicated cytokine release and T cell activation patterns.

Conclusion: The constructed multi-epitope DNA construct was predicted to be antigenic and non-toxic, with structural stability and immunogenic potential. These computational results support prioritizing this construct for experimental expression and murine preclinical validation (immunogenicity, safety, and protection).

Please cite this article as: Sadeh S, Ghaemi A, Arbab Soleimani N, Moghbeli M, BahramAli G. Immunoinformatic and Structural Design of an Influenza A DNA Vaccine Based on Conserved Nucleoprotein Epitopes and Heat Shock Protein 70 Adjuvant. Iran J Med Sci. doi: 10.30476/ijms.2026.109702.4506.

Keywords • Immunoinformatics • Influenza virus strains • Molecular dynamics • Cell-penetrating peptide

Introduction

Influenza A remains a major public health challenge because of its antigenic variability and limited cross-protection provided by strain-matched vaccines.¹

Deoxyribonucleic acid (DNA) vaccines can induce antigen expression in host cells, but their immunogenicity is often limited by inefficient intracellular delivery.² Accordingly, while delivery remains a key limitation, our DNA-encoded multi-epitope construct was designed to enhance antigen expression and availability through codon optimization, a cytomegalovirus (CMV) promoter-based expression cassette, vascular endothelial growth factor 183 (VEGF-183)-mediated endoplasmic reticulum (ER) and secretory routing, and heat shock protein 70 (Hsp70)-based immunostimulatory elements. Future validation may pair these optimizations with electroporation to improve DNA uptake and expression.³

Among conserved influenza A antigens, Nucleoprotein (NP) is an attractive vaccine target because of its conservation and role in cellular immune responses.¹ Leveraging *in silico* approaches offers promising solutions to overcome challenges related to vaccine immunogenicity by identifying and predicting potential epitopes within pathogens.⁴

Recent immunoinformatics studies have proposed influenza multi-epitope constructs with different targets and validation levels. Bazhan and others supported the feasibility of DNA-encoded multi-epitope strategies by reporting an influenza epitope-based T cell antigen with immunogenicity and partial protection in BALB/c mice, while Yuan and others proposed a broad-spectrum hemagglutinin (HA), neuraminidase (NA)-based multi-epitope construct evaluated through *in silico* structural and immune analyses.^{5, 6} In this context, we propose an NP-centered DNA multi-epitope construct and evaluate it using a structure-informed analysis, including molecular dynamics simulation (MD) refinement of the modeled construct followed by docking analyses against adaptive and innate targets, including human leukocyte antigen (HLA) class I/II structures and toll-like receptor 2 (TLR2)/TLR4. With this in mind, our construct was designed around conserved NP epitopes due to their essential role in viral replication and immunogenicity.⁷

An HA-derived cell-penetrating peptide (CPP) was incorporated to enhance intracellular delivery of the vaccine construct through direct or endocytic uptake pathways.⁸

Signal peptides facilitate cotranslational

targeting to the ER and subsequent secretion, thereby supporting antigen processing and presentation. In this study, we employed the signal peptide from VEGF-183 to direct the multi-epitope construct to the ER and support efficient expression and secretion.⁹

To enhance immunogenicity in the multi-epitope construct, adjuvants are commonly employed to increase antigen uptake, processing, and lymphocyte activation.¹⁰ The microbial Hsp70(407–426) motif has been reported to stimulate cytokine production and promote dendritic cell maturation, thereby supporting antigen-presenting cell activation and T cell-oriented immune induction, an immune direction often prioritized in NP-based influenza vaccine strategies.¹¹ Accordingly, we incorporated Hsp70 407–426 as a C-terminal adjuvant component in our construct, consistent with prior use of this peptide epitope as an adjuvant element in vaccine contexts.¹² This study aims to design a DNA-encoded multi-epitope construct against influenza and to evaluate its *in silico* immunogenic and physicochemical properties, ultimately informing the development of a vaccine candidate to reduce influenza infection.

Materials and Methods

This bioinformatics and *in silico* study was conducted in 2025 at the Pasteur Institute of Iran, Tehran, Iran.

Protein Sequence Retrieval

The sequence of the nucleoprotein (NP) of Influenza A virus (A/Puerto Rico/8/1934, H1N1) with the GenBank Reference Sequence NP_040982.1 was retrieved for construct design.

Epitope Selection

Segments NP296-304, NP200-208, NP 418-426, NP 381-395 and NP 364-378 were selected for import into our peptide. The listed epitopes were predicted and analyzed using HLA class I/II and murine BALB/c (H-2d) major histocompatibility complex (MHC) class I/II and BepiPred-2.0 tools on the Immune Epitope Database (IEDB) web server.

Prediction of T cell Epitopes for human and murine BALB/c (H-2d)

To predict influenza A NP-derived T cell epitopes for both HLA and BALB/c (H-2d) MHC alleles, we used the IEDB Analysis Resource binding prediction tools. MHC class I binding was predicted using NetMHCpan 4.1 EL (<http://tools.iedb.org/mhci/>), and MHC class II binding was predicted using NetMHCIIpan 4.1 EL (<http://tools.iedb.org/mhci/>).

tools.iedb.org/mhcii/). Predictions for MHC-I and MHC-II epitopes were performed separately for HLA and BALB/c (H-2d) allele panels. The BALB/c (H-2d) alleles were added to allow *in vivo* validation in mice, and HLA alleles were chosen according to their frequency in the Iranian population. The epitope length for MHC-I in both BALB/c mice and humans was primarily 9-mers, and for MHC-II, the epitope length was 15-mers.

H2-Ld, H2-Kd, and H2-Dd belong to MHC class I, while H2-IEd and H2-IAd belong to BALB/c MHC class II. The HLA alleles HLA-B 0702, HLA-B 5101, HLA-B 0801, HLA-B 5301, HLA-A 2601, HLA-A 3101, HLA-A 3301, HLA-A 1101, HLA-A 68:01 and HLA-B 3501 for MHC class I and HLA-DRB*11, HLA-DRB1*08, HLA-DRB*04 for MHC class II were common in the Iranian population based on the accessible report (<http://www.allelefrequencies.net>).¹³

Sorting of the conserved peptides with predicted binding affinity to HLA class I/II and murine BALB/c (H-2d) MHC class I/II molecules was performed using their percentile rank, selecting peptides with a rank ≤ 1 .

B Cell Epitope Prediction

Epitope prediction for linear B cells was performed using the IEDB B cell epitope prediction tool. In the linear method, prediction was performed using BepiPred-2.0, Chou and Fasman beta-turn, amino-surface accessibility, Karplus and Schulz flexibility, Kolaskar and Tongaonkar antigenicity, and Parker hydrophilicity.¹⁴

Antigenicity Evaluation

Antigenicity was assessed at two levels using alignment-free predictors. First, individual shortlisted epitopes (MHC-I 9-mers, MHC-II 15-mers, and linear B cell epitopes) were screened using VaxiJen v2.0 (<http://scratch.proteomics.ics.uci.edu>). A threshold of 0.4 is commonly used, where sequences scoring >0.4 are classified as antigenic; therefore, epitopes meeting this cutoff were retained for construct assembly. Second, the final multi-epitope construct was evaluated for overall antigenicity using both VaxiJen v2.0 (cutoff=0.4) and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu>).¹⁵

Allergenicity and Toxicity Prediction

Allergenicity of the selected epitopes was assessed using AllgPred (<https://webs.iitd.edu.in/raghava/allgpred2/batch.html>). Allergenicity of the final multi-epitope construct was evaluated using AllergenFP v1.0 ([\[webs.iitd.edu.in/raghava/toxinpred/index.htm\]\(https://webs.iitd.edu.in/raghava/toxinpred/index.htm\)\).¹⁶ Moreover, the toxic effects of peptides were predicted using ToxinPred \(<https://webs.iitd.edu.in/raghava/toxinpred/index.htm>\).¹⁶](https://ddg-pharmfac.</p>
</div>
<div data-bbox=)

Epitope Conservancy and Global Population Coverage Analysis

The epitope conservancy analysis was performed using the reference-based identity/sequence mapping method using the IEDB Epitope Conservancy tool (<http://tools.iedb.org/conservancy/>). The analysis was conducted for the entire set of epitope classes incorporated into the final construct, which are MHC-I/CTL, MHC-II/HTL, and the B cell epitope, by inputting the reference influenza A NP sequence (GenBank accession number NP_040982.1) and the final construct sequence into the tool. Population coverage was analyzed using the IEDB Population Coverage tool (<http://tools.iedb.org/population/>) for HLA-restricted T cell epitopes (MHC-I and MHC-II) in Iranian and global populations; population coverage was not assessed for B cell epitopes because they are not HLA-restricted.

Identification of Potential CPPs from Influenza HA

For the influenza A virus hemagglutinin (HA) protein, specifically the H1N1 subtype, bioinformatics approaches were undertaken to identify CPPs and non-CPPs.¹⁷

Accordingly, we analyzed the reference genome of the influenza A strain A/Puerto Rico/8/1934 (H1N1; NCBI Taxonomy ID 211044). To identify new CPPs, we first retrieved the sequence in FASTA format from the NCBI server (GenBank accession NP_040980.1). Then, the CellPPD web server (<https://webs.iitd.edu.in/raghava/cellppd/index.html>) was used to identify cell-penetrating peptides. The submitted influenza hemagglutinin sequence was analyzed to identify potential CPPs and assigned an SVM score indicating its probability of cell penetration. Prediction was performed using the default SVM threshold value of -0.1 implemented in the CellPPD server. The physicochemical properties of the identified CPP peptides, including molecular weight, isoelectric point (pI), hydrophobicity, net charge, and amphipathicity, were also analyzed using the CellPPD web server.⁸

The C2Pred web server (<http://lin-group.cn/server/C2Pred>) was used to further evaluate the cell penetration probability of each peptide. The CPPs identified by the CellPPD server are sent to the C2Pred server in FASTA format.

Then, to check the absorption efficiency of known CPPs in the previous step, the MLCPP web server (www.thegleelab.org/MLCPP/) was used. In this way, all the CPPs that were

identified in the previous step using the CellPPD server were sent to the MLCPP web server.⁸ For cellular localization/orientation assessment, the TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>) was used to predict transmembrane topology and inside/outside orientation probabilities of the candidate peptides.⁸

Physicochemical Properties, Toxicity, Sensitivity, and Immunogenicity Evaluation of CCP Peptides

The ProtParam server (<https://web.expasy.org/protparam/>) was used to evaluate the grand average of hydropathy (GRAVY), instability index, and half-life of CPPs. The AllergenFP web server (<http://ddg-pharmfac.net/AllergenFP/>) was used to check the sensitivity of CPPs. To check the toxicity of CPPs, the ToxinPred web server (<https://webs.iitd.edu.in/raghava/toxinpred/algo.php>) was used.¹⁷

IEDB Immunogenicity Predictor server (<http://tools.iedb.org/immunogenicity/>) was used to check the immunogenicity of CPPs.¹⁷

Designing a Multi-Epitope Construct

An N-terminal influenza hemagglutinin-derived CPP was incorporated as an intracellular delivery tag to enhance cellular association/uptake of the expressed multi-epitope construct. However, its actual contribution to cellular internalization, intracellular trafficking, endosomal escape, and biodistribution has not been experimentally evaluated and requires further validation.⁸

Vascular endothelial growth factor (VEGF) was incorporated as a signaling peptide. In addition, the tetanus toxoid (TT) helper epitope (830-844) was incorporated as a helper epitope to enhance the immunogenicity of the vaccine by stimulating CD4⁺ lymphocytes.¹⁸ Both components were placed at the N-terminus of our multi-epitope construct. Subsequently, to stimulate cellular immunity, T cell MHC-I and T cell MHC-II epitopes were placed at NP296-304, NP200-208, NP418-426, and NP381-395, respectively. Then, to help stimulate humoral immunity, the B cell linear epitope NP 364-378 and the NP 296-304 epitope, which was shared between the B cell linear epitope and human MHC class I, were implemented. GGGs linkers were used as flexible spacers between T- and B cell epitopes to reduce structural interference, while acknowledging that junctional peptides may potentially form at epitope-linker boundaries.¹⁹

To enhance the immunostimulatory capacity of the construct in the context of a DNA multi-epitope construct design, two tandem repeats of the microbial Hsp70(407–426) peptide were incorporated as a C-terminal adjuvant element.

Accordingly, the Hsp70(407–426) repeats were positioned at the C-terminus and connected to the upstream epitope string via an EAAAK linker to provide spacing and reduce potential structural interference with the selected NP epitopes.¹²

The Physicochemical Parameters of Multi-Epitope Construct

The physicochemical properties of the final construct, including molecular weight, theoretical isoelectric point (pI), aliphatic index, total average hydropathic (GRAVY), extinction coefficient, instability index, and estimated half-life, were calculated using ProtParam Server (<http://web.expasy.org/protparam>).

Prediction of Secondary Structure

The analysis was performed using the SOPMA servers and PSIPRED v4.0118. For SOPMA, the output width was set to 70. The parameters for the number of conformational states, similarity threshold, and window width were defined as 4 (helix, sheet, turn, and coil), 8, and 17, respectively.²⁰

Tertiary Structure Prediction and Modeling Using AlphaFold

The tertiary structure was predicted and modeled utilizing the AlphaFold server (<https://alphafold.ebi.ac.uk/>) and was presented in Protein Data Bank (PDB) format.²¹

Molecular Dynamics (MD) Simulation, Refinement, and in Silico Structural Assessment

MD simulation was used to refine and stabilize the predicted structure. The initial 3D model of the full-length multi-epitope construct (AlphaFold output) was used as the starting structure for MD simulation, and MD was performed before refinement to obtain a post-stabilization construct conformation. For this purpose, MD simulation was conducted using the GROMACS 2023.4 software suite and the CHARMM36 force field, in conjunction with the TIP3P water model. System neutralization and the achievement of a physiological ionic strength of 0.14 M were accomplished by adding appropriate quantities of sodium (Na⁺) and chloride (Cl⁻) ions.

Initially, energy minimization was performed to relax internal constraints, utilizing the steepest descent integrator with the Verlet cutoff scheme until system convergence was achieved, defined as the maximum force falling below 1000.0 kJ/mol/nm. The equilibration process was then carried out within the canonical (NVT) and isobaric (NPT) ensembles under positional restraints for 500 and 1000 ps, respectively.

The V-rescale thermostat, a modified Berendsen thermostat, was employed to maintain a constant volume and temperature of 310 K, while a C-rescale pressure coupling algorithm was used to maintain a constant pressure of 1 bar.

Subsequently, an MD production run was executed for 200 ns with a 2 fs time step under conditions of 310 K and 1 atm. During the production runs, temperature and pressure coupling were managed using the Nosé-Hoover thermostat and the Parrinello-Rahman barostat algorithms, respectively. The equations of motion were integrated via the leapfrog algorithm, with all bonds constrained using the LINCS algorithm. The generated trajectories were analyzed employing built-in GROMACS utilities. Structural stability was primarily monitored by root-mean-square deviation (RMSD); root-mean-square fluctuation (RMSF) was computed over the post-stabilization interval used in the results.

Following MD stabilization (RMSD convergence), the final frame from the stabilized portion of the trajectory was extracted and used as the input for the GalaxyRefine server.²²

Finally, *in silico* structural quality assessment was conducted utilizing the ProSA web server and the SAVES v6.0 server (<https://saves.mbi.ucla.edu>) to compare the initial AlphaFold model with the post-MD refined model.²³ The Ramachandran plot and ERRAT chart were created by the SAVES v6.0 PROCHECK and ERRAT tools.

Prediction of Cleavage Sites and TAP Transporter Binding Analysis

To predict proteasome cleavage sites and evaluate transporter-associated with antigen processing (TAP) transporter binding affinity within the multi-epitope construct, we utilized the IEDB (Immune Epitope Database) tool called the Proteasomal Cleavage/TAP Transport/MHC Class I Combined Predictor, accessed via the provided link (<http://tools.iedb.org/processing/>).

This analysis was intentionally restricted to the final MHC class I-restricted (CTL) epitopes included in the construct, because proteasomal cleavage and TAP-dependent transport are defining steps of the endogenous MHC-I antigen-processing pathway (cytosolic proteasomal processing followed by TAP-mediated peptide transport into the ER for MHC-I loading). In contrast, MHC class II (HTL) epitopes are predominantly generated via endosomal/lysosomal proteolysis and are not TAP-dependent. Therefore, TAP/proteasome-based processing predictors are not directly applicable to MHC-II epitope processing.

Molecular Docking and Analysis of Predicted Binding Affinity

The 3D structures used included TLR2 (PDB ID: 2Z7X, chain A), TLR4 (PDB ID: 4G8A, chain A), HLA class I, and HLA class II, which were obtained from the Protein Data Bank (PDB). Peptide-MHC docking was performed for the final shortlisted CTL and HTL epitopes with HLA class I (MHC-I) and HLA class II (MHC-II), respectively, while protein-protein docking was carried out only for the full-length construct with TLR2/TLR4. Full-length construct docking with HLA was not performed because HLA complexes present processed peptides rather than intact proteins. Thus, the HLA-related docking analysis was restricted only to the epitope level.²⁴ TLR2 and TLR4 were selected due to their roles in innate immunity and Hsp70-mediated immunostimulatory activity.¹¹

All ligands and water molecules linked to these structures were subsequently removed.²⁵ The results were then submitted to HADDOCK 2.4 for the docking process. The optimal cluster was determined by analyzing the lowest HADDOCK score and was further refined through the HADDOCK correction interface. The PRODIGY web server was employed to estimate the binding affinity of each docked complex by choosing the most refined structure from the optimal cluster. The binding affinity of the docked complexes was estimated using this server at a temperature of 37 °C, with the results presented as K_d (dissociation constant) values. Additionally, the ΔG (Gibbs free energy) values in kcal mol⁻¹ for the complexes were determined as well. Interactions between the multi-epitope construct and target receptors were visualized through PyMOL, with mapping conducted using Protein Data Bank Summary (PDBsum).²⁶

Codon Optimization and in Silico Simulation of Vaccine Construct

The Java Codon Adaptation Tool (JCat) facilitated the processes of codon optimization and reverse translation.²⁷ Because the construct is ultimately aimed for human use, the coding sequence was codon-optimized for *Homo sapiens*; BALB/c (H-2d) allele screening was performed to enable murine preclinical evaluation and to capture shared murine-human epitope coverage during epitope selection. Following codon optimization, a cDNA sequence of the multi-epitope construct was generated for high-level expression in mammalian systems. Guanine-cytosine (GC) content and codon adaptation index (CAI) were used to evaluate predicted protein expression. Furthermore, restriction enzyme sites *BamHI* and *EcoRI* were

incorporated at the N- and C-terminal ends of the optimized codon sequence, respectively, for *in silico* cloning in the pcDNA3.1(+) vector using SnapGene 6.2 software. The CMV-driven pcDNA3.1(+) backbone is suitable for mammalian expression and is routinely used for preclinical DNA vaccination studies in mice.²⁸

Simulation of the Immune System

Immune simulation was performed using C-IMMSIM to assess the predicted immunogenicity of the multi-epitope construct.²⁹ The most prevalent HLA alleles in the target population were selected for the study. The multi-epitope construct was administered three times at 1, 42, and 84 hours under *in silico* conditions, with each time step corresponding to 8 hours in *in vivo* conditions. Each injection contained 1000 molecules. The parameters for the random seed, simulation volume, and simulation step were established at 12345, 10 μ L, and 300, respectively.

Results

Multi-Epitope Construct: Design and *in Silico* Properties

The final multi-epitope construct targeted influenza A NP and included an HA-derived

CPP, VEGF-183 signal peptide, TT helper epitope, selected NP epitopes, and Hsp70 adjuvant repeats. The amino acid sequence and structural organization are shown in figure 1.

MHC I and II Peptide Binding Prediction

A total of seven epitopes in three alleles (H2-Kd, H2-Ld, H2-Dd) were identified in the analysis of murine MHC class I (H-2d) alleles in BALB/c. Among the epitopes, the LPFDRTTVM epitope was selected. The LPFDRTTVM epitope was notable because it was associated with multiple alleles and was shared between the murine MHC class I allele (H2-Ld) and the HLA class I alleles HLA-B07:02, HLA-B35:01, HLA-B51:01, HLA-B08:01, HLA-B53:01, and HLA-A26:01 (supplementary table 1).

Among the epitopes identified for HLA class I alleles, LPFDRTTVM, GINDRNFWR, and SLVGIDPFR epitopes were selected, each of which was associated with several HLA alleles (supplementary table 2). We investigated murine MHC class II (I-Ad/I-Ed) and HLA class II, focusing on longer peptides (15-mers). A total of eight epitopes were identified in two alleles (H2-IEd, H2-IAd) in MHC class II analysis of BALB/c. Among the predicted epitopes, we chose the RSRYWAI TRSGGNT epitope (supplementary table 3).

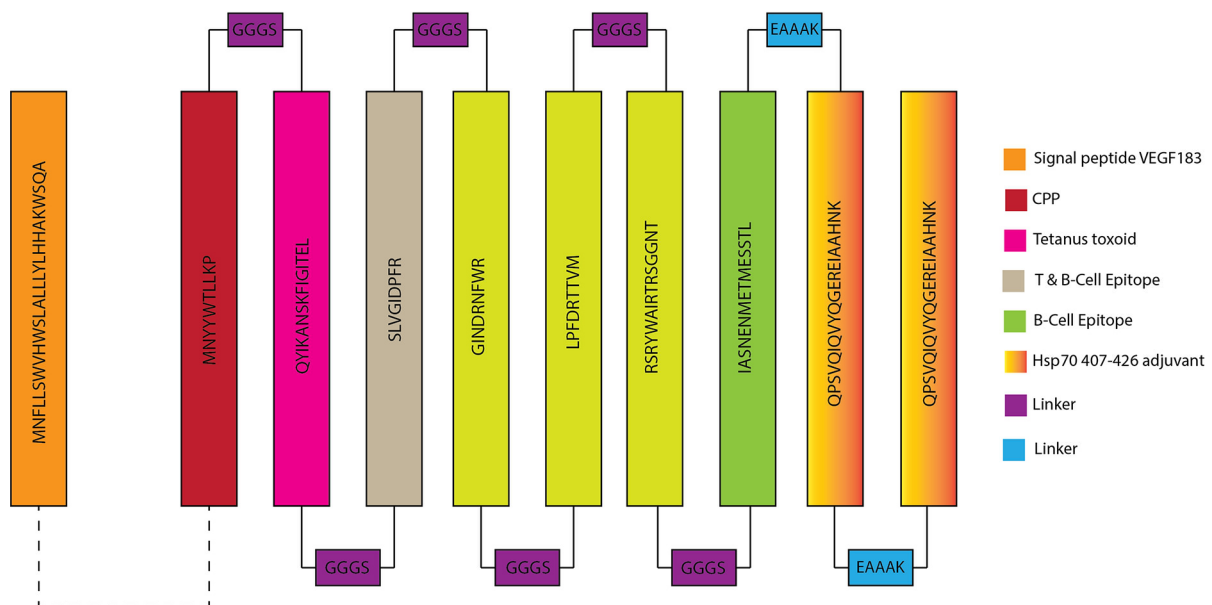


Figure 1: This figure shows the designed multi-epitope construct. The construct begins with a VEGF-183 signal peptide (orange) at the N-terminus to facilitate entry into the secretory pathway; this region is cleaved post-translationally (dashed line). The remaining secretory structure, after removal of the signal peptide, serves as the main scaffold for immunoinformatic analyses. Following this are a CPP (dark red), a tetanus toxoid (TT) helper epitope (830–844) (pink), and several immunodominant T cell and B cell epitopes (gray, yellow-green, and green), all connected via flexible GGGG linkers (purple). The C-terminal region contains two tandem repeats of the Hsp70(407–426) peptide (gradient orange-yellow), linked via a rigid EAAAK linker (blue). This final construct, excluding the cleaved signal peptide, serves as the core vaccine model for structural modeling and immunoinformatic evaluation. VEGF-183: Vascular endothelial growth factor isoform 183; CPP: Cell-penetrating peptide; TT: Tetanus toxoid; Hsp70: Heat shock protein 70; GGGG: Glycine–glycine–glycine–serine; EAAAK: Glutamate–alanine–alanine–alanine–lysine

Among the HLA class II epitopes, 11 epitopes were identified. Among them, the RSRYWAI TRSGGNT epitope was selected. The RSRYWAI TRSGGNT epitope was notable because it was associated with multiple alleles and was shared between the mouse allele (H2-IE_d) and the human alleles HLA-DRB1*08:02, HLA-DRB1*11:01, and HLA-DRB1*04:01 (supplementary table 4).

All the mentioned epitopes from MHC classes I and II in human and BALB/c mice had percentile scores ≤ 0.95 , indicating high affinity for MHC molecules; several epitopes below 0.1 were identified among them, indicating very high binding affinity. All selected epitopes were predicted to be antigenic, non-toxic, and non-allergenic.

B Cell Epitope Analysis

Fifteen linear B cell epitopes were initially predicted, of which six passed antigenicity, toxicity, and allergenicity screening. Finally, the epitope IASNENMETMESSTL (15-mers) was selected for the multi-epitope candidate vaccine, based on the fact that the epitope was non-toxic and non-allergenic (supplementary table 5). Comparison of MHC-I and linear B cell epitope predictions showed that the SLVGIDPFR epitope identified for MHC-I T cells shares overlap with the linear epitope YSLVGIDPFRLLQNSQVYSLI from residues 2 to 10. Therefore, we utilized this overlap and included the SLVGIDPFR epitope in the structure of our multi-epitope construct to ensure its capability to stimulate T cells through the MHC-I pathway while also acting as a linear epitope for B cells (supplementary tables 2 and 5).

Population Coverage and Conservancy Analysis

The selected HLA class I (CTL) and HLA class II (HTL) epitopes (with murine BALB/c H-2d overlap captured during the prediction step) and the retained linear B cell epitope showed 100% sequence identity in a reference-based assessment against the NP reference sequence (GenBank: NP_040982.1) and the final construct sequence, indicating correct epitope mapping to the design templates. The results are presented in supplementary table 6. Additionally, population coverage percentage (PPC%) in the Iranian and world populations was computed only for the HLA-restricted T cell epitopes (MHC-I and MHC-II) using the IEDB Population Coverage tool (<http://tools.iedb.org/population/>), according to which the highest coverage in the Iranian population is related to the HLA-A11:01 allele of human MHC class I (Iran PPC%=27.07%). The highest coverage in the world dataset is associated with the HLA-DRB115:01 allele of

human MHC class II (world PPC%=18.41%), as shown in tables 2 and 4.

Identification and Biophysical Characterization of Predicted CPPs

The CellPPD server identified 24 potential CPPs with an SVM threshold of -0.1, and their physicochemical properties, including hydrophobicity, charge, isoelectric point (pI), and molecular weight, were assessed. Among these, the peptide MNYYWTLLKP was selected based on its favorable properties: an SVM score of 0.07, pI of 8.83, molecular weight of 1328.74 Da, and net positive charge of +1.00 (supplementary table 7).

Further *in silico* screening with C2Pred and MLCPP supported its CPP classification. The C2Pred server is defined based on a threshold of 0.5, such that peptides below this threshold are identified as non-CPPs and peptides above this threshold are identified as CPPs. MNYYWTLLKP was consistently classified as a CPP by CellPPD, C2Pred, and MLCPP, supporting its selection. Uptake efficiency of candidate CPPs was assessed with the MLCPP server. Consistent with these predictions, MNYYWTLLKP showed moderate MLCPP-predicted uptake efficiency (0.0174) and a TMHMM total N-in probability of 0.4146. Of the 24 peptides, 22 were predicted as inside by TMHMM (supplementary table 8; supplementary figure 1).

CPP Immunogenicity, Allergenicity, Toxicity

The immunogenicity, toxicity, and sensitization of CPPs were evaluated, according to which the MNYYWTLLKP peptide was predicted to be non-toxic and non-allergenic. Low immunogenicity may reduce unwanted immune responses and improve CPP efficiency as delivery vehicles. The immunogenicity of the selected peptide was -0.22201. ProtParam analysis predicted a long half-life of over 10 hours in *E. coli* and 30 hours in mammalian cells, along with hydrophilic properties (negative GRAVY), suggesting stability in biological fluids (supplementary table 9).

Prediction of Vaccine Antigenicity and Allergenicity

The assessed antigenicity of the designed multi-epitope construct by VaxiJen v2.0 (threshold=0.4) was 0.54. ANTIGENpro score was 0.933. For allergenicity prediction, both AlgPred and AllergenFP v1.0 predicted the construct to be non-allergenic. Subsequent toxicity prediction was performed using ToxinPred, which predicted the multi-epitope construct to be non-toxic.

Table 1: Physicochemical properties, antigenicity, allergenicity, and toxicity assessment of the multi-epitope construct

Characteristics	Finding	Remark
Number of amino acids	156	Suitable
Molecular weight	16688.52	Average
Theoretical pI	9.47	Alkaline
Chemical formula	C723H1134N218O230S4	-
Extinction coefficient (at 280 nm in H ₂ O)	25440	-
Estimated half-life (mammalian reticulocytes, <i>in vitro</i>)	30 hours	-
Estimated half-life (yeast cell, <i>in vivo</i>)	>20 hours	-
Estimated half-life (<i>E. coli</i> , <i>in vivo</i>)	>10 hours	-
Aliphatic index of vaccine	61.99	Thermostable
Grand average of hydropathicity (GRAVY)	-0.600	Hydrophilic
Antigenicity vaxijen v2.0	0.5447	Antigen threshold for this model: 0.4
ANTIGENpro	0.933298	Antigen
Allergenicity (AlgPred)	NON-ALLERGEN Score=-0.78769882 Threshold=-0.4	Non-allergen
Allergenicity (allergenfp v1.0)	Non-allergen	Non-allergen
ToxinPred	Non-toxic	Non-toxic

Construct-level antigenicity was evaluated using VaxiJen v2.0 (cutoff=0.4) and sequence-based ANTIGENpro. Allergenicity was evaluated using AlgPred and AllergenFP v1.0. Toxicity was evaluated using ToxinPred. *pI*: Isoelectric point; *GRAVY*: Grand average of hydropathy

Physicochemical Evaluation of the Vaccine

The physicochemical properties of the designed multi-epitope construct, including half-life, amino acid composition, molecular weight, aliphatic index, theoretical pI, GRAVY, and the *in silico*-predicted immunological properties (antigenicity and allergenicity), were assessed and are presented in table 1.

Secondary Structure Analysis

Secondary structure prediction by the SOPMA server shows that the protein composition consists of 20.51% alpha-helix, 31.41% extended strand, 12.18% beta turn, and 35.90% random coil. It showed a balanced secondary structure contributing to stability and flexibility. Moreover, the secondary structure of

the multi-epitope construct was obtained by the Psipred server, as shown in figure 2.

Predicted Tertiary Structure: Selected AlphaFold Model

The 3D model of the multi-epitope construct was generated using the AlphaFold server and visualized in PyMOL (<https://pymol.org/2/>). Among the server outputs, the top-ranked model was retained as the working structure for downstream analyses and is shown in (figure 3A).

MD Analysis and in Silico Structural Assessment

Specifically, the initial AlphaFold-predicted 3D model of the full-length multi-epitope construct was simulated for 200 ns in GROMACS before refinement (figure 3A).

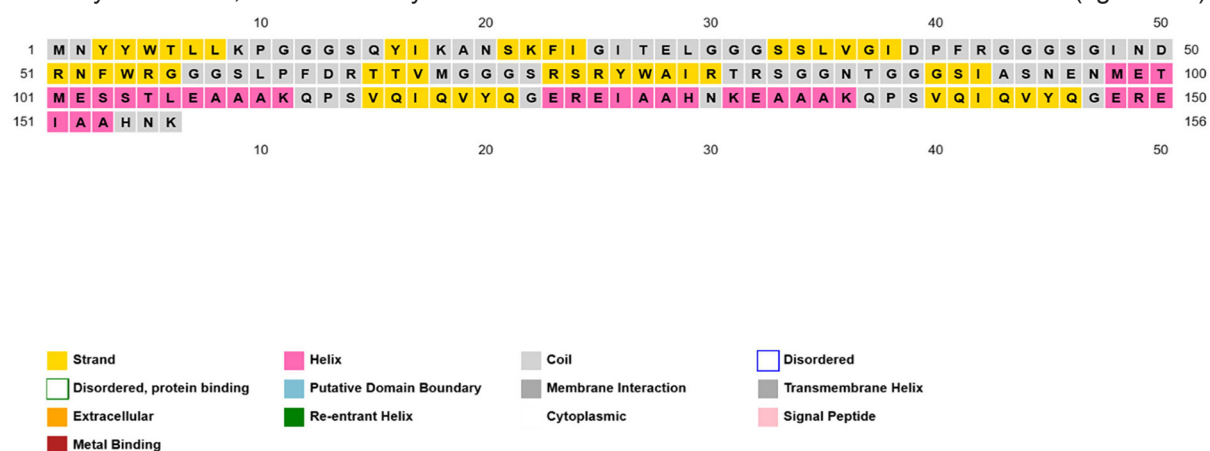


Figure 2: The amino acid sequence map of the final multi-epitope construct is shown here. It illustrates the predicted secondary structure of each amino acid residue using color shading: yellow for extended strands (β -sheets), pink for α -helices, and gray for random coils or loops. This schematic shows the linear arrangement and connectivity of the vaccine elements used in subsequent structural and immunoinformatics analyses.

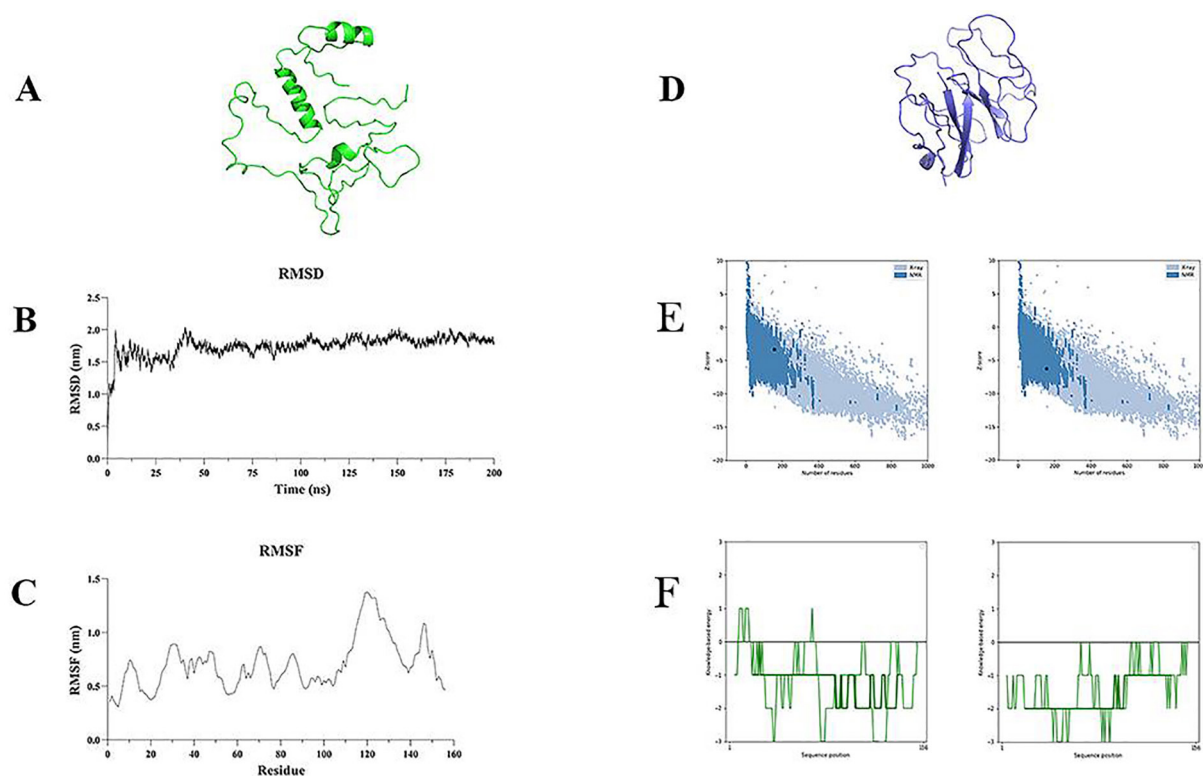


Figure 3: Structural and dynamic analyses of the designed multi-epitope construct are shown. (A) 3D structure of the designed multi-epitope construct in its initial form (AlphaFold output; pre-MD). (B) Final refined structure of the designed multi-epitope construct after energy minimization (post-MD and GalaxyRefine). (C&D) ProSA Z-score plots, showing the overall quality of the construct before (left) and after refinement (right), highlighting improvements in structural stability and accuracy. (E) RMSD plot for the full-length construct from a 200-ns molecular dynamics run. This shows the trajectory stabilizes in about 40 ns. (F) RMSF plot showing residue-specific flexibility, highlighting relatively stable and flexible regions in the simulation within the multi-epitope construct. MD: Molecular dynamics; RMSD: Root-mean-square deviation; RMSF: Root-mean-square fluctuation; ProSA: Protein structure analysis.

According to the RMSD plot, the trajectory reached a steady state after ~40 ns, suggesting simulated stabilization with minimal fluctuation (figure 3B). Therefore, RMSF analysis was conducted within this interval. The RMSF graph identified flexible regions mainly corresponding to loop regions (figure 3C). Following RMSD convergence, the stabilized MD frame was extracted and refined using GalaxyRefine (figure 3D). The ProSA Z-score of the initial model was -3.43 and improved to -6.75 after refinement (figure 3E). The ProSA local quality plot is shown in figure 3F.

The superimposed initial and refined structures are shown in figure 4A. For the initial model, the Ramachandran plot indicated that 54.8% of residues were found within the most favored region, with 44.4% within the allowed region, and 0.8% within the disallowed region (figure 4B). However, the refined model showed 79% of residues within the most favored region, with 18.9% being placed in the allowed region, and 2.4% remaining within the disallowed region (figure 4C). The overall ERRAT quality factor of the initial model was 26.786% (figure 4D),

which increased to 70.417% after GalaxyRefine refinement (figure 4E).

Cleavage sites and TAP Transport Efficiency Analysis

Selected peptides were analyzed for proteasomal cleavage and TAP transport suitability. The IEDB MHC class I prediction tools identified LPFDRTTVM, GINDRNFWR, and SLVGIDPFR peptides with favorable binding characteristics to multiple HLA class I alleles.

Subsequent Cleavage/TAP transport analysis revealed that LPFDRTTVM exhibits low affinity for the antigen processing-associated transporter (TAP). However, LPFDRTTVM was predicted to be efficiently processed by the proteasome, as evidenced by a proteasome score of 1.29. This high score suggests efficient generation during antigen processing.

GINDRNFWR and SLVGIDPFR peptides showed a significantly higher TAP score than the first peptide; their scores were 0.60 and 0.73, respectively, suggesting a stronger binding affinity to TAP. This enhanced affinity suggests better ER transport and MHC-I

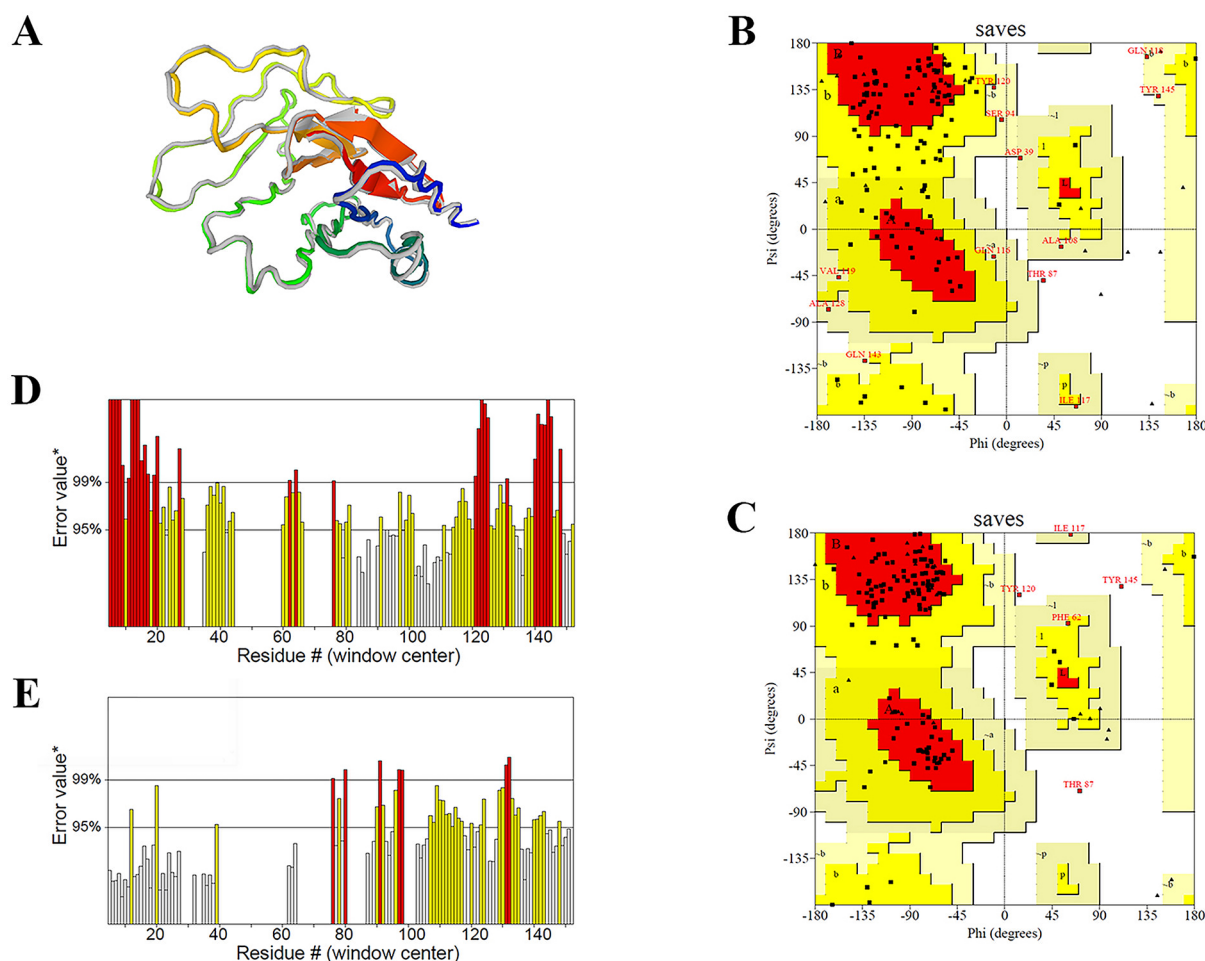


Figure 4: Structural quality assessment and refinement of the designed multi-epitope construct are shown. (A) Superimposed three-dimensional structures comparing the pre-MD AlphaFold model and the post-MD refined model. (B) Ramachandran plot prior to refinement showing residues in disallowed regions. (C) Ramachandran plot after refinement showing improved residue distribution in favored regions. (D) ERRAT quality factor plot before refinement indicating structural inconsistencies. (E) ERRAT plot after refinement showing improved structural reliability. MD: Molecular dynamics; ERRAT: Error recognition and analysis tool

presentation. The proteasome score was 1.06 for the GINDRNFWR peptide and 1.18 for the SLVGIDPFR peptide, slightly lower than LPFDRTTVM, but still reflecting a reasonable potential for processing by the proteasome. LPFDRTTVM showed the highest proteasome score, whereas GINDRNFWR and SLVGIDPFR showed higher TAP scores. The results are shown in supplementary table 10.

Molecular Docking and Analysis of Binding Affinity

We docked the final CTL and HTL peptides to the corresponding HLA class I and HLA class II molecules to reflect peptide presentation, and in parallel docked the full multi-epitope construct to TLR2 and TLR4 to probe a plausible *in silico* innate receptor interaction interface. The statistical parameters and their values for each docked vaccine–receptor complex are presented in table 2, including the HADDOCK

score, RMSD, van der Waals/electrostatic and desolvation energies, buried surface area (BSA), and Z-score. Docking interfaces and interacting amino acid residues are shown in figures 5A–B. Applying these criteria to the top clusters, the HLA class II–RSRYWAIRTRSGGNT complex ranked highest among epitope–HLA pairs, consistent with low RMSD and substantial buried surface area, supporting predicted stable HLA class II presentation. RSRWAIRTRSGGNT–HLA class II and vaccine–TLR4 showed the most favorable docking profiles.

The PDBsum results indicate that the complex between the vaccine and TLR2 consists of 32 residues from the vaccine and 27 residues from TLR2. The interface area for the vaccine is 1362 Å², while for TLR2, it is 1243 Å². The vaccine–TLR4 complex exhibited interactions between 39 vaccine residues and 48 residues of TLR4, with an interface area measuring 2040 Å² for the vaccine and 1990 Å² for TLR4.

Table 2: HADDOCK results for peptide–human leukocyte antigen (HLA) and vaccine–toll-like receptor (TLR) interactions

HADDOCK results	Complexes						
	HLA class I and SLVGIDPFR	HLA class I and GINDRNFWR	HLA class I and LPFDRTTVM	HLA class II and RSRYWAI RTRSGGNT	HLA class II and IASNNMETMESSTL	Vaccine and TLR2	Vaccine and TLR4
HADDOCK score	-97.5±2.3	-117.7±4.8	-99.0±5.2	-138.9±3.9	-125.2±7.6	-18.4±2.5	-55.6±4.6
Cluster size	145	82	43	183	78	8	18
RMSD from the overall lowest-energy structure	0.3±0.2	0.2±0.1	2.2±0.0	0.3±0.2	0.8±0.5	12.4±0.1	10.1±0.1
Van der Waals energy	-59.7±2.4	-50.2±1.3	-56.0±1.8	-81.4±5.3	-80.5±2.8	-71.4±8.1	-109.6±9.1
Electrostatic energy	-84.0±15.8	-160.6±8.4	-108.1±9.9	-185.3±24.4	-165.8±29.7	-621.0±17.6	-577.3±93.2
Desolvation energy	-22.1±3.1	-37.5±3.1	-22.8±2.8	-20.9±5.2	-11.8 ± 1.5	29.0±6.8	20.5±6.7
Restraints violation energy	11.5±15.8	20.7±14.0	14.0±17.4	3.7±2.5	1.2±0.9	1481.6±77.1	1489.6±108.6
Buried surface area	1498.9±13.5	1493.4±68.5	1554.7±52.2	2255.1±71.8	2139.7±69.8	3490.7±195.1	3898.3±182.2
Z-score	-1.4	-1.0	-1.2	-1.1	-1.5	-1.4	-1.6

Best water-refined clusters are reported as mean±SD, including cluster size, RMSD, binding energies, BSA, and Z-score. Ranking highlights RSRYWAI RTRSGGNT–HLA class II and a preferential vaccine–TLR4 interface. TLR models show broader dispersion by RMSD/restraint statistics and warrant routine experimental follow-up. HLA: Human leukocyte antigen; TLR: Toll-like receptor; RMSD: Root-mean-square deviation; BSA: Buried surface area; SD: Standard deviation

Table 3: PRODIGY-estimated binding affinities (ΔG and K_d) for epitope–human leukocyte antigen (HLA) class I/II and vaccine–toll-like receptor (TLR) complexes

Complex	Gibbs free energy (kcal mol ⁻¹)	K_d (M)
HLA class I vs SLVGIDPFR	-9.8	6.4E-08
HLA class I vs GINDRNFWR	-8.6	4.8E-07
HLA class I vs LPFDRTTVM	-9.6	9.1E-07
HLA class II vs RSRYWAI RTRSGGNT	-12.2	1.3E-09
HLA class II vs IASNNMETMESSTL	-10	1.1E-08
Vaccine vs TLR2	-9.2	1.3E-07
Vaccine vs TLR4	-21.0	1.28E-13

Thermodynamic estimates translate the modeled interfaces into a common scale for ranking interactions (more negative ΔG and lower K_d indicate stronger binding). The rankings favor the vaccine–TLR4 complex and, among peptides, the HLA class II–RSRYWAI RTRSGGNT pair, an orthogonal check to HADDOCK that motivates experimental confirmation. ΔG : Gibbs free energy; K_d : Dissociation constant; HLA: Human leukocyte antigen; TLR: Toll-like receptor; M: molar

The molecular interaction developed between vaccine–TLR2 involves three salt bridges, 16 hydrogen bonds, and 131 non-bonding contacts. However, the interaction with TLR4 involves five salt bridges, 24 hydrogen bonds, and 270 non-bonding contacts. Most vaccine–receptor hydrogen bonds were 2–3 Å, suggesting strong interactions. Results are presented in table 3.

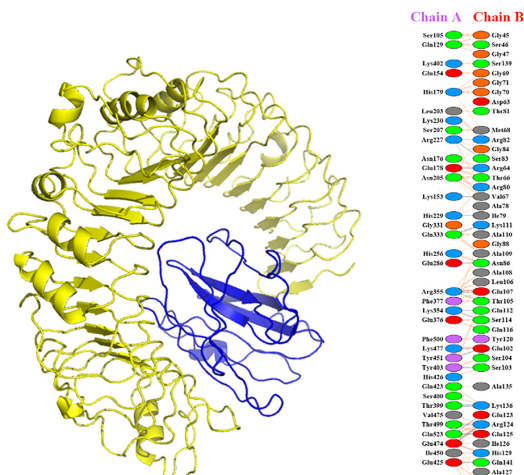
PRODIGY analysis estimated ΔG and K_d values for peptide–HLA class I/II and vaccine–TLR2/TLR4 complexes. Lower K_d values indicated stronger predicted vaccine–receptor affinity (table 2). The binding interactions analyzed show favorable predicted affinities, with ΔG ranging from -8.6 to -21.0 kcal/mol and dissociation constants (K_d) in the nanomolar to picomolar range. Notably, the strongest binding is observed between Vaccine and TLR4, with a

ΔG of -21.0 kcal/mol and a K_d of 1.28E-13 M. Furthermore, among peptides, the strongest interaction is between HLA class II and RSRYWAI RTRSGGNT (ΔG = -12.2 Kcal/mol, K_d =1.3E-09 M).

Optimization of Codons and Computational Cloning of a Vaccine

The candidate vaccine sequence was codon-optimized for expression in *Homo sapiens*, whereas BALB/c (H-2d) allele screening was used only during epitope selection to support murine preclinical testing and to identify epitopes with murine–human overlap. The optimized codon sequence, consisting of 549 nucleotides, exhibited a guanine-cytosine content (GC-content) of 68.45 and a codon adaptation index (CAI Value) of 0.94, as detailed in the supplementary material.

A



B

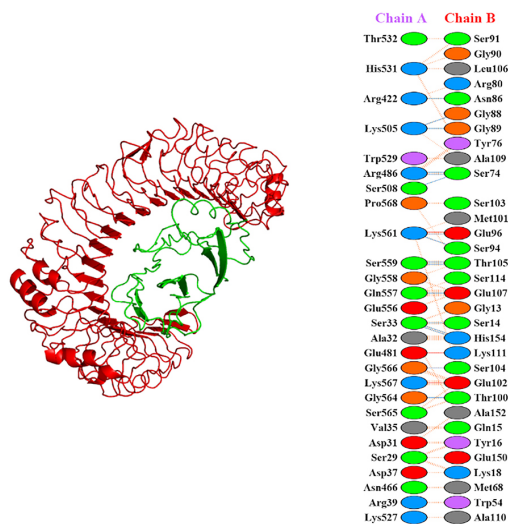


Figure 5: Molecular docking examination of the multi-epitope construct with TLR2 and TLR4 is shown. (A) Docking-predicted conformation of the multi-epitope construct (blue) with TLR2 receptor (yellow). The right panel lists the interacting residues between the vaccine (Chain A) and TLR2 (Chain B). (B) Predicted binding conformation of the multi-epitope construct (green) docked to the TLR4 receptor (red). The right panel displays the corresponding interacting residues between the vaccine (Chain A) and TLR4 (Chain B). TLR2: Toll-like receptor 2; TLR4: Toll-like receptor 4

Figure 6 presents the output from SnapGene software, illustrating the cloning of the multi-epitope construct coding sequence within the pcDNA3.1(+) vector under the CMV promoter/enhancer between the *Bam*HI and *Eco*RI sites to create a recombinant plasmid.

In Silico Immune Simulation

The immune simulation results indicated

an increasing level of IgM+IgG, IgM, IgG1, and IgG1+IgG2, particularly in secondary and tertiary immune responses (figure 7A). Consistent with the antibody-response pattern in figure 7A, plasma B cell (PLB) dynamics also increased after repeated exposures (figure 7B). After secondary and tertiary injections, TH memory cells showed higher levels than TH nonmemory cells (figure 7G). The elevated level of IFN- γ in the third injection may support a cytokine profile compatible with macrophage activation and MHC-II-associated antigen presentation pathways (figure 7C). The elevated level of TGF- β and interleukin-10 (IL-10) is consistent with regulatory and anti-inflammatory signaling in the *in silico* simulation on the designed multi-epitope construct (figure 7C). Active TH cells with the designed multi-epitope construct had a significant increase compared to resting TH cells (figure 7E). Moreover, the total count of B cells was increased after repeating exposure to the designed multi-epitope construct (figure 7D). In addition, the simulation suggested engagement of the cellular immune arm, as reflected by the Tc (cytotoxic T cell) population dynamics over time and the distribution of Tc cells across activation states following immunisation (figure 7F) and the T cell-oriented cytokine signals captured in figure 7C (e.g., IFN- γ and IL-2). Together with the observed helper T cell expansion/memory (figures 7E, 7G), these patterns suggested helper and cytotoxic T cell engagement in the *in silico* simulation.

Discussion

In this study, we designed an NP-centered DNA multi-epitope construct incorporating conserved CTL, HTL, and B cell epitopes together with an HA-derived CPP, VEGF-183 signal peptide, TT helper epitope, and Hsp70(407–426) adjuvant repeats. Overall, the construct showed favorable predicted antigenicity, non-allergenicity, non-toxicity, reference-based epitope conservancy, structural stability after MD refinement, favorable peptide–HLA and vaccine–TLR docking profiles, and immune simulation patterns consistent with antibody and T cell responses. These findings support the construct as a computationally prioritized vaccine candidate requiring experimental validation. Compared with Bazhan and others, who used experimentally confirmed T cell epitopes in a bivalent H1N1/H3N2 DNA vaccine, the present study combined predicted CTL, HTL, and B cell epitopes and additionally assessed TAP transport and proteasomal processing to support MHC-I presentation.⁵ Previous *in silico* influenza vaccine studies have also used HA/NA-, M2-, or NP-based designs.^{6,30,31}

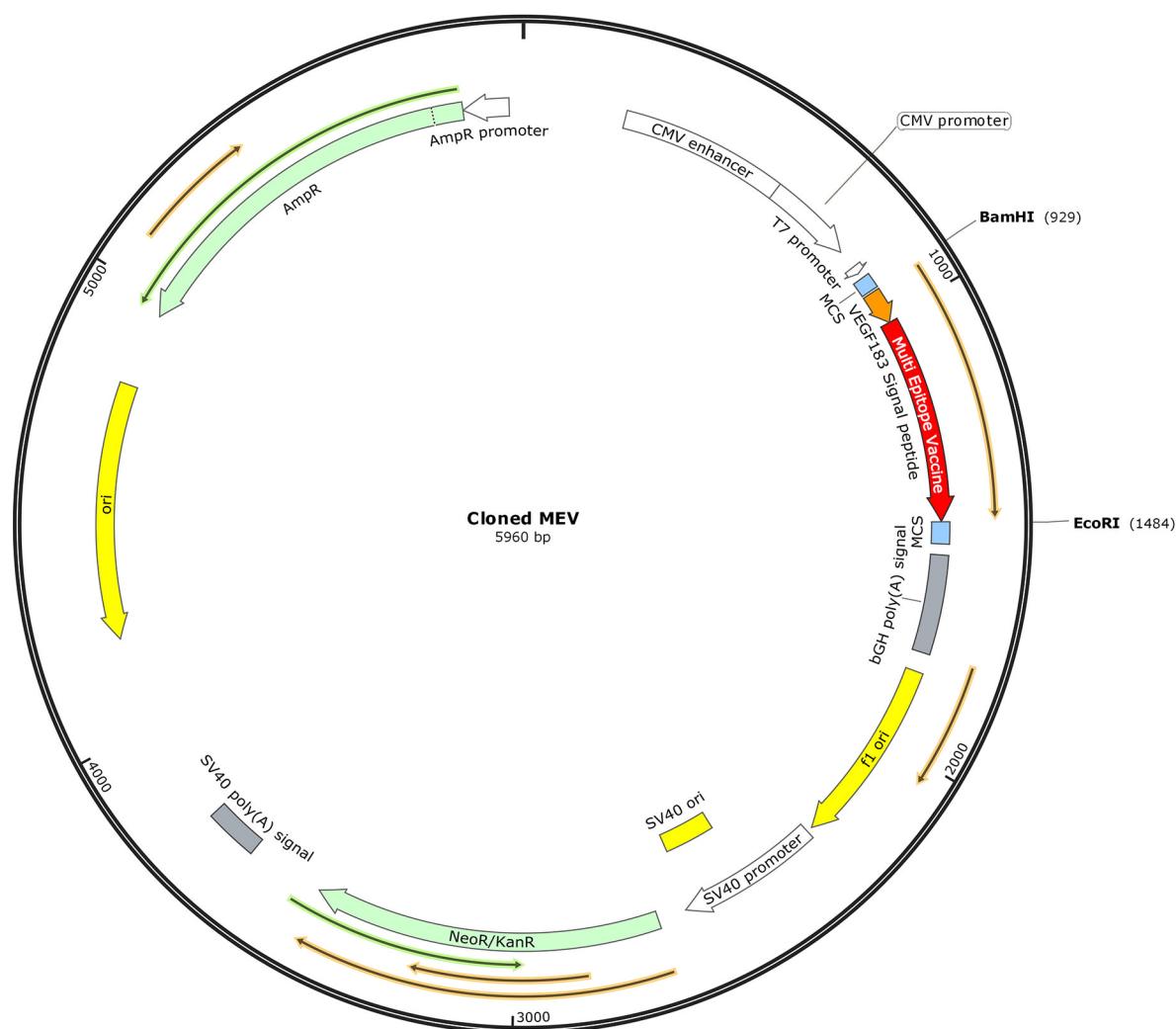


Figure 6: Plasmid map of the recombinant construct encoding the multi-epitope vaccine is shown. The codon-optimized multi-epitope construct coding sequence was successfully inserted between the *Bam*HI and *Eco*RI restriction sites under the control of a CMV promoter. The construct includes essential regulatory elements such as the T7 promoter, VEGF-183 signal peptide, and 6×His tag for expression and purification. Additional features such as AmpR, NeoR/KanR, SV40 poly(A) signal, and SV40 ori are incorporated to ensure selection and replication in bacterial and mammalian systems. CMV: Cytomegalovirus; VEGF-183: Vascular endothelial growth factor isoform 183; AmpR: Ampicillin resistance; NeoR/KanR: Neomycin/kanamycin resistance; SV40: Simian virus 40.

By focusing on conserved NP epitopes and assessing both humoral and cellular immune responses, this study extends previous work.

The HA-derived CPP MNYYWTLLKP was incorporated to support cellular uptake of the expressed construct and support antigen processing and presentation.³⁰ However, uptake and endosomal escape are distinct processes, and endosomal escape was not evaluated in this *in silico* study.

The CPP was positioned immediately downstream of the VEGF-183 signal peptide so that, after signal peptide cleavage, it remains at the N-terminus of the mature expressed protein and is discussed strictly as an intracellular delivery-facilitating tag (not as a tissue-targeting or antiviral element).

Because the CPP is derived from influenza

HA, it preserves sequence continuity with a viral antigen. Any antibody responses to this HA-derived segment and their relevance to protection remain speculative and require experimental confirmation. Accordingly, the CPP is discussed here strictly as a delivery-facilitating element rather than a validated targeting or antiviral component.

The VEGF-183 signal peptide was included to support ER targeting and secretory routing of the expressed construct.³¹ The TT helper epitope was included to support the activation of CD4⁺ T cells, while two Hsp70 repeats (407-426) were used as a molecular adjuvant.³² This fragment increases the synthesis of inflammatory cytokines and facilitates the maturation of dendritic cells, thereby enhancing the overall immune response to the vaccine.¹²

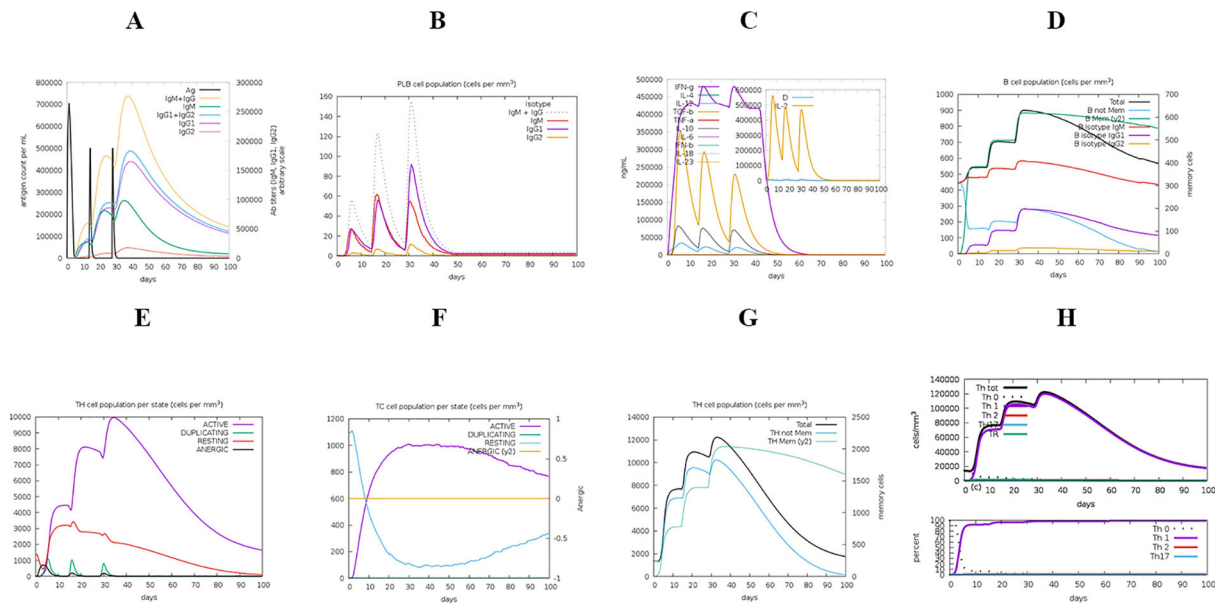


Figure 7: *In silico* immune simulation of the multi-epitope construct is presented. A) Antibody titers (IgM, IgG, IgG1, IgG2) after 100 days with a typical primary and an enhanced secondary antibody response profile. B) Plasma B cell population by state (active, rest) with time, revealing activation and antibody production kinetics. C) Cytokine secretion levels (e.g., IFN- γ , IL-2, IL-10, IL-12, IL-4, TNF- α) throughout the simulation period, highlighting the pro-inflammatory and regulatory responses. D) Total size of all B cell, including naïve, activated, memory, over time course to reveal development and persistence of humoral immune response. E) TH population by state over 100 days (active, duplicating, resting, anergic) across primary, secondary, and tertiary responses, showing a booster-driven expansion of active TH cells followed by contraction and persistence of the resting/anergic pools. F) TC population with time by state (active, duplicating, resting) depicting kinetics of T cytotoxic cell response to immunisation. G) Memory and active TH cell population dynamics, providing insights into long-term helper T cell-mediated immunity. H: Differentiation of TH subset, such as TH0, TH1, TH2, and TH17, changes with time to represent polarization of immune responses to different T-helper lineages. IgM: Immunoglobulin M; IgG: Immunoglobulin G; TH: T helper cell; TC: Cytotoxic T cell; IFN- γ : Interferon gamma; IL: Interleukin; TNF- α : Tumor necrosis factor alpha

A recent study showed that tandem mHSP70(407–426) repeats enhanced vaccine efficacy by strengthening CTL responses in mouse models.³³ Flexible (GGGS) and rigid (EAAAK) linkers were utilized (as detailed in the methods) to ensure proper spacing and folding of vaccine domains, which helps support the construct's stable expression.³⁴

Notably, our construct's physicochemical profile, together with predicted HLA class I/II binding and B-/T cell epitope features, supported its prioritization as a vaccine candidate. The selected murine and human MHC-I epitopes showed low percentile ranks and favorable antigenicity/safety profiles, supporting their prioritization for experimental validation. Selection of LPFDRTTVM as a non-toxic and non-allergenic epitope reinforces its relevance to vaccine development and may positively contribute to the overall vaccine safety profile.

For HLA class I analysis, broad-based allelic coverage covering HLA-B*35:01 and HLA-B*07:02 suggests predicted broad population coverage across the analyzed HLA allele panel. The low percentile ranks of epitopes such as LPFDRTTVM, GINDRNFWR, and SLVGIDPFR emphasize their strong binding affinity, which is essential for effective antigen presentation and

subsequent T cell activation. The murine MHC class II (I-Ad/I-Ed) and HLA class II analysis, focusing on longer peptides (15-mers), reveals a spectrum of binding affinities and good antigenicity scores. In the analysis of BALB/c mice, eight epitopes were identified in alleles (H2-IEd, H2-IAd) with low percentile ranks and high antigen scores for graft affinity, providing suitable candidates for experimental validation. The RSRYWAI RTRSGGNT epitope selection had a low percentile rank and was shared between murine MHC class II (H2-IEd) and multiple HLA class II alleles (HLA-DRB108:02, HLA-DRB111:01, HLA-DRB1*04:01). In addition, the abundance of the mentioned epitope in the human allele has made it a suitable epitope for predicted antigen presentation and subsequent activation of T cells as a suitable candidate for the designed vaccine.

The overlap of SLVGIDPFR with a predicted linear B cell epitope suggests potential dual relevance for cellular and humoral responses, while IASNENMETMESSTL was retained as a non-toxic and non-allergenic B cell epitope.

We also assessed HLA allele frequencies to ensure the vaccine's relevance across populations. Notably, one of our top epitopes (GINDRNFWR) is predicted to cover a

substantial portion of the target population's alleles (including roughly a quarter of Iranian individuals), supporting the broad applicability of our vaccine candidate.

Our analysis of antigen processing (proteasomal cleavage/TAP transfer/MHC class I) also showed that the epitopes selected by the IEDB HLA class I/II prediction tools exhibited allele-specific high-affinity binding, together providing complementary coverage across multiple HLA alleles. Moreover, they have complementary strengths. For example, one epitope (LPFDRTTVM) is highly efficiently processed (although less efficiently transported by TAP), while other epitopes (GINDRNFWR and SLVGIDPFR) have higher TAP affinity but slightly lower processing efficiency. This balance suggests that these epitopes, in combination, can ensure predicted HLA class I presentation by covering multiple steps of the processing pathway and may improve immune-response coverage when used together.³⁵ As an additional design consideration, we note that epitope-linker junctions could theoretically generate neo-epitopes; a dedicated junctional scan of overlapping 9-mers/15-mers spanning all junctions would be a useful future screening step.³⁶

The secondary structure analysis showed that the vaccine has alpha helices, extended strands, beta turns, and random coil structures. We subjected the predicted structure required for stability and energy minimization to molecular dynamics simulations for 200 nanoseconds. The structure showed simulated stabilization after ~45 ns, as indicated by the RMSD, providing a post-stabilization conformation for subsequent docking. The 3D structure of the vaccine was predicted using reliable deep learning software and subsequently reconstructed with additional software, focusing on conformational shape, energy minimization, and atomic distances. The refinement and *in silico* validation of the vaccine model suggested improved structural quality. The findings were further validated by predictions from the Ramachandran plot.³⁷

Although our multi-epitope construct includes both T cell and B cell epitopes to elicit adaptive immunity, we applied a docking-based approach tailored to the nature of each immune receptor. Since MHC molecules present processed peptides, docking the full multi-epitope construct with MHC would not be biologically meaningful. Therefore, docking was performed at the epitope level for both representative HLA class I and HLA class II structures. In contrast, TLRs recognize intact extracellular protein structures; thus, we docked the full multi-epitope construct

to TLR2 and TLR4 to explore predicted *in silico* interfaces relevant to innate receptor engagement. We limited this analysis to TLR2 and TLR4 because they are well-studied surface TLRs implicated in APC activation and in shaping adaptive immunity, and because Hsp70-related immunostimulatory signaling has been reported to involve TLR4 and, in some contexts, TLR2.¹¹ Evaluation of additional TLRs was outside the scope of this work and remains for future studies. We performed protein-peptide docking to investigate peptide-HLA class I and peptide-HLA class II interactions with selected peptides, and protein-protein docking to evaluate interactions between the full multi-epitope construct and TLR2/TLR4. The docking results showed favorable scores and negative ΔG values for peptide-HLA class I, peptide-HLA class II, and vaccine-TLR complexes, with the strongest predicted interaction observed for TLR4. The docking analysis of the vaccine with receptors indicates the presence of salt bridges and hydrogen bonds, which provide *in silico* support for interaction. Among the dockings performed, the top epitope-HLA result was the HLA class II-RSRYWAIRTRSGGNT complex, and for the full construct the interface appeared more favorable with TLR4 than TLR2. However, broader TLR docking dispersion warrants cautious interpretation. These conclusions are model-based and await experimental validation.

Finally, an immune simulation was performed to replicate standard immune responses, showing a gradual rise in immunoglobulins (IgM+IgG, IgM, IgG1, and IgG1+IgG2) after multiple exposures to the engineered Multi-epitope construct, especially noted during secondary and tertiary immune responses. This is consistent with the predicted production of memory B cells. Additionally, the comparison of elevated TH memory cell levels to non-memory TH cells following secondary and tertiary injections, combined with the rise in interferon-gamma (IFN- γ) levels, may support a cytokine profile compatible with macrophage activation and MHC-II-associated antigen presentation pathways. The observed increase in TGF- β and IL-10 indicates a potential anti-inflammatory role for the multi-epitope construct. The immune simulation results indicated a notable activation of TH cells in comparison to resting TH cells, along with an increase in the overall B cell count following repeated exposures. The data suggest that three injections of the developed multi-epitope construct may induce a robust immune response *in silico*. Overall, our findings support the feasibility of a multi-epitope DNA vaccine centered on conserved

NP regions and augmented with an HA-derived CPP, the TT830–844 helper peptide, and Hsp70(407–426) to broaden immune activation. Integrated *in silico* analyses, including epitope prioritization, antigen-processing and binding predictions, structural modeling/docking, and immune-response simulations, indicate a stable construct with favorable immunogenic potential. The limitations of this study are as follows: the study is based entirely on *in silico* analyses, meaning that the findings require empirical validation before experimental or practical translation. Although structural refinement was performed to improve the stereochemical quality of the designed construct, Ramachandran plot analysis revealed that 79% of the residues in the refined model were located within the most favored regions. This percentage remains below the commonly accepted benchmark of 90%, indicating that further structural optimization is required. Therefore, the stereochemical quality of the predicted three-dimensional structure should be considered a limitation of the present study. Future investigations should focus on additional structural refinement and validation approaches to improve the overall quality and reliability of the three-dimensional model before proceeding with further experimental evaluations. Epitope conservancy was assessed in a reference-based manner rather than across a broad panel of circulating influenza A sequences; the 3D structure, docking interfaces, and PRODIGY affinity estimates are model-dependent predictions and do not demonstrate receptor engagement or biological function; and C-IMMSIM outputs provide computational approximations of immune dynamics. In addition, antigen expression/secretion, delivery efficiency constraints (e.g., cellular uptake and endosomal escape), and safety/protective efficacy were not experimentally evaluated. Therefore, these findings should be interpreted as computational prioritization and warrant *in vitro* expression validation and *in vivo* immunogenicity and challenge studies to confirm immunogenicity, safety, and protection.

Conclusion

In this study, we designed an influenza A NP-based DNA multi-epitope construct incorporating an HA-derived CPP, a VEGF-183 signal peptide, a TT helper epitope, and two Hsp70(407–426) repeats, and evaluated it using an immunoinformatics and structure-guided *in silico* workflow. The shortlisted CTL/HTL and linear B cell epitopes were predicted antigenic, non-toxic, and non-allergenic, and

showed reference-based epitope mapping/conservancy. MD analysis suggested simulated stabilization of the modeled construct, and *in silico* docking/affinity estimates supported favorable predicted interfaces for selected epitope–HLA class I/II complexes and construct–TLR2/TLR4 complexes, with stronger modeled interaction metrics for TLR4. Immune simulation outputs further suggested antibody and T cell–associated response patterns across repeated *in silico* exposures. Collectively, these computational results prioritize the proposed construct for subsequent experimental validation. Accordingly, experimental validation should include cloning the human codon-optimized insert into pcDNA3.1(+), confirming mammalian expression, and assessing immunogenicity, safety, and protection in BALB/c influenza challenge studies.

Acknowledgment

This manuscript was extracted from a student's thesis, Sanaz Sadeh. We would like to gratefully thank the Influenza and Respiratory Viruses Department, Pasteur Institute of Iran for their scientific support in the study. This work was supported by the INSF grant number: Iran National Science Foundation (INSF) - 4046527 to Amir Ghaemi.

Authors' Contribution

S.S: Conceptualization, study design, formal analysis, data interpretation, drafting, and reviewing the manuscript; A.G: Conceptualization, study design, project administration, supervision, validation, drafting, and reviewing the manuscript; N.A.S: Conceptualization, study design, project administration, drafting, and reviewing the manuscript; M.M: Conceptualization, study design, formal analysis, drafting, and reviewing the manuscript; G.B: Study design, project administration, supervision, data curation, drafting, and reviewing the manuscript; All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work, ensuring that any questions related to the accuracy or integrity of any part of the study are appropriately investigated and resolved.

Declaration of AI

The authors used artificial intelligence–assisted tools, including Grammarly and ChatGPT, only for grammar correction, and manuscript editing. All scientific content, interpretations, analyses,

and conclusions were developed, reviewed, and approved solely by the authors. No artificial intelligence tool was used for data generation, and analysis.

Conflict of Interest: None declared.

References

- 1 Parvez S, Pathrathota A, Uppar AL, Yadagiri G, Mudavath SL. Influenza Virus: Global Health Impact, Strategies, Challenges, Role of Nanotechnology in Influenza Vaccine Development. *Vaccines (Basel)*. 2025;13. doi: 10.3390/vaccines13090890. PubMed PMID: 41012096; PubMed Central PMCID: PMC12474260.
- 2 Lu B, Lim JM, Yu B, Song S, Neeli P, Sobhani N, et al. The next-generation DNA vaccine platforms and delivery systems: advances, challenges and prospects. *Front Immunol*. 2024;15:1332939. doi: 10.3389/fimmu.2024.1332939. PubMed PMID: 38361919; PubMed Central PMCID: PMC10867258.
- 3 Suschak JJ, Williams JA, Schmaljohn CS. Advancements in DNA vaccine vectors, non-mechanical delivery methods, and molecular adjuvants to increase immunogenicity. *Hum Vaccin Immunother*. 2017;13:2837-48. doi: 10.1080/21645515.2017.1330236. PubMed PMID: 28604157; PubMed Central PMCID: PMC5718814.
- 4 Bahrami AA, Payandeh Z, Khalili S, Zakeri A, Bandehpour M. Immunoinformatics: In Silico Approaches and Computational Design of a Multi-epitope, Immunogenic Protein. *Int Rev Immunol*. 2019;38:307-22. doi: 10.1080/08830185.2019.1657426. PubMed PMID: 31478759.
- 5 Bazhan SI, Antonets DV, Starostina EV, Ilyicheva TN, Kaplina ON, Marchenko VY, et al. In silico design of influenza a virus artificial epitope-based T-cell antigens and the evaluation of their immunogenicity in mice. *J Biomol Struct Dyn*. 2022;40:3196-212. doi: 10.1080/07391102.2020.1845978. PubMed PMID: 33222632.
- 6 Yuan L, Li X, Li M, Bi R, Li Y, Song J, et al. In silico design of a broad-spectrum multi-epitope vaccine against influenza virus. *Int J Biol Macromol*. 2024;254:128071. doi: 10.1016/j.ijbiomac.2023.128071. PubMed PMID: 37967595.
- 7 Santak M, Matic Z. The Role of Nucleo-protein in Immunity to Human Negative-Stranded RNA Viruses-Not Just Another Brick in the Viral Nucleocapsid. *Viruses*. 2022;14. doi: 10.3390/v14030521. PubMed PMID: 35336928; PubMed Central PMCID: PMC8955406.
- 8 Sadeh S, Ghaemi A, Soleimani NA, Moghbeli M, BahramAli G. Designing influenza virus-derived cell-penetrating peptides for antigen delivery: Integrating uptake efficiency, safety, and receptor targeting. *PLoS One*. 2025;20:e0338028. doi: 10.1371/journal.pone.0338028. PubMed PMID: 41364690; PubMed Central PMCID: PMC12688144.
- 9 Apte RS, Chen DS, Ferrara N. VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell*. 2019;176:1248-64. doi: 10.1016/j.cell.2019.01.021. PubMed PMID: 30849371; PubMed Central PMCID: PMC6410740.
- 10 Facciola A, Visalli G, Lagana A, Di Pietro A. An Overview of Vaccine Adjuvants: Current Evidence and Future Perspectives. *Vaccines (Basel)*. 2022;10. doi: 10.3390/vaccines10050819. PubMed PMID: 35632575; PubMed Central PMCID: PMC9147349.
- 11 Hu B, Liu G, Zhao K, Zhang G. Diversity of extracellular HSP70 in cancer: advancing from a molecular biomarker to a novel therapeutic target. *Front Oncol*. 2024;14:1388999. doi: 10.3389/fonc.2024.1388999. PubMed PMID: 38646439; PubMed Central PMCID: PMC11026673.
- 12 Xu M, Zhang Y, Dong W, Jiang L, Zhang J, Yu P, et al. Two tandem repeats of mHSP70(407-426) enhance therapeutic antitumor effects of a recombinant vascular endothelial growth factor (VEGF) protein vaccine. *Life Sci*. 2018;201:102-10. doi: 10.1016/j.lfs.2018.03.039. PubMed PMID: 29572180.
- 13 Esmaeili A, Rabe SZT, Mahmoudi M, Rastin M. Frequencies of HLA-A, B and DRB1 alleles in a large normal population living in the city of Mashhad, Northeastern Iran. *Iran J Basic Med Sci*. 2017;20:940-3. doi: 10.22038/IJBMS.2017.9117. PubMed PMID: 29085586; PubMed Central PMCID: PMC5651480.
- 14 Jalalvand A, Fotouhi F, Bahramali G, Bambai B, Farahmand B. In silico design of a trivalent multi-epitope global-coverage vaccine-candidate protein against influenza viruses: evaluation by molecular dynamics and immune system simulation. *J Biomol Struct Dyn*. 2025;43:1522-38. doi: 10.1080/07391102.2023.2292293. PubMed PMID: 38088331.
- 15 Alnajran H, Awadalla M, Aldakheel FM, Alam I, Momin AA, Alturaiki W, et al. Design of a peptide-based vaccine against human

- respiratory syncytial virus using a reverse vaccinology approach: evaluation of immunogenicity, antigenicity, allergenicity, and toxicity. *Front Immunol.* 2025;16:1546254. doi: 10.3389/fimmu.2025.1546254. PubMed PMID: 40226615; PubMed Central PMCID: PMC11986473.
- 16 Tan YC, Lahiri C. Promising *Acinetobacter baumannii* Vaccine Candidates and Drug Targets in Recent Years. *Front Immunol.* 2022;13:900509. doi: 10.3389/fimmu.2022.900509. PubMed PMID: 35720310; PubMed Central PMCID: PMC9204607.
 - 17 Kardani K, Bolhassani A. Exploring novel and potent cell penetrating peptides in the proteome of SARS-COV-2 using bioinformatics approaches. *PLoS One.* 2021;16:e0247396. doi: 10.1371/journal.pone.0247396. PubMed PMID: 33606823; PubMed Central PMCID: PMC7894964.
 - 18 Lu Y, Mekoo DJ, Ouyang K, Hu X, Liu Y, Lin M, et al. Strong humoral response elicited by a DNA vaccine targeting gastrin-releasing peptide with optimized adjuvants inhibits murine prostate carcinoma growth in vivo. *Endocr Relat Cancer.* 2009;16:1171-84. doi: 10.1677/ERC-09-0058. PubMed PMID: 19648182.
 - 19 van Rosmalen M, Krom M, Merks M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry.* 2017;56:6565-74. doi: 10.1021/acs.biochem.7b00902. PubMed PMID: 29168376; PubMed Central PMCID: PMC6150656.
 - 20 Han L, Wang X-F, Zhang S-J, Wang Y-Z, Zheng Z-H, Zhang W-J, et al. Bioinformatic analysis of the MprA protein in *Mycobacterium tuberculosis*. *CABI.* 2020: 1171-6.
 - 21 Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 2024;52:D368-D75. doi: 10.1093/nar/gkad1011. PubMed PMID: 37933859; PubMed Central PMCID: PMC10767828.
 - 22 Lee GR, Won J, Heo L, Seok C. GalaxyRefine2: simultaneous refinement of inaccurate local regions and overall protein structure. *Nucleic Acids Res.* 2019;47:W451-W5. doi: 10.1093/nar/gkz288. PubMed PMID: 31001635; PubMed Central PMCID: PMC6602442.
 - 23 Bagga P, Singh S, Ram G, Kapil S, Eerapagula R, Puri P, et al. Molecular Dynamics Simulation and Structural Insights into The SMN1 Protein Involved in the Pathogenesis of Spinal Muscular Atrophy. *Journal of Neonatal Surgery.* 2025;14:1066-77. doi: 10.63682/jns.v14i13S.3935.
 - 24 Pishesha N, Harmand TJ, Ploegh HL. A guide to antigen processing and presentation. *Nat Rev Immunol.* 2022;22:751-64. doi: 10.1038/s41577-022-00707-2. PubMed PMID: 35418563.
 - 25 Mou SI, Sultana T, Chatterjee D, Faruk MO, Hosen MI. Comprehensive characterization of coding and non-coding single nucleotide polymorphisms of the Myoneurin (MYNN) gene using molecular dynamics simulation and docking approaches. *PLoS One.* 2024;19:e0296361. doi: 10.1371/journal.pone.0296361. PubMed PMID: 38165846; PubMed Central PMCID: PMC10760682.
 - 26 Honorato RV, Trellet ME, Jimenez-Garcia B, Schaarschmidt JJ, Giuliani M, Reys V, et al. The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nat Protoc.* 2024;19:3219-41. doi: 10.1038/s41596-024-01011-0. PubMed PMID: 38886530.
 - 27 Demissie EA, Park SY, Moon JH, Lee DY. Comparative Analysis of Codon Optimization Tools: Advancing toward a Multi-Criteria Framework for Synthetic Gene Design. *J Microbiol Biotechnol.* 2025;35:e2411066. doi: 10.4014/jmb.2411.11066. PubMed PMID: 40223268; PubMed Central PMCID: PMC12010093.
 - 28 Chehelgerdi M, Doosti A. Effect of the cagW-based gene vaccine on the immunologic properties of BALB/c mouse: an efficient candidate for *Helicobacter pylori* DNA vaccine. *J Nanobiotechnology.* 2020;18:63. doi: 10.1186/s12951-020-00618-1. PubMed PMID: 32316990; PubMed Central PMCID: PMC7175550.
 - 29 Cheng P, Xue Y, Wang J, Jia Z, Wang L, Gong W. Evaluation of the consistence between the results of immunoinformatics predictions and real-world animal experiments of a new tuberculosis vaccine MP3RT. *Front Cell Infect Microbiol.* 2022;12:1047306. doi: 10.3389/fcimb.2022.1047306. PubMed PMID: 36405961; PubMed Central PMCID: PMC9666678.
 - 30 Rostaminia S, Aghaei SS, Farahmand B, Nazari R, Ghaemi A. Computational Design and Analysis of a Multi-epitope Against Influenza A virus. *Int J Pept Res Ther.* 2021;27:2625-38. doi: 10.1007/s10989-021-10278-w. PubMed PMID: 34539293; PubMed Central PMCID: PMC8435298.
 - 31 Rostaminia S, Aghaei SS, Farahmand B,

- Nazari R, Ghaemi A. In-Silico Design of a Multi epitope Construct Against Influenza A Based on Nucleoprotein Gene. *International Journal of Peptide Research and Therapeutics*. 2022;28:117. doi: 10.1007/s10989-022-10418-w.
- 32 Lee LYY, Izzard L, Hurt AC. A Review of DNA Vaccines Against Influenza. *Front Immunol*. 2018;9:1568. doi: 10.3389/fimmu.2018.01568. PubMed PMID: 30038621; PubMed Central PMCID: PMC6046547.
- 33 Xu M, Zhou L, Zhang Y, Xie Z, Zhang J, Guo L, et al. A Fixed Human Umbilical Vein Endothelial Cell Vaccine With 2 Tandem Repeats of Microbial HSP70 Peptide Epitope 407-426 As Adjuvant for Therapy of Hepatoma in Mice. *J Immunother*. 2015;38:276-84. doi: 10.1097/CJI.0000000000000091. PubMed PMID: 26261891.
- 34 Shahid F, Ashfaq UA, Javaid A, Khalid H. Immunoinformatics guided rational design of a next generation multi epitope based peptide (MEBP) vaccine by exploring Zika virus proteome. *Infect Genet Evol*. 2020;80:104199. doi: 10.1016/j.meegid.2020.104199. PubMed PMID: 31962160.
- 35 Bahrami AA, Bandehpour M, Khalesi B, Kazemi B. Computational Design and Analysis of a Poly-Epitope Fusion Protein: A New Vaccine Candidate for Hepatitis and Poliovirus. *International Journal of Peptide Research and Therapeutics*. 2020;26:389-403. doi: 10.1007/s10989-019-09845-z.
- 36 Dorigatti E, Schubert B. Graph-theoretical formulation of the generalized epitope-based vaccine design problem. *PLoS Comput Biol*. 2020;16:e1008237. doi: 10.1371/journal.pcbi.1008237. PubMed PMID: 33095790; PubMed Central PMCID: PMC7652351.
- 37 Varghese S, Thomas B. Homology modelling and structure evaluation of human GRIN 1 protein. *Aureole*. 2021;13:40-5.