INTERNATIONAL JOURNAL OF APPLIED AND CLINICAL RESEARCH

IJACR

Research paper

Immunoinformatics-Based Multi-Epitope Vaccine Design Against Influenza A (H1N1) Hemagglutinin

Roshni Khan[1]*, Salman Khan[2]

1. Center of Biotechnology and Microbiology, University of Peshawar, University Campus, Peshawar 25120, Khyber Pakhtunkhwa, Pakistan
2. Department of Chemistry, Islamia College Peshawar, Jamrud Road, University Campus, Peshawar 25120, Khyber Pakhtunkhwa, Pakistan

Volume No 04, Issue 03, 2026

Received: 15th June 2026
Accepted: 25th July 2026
Published: 27th July 2026
Doi: https://doi.org/10.66222/IJACR.04.03.67

How to cite
Khan R, Khan S. Immunoinformatics-based multi-epitope vaccine design against Influenza A (H1N1) hemagglutinin. *Int J Appl Clin Res.* 2026;4(3):1-15.
Corresponding author
Roshni khan
Center of Biotechnology and Microbiology, University of Peshawar, University Campus, Peshawar 25120, Khyber Pakhtunkhwa, Pakistan
Email: roshnimicro058@gmail.com

Abstract: Influenza A (H1N1) remains a significant global health threat due to its high mutation rate and antigenic variability, which limit the long-term efficacy of conventional strain-specific vaccines. This study employed an Immunoinformatics approach to design a broadly protective multi-epitope vaccine targeting the hemagglutinin (HA) protein. **Methods:** The HA protein sequence was analyzed for physicochemical properties, antigenicity, and epitope prediction. Promising B-cell and T-cell epitopes were selected and assembled into a multi-epitope vaccine construct. Structural modeling, molecular docking with Toll-like receptor 3 (TLR3), molecular dynamics simulation, population coverage analysis, codon optimization, and in silico cloning were performed to evaluate the vaccine candidate. **Results:** The HA protein exhibited favorable physicochemical characteristics and strong antigenicity. The final vaccine construct was predicted to be highly antigenic, non-allergenic, and non-toxic, with a global population coverage of 81.68%. Structural validation confirmed model quality, while docking and molecular dynamics analyses demonstrated stable interactions with TLR3, indicating its potential to induce robust immune responses. Codon optimization and *in-silico* cloning suggested efficient expression in the host system. **Conclusion:** The designed HA-based multi-epitope vaccine demonstrated promising immunogenicity, safety, structural stability, and broad population coverage in silico. These findings support its potential as a vaccine candidate against Influenza A (H1N1), warranting further experimental validation through in vitro and in vivo studies.

Keywords: Influenza A (H1N1); Hemagglutinin; Multi-epitope vaccine; Immunoinformatics; Reverse vaccinology; Molecular docking; Molecular dynamics simulation; Epitope prediction.

Introduction

Influenza A virus (IAV) remains one of the most significant respiratory pathogens, causing seasonal epidemics and recurrent pandemics that pose substantial public health and socioeconomic challenges worldwide. Annual influenza epidemics are associated with millions of severe infections and hundreds of thousands of deaths, while the virus continues to evolve through antigenic drift and antigenic shift, facilitating the emergence of novel strains with pandemic potential [1–3]. The segmented negative-sense RNA genome of IAV enables frequent genetic reassortment, resulting in continuous viral evolution and reduced effectiveness of existing preventive strategies [4]. Among the viral proteins, hemagglutinin (HA) is the principal surface glycoprotein responsible for host-cell attachment and membrane fusion during viral entry. As the primary target of neutralizing antibodies, HA plays a pivotal role in viral infectivity and immune recognition, making it the most attractive antigen for influenza vaccine development [5,6]. However, the high mutation rate within the immunodominant regions of HA often leads to immune escape, requiring annual reformulation of seasonal influenza vaccines and limiting their cross-protective efficacy against newly emerging vari-

ants [7,8]. Recent advances in computational biology have accelerated vaccine development through reverse vaccinology and Immunoinformatics. These approaches enable rapid identification of immunodominant B-cell and T-cell epitopes with high antigenicity while simultaneously evaluating allergenicity, toxicity, physicochemical properties, structural stability, and population coverage before experimental validation [9–11]. Compared with conventional vaccines, multi-epitope vaccines offer several advantages, including improved safety, enhanced specificity, broad-spectrum immune responses, and reduced production time and cost [12,13]. Furthermore, molecular docking and molecular dynamics simulations provide valuable insights into vaccine–receptor interactions, structural stability, and the potential to activate innate immune receptors such as Toll-like receptor 3 (TLR3), which plays a crucial role in antiviral immunity [14,15]. Despite continuous advances in influenza vaccine development, the rapid genetic evolution and antigenic diversity of Influenza A (H1N1) continue to compromise the efficacy of conventional strain-specific vaccines, highlighting the need for broadly protective and durable vaccine strategies. To address this challenge, the present study employed a comprehensive Immunoinformatics-driven reverse vaccinology approach to rationally design a novel HA-based multi-epitope vaccine candidate.

Materials and Methods

Sequence Retrieval and Preliminary Analysis

The amino acid sequence of the hemagglutinin (HA) protein of Influenza A (H1N1) virus was retrieved in FASTA format from the UniProt Knowledgebase (UniProt) and the National Center for Biotechnology Information (NCBI) Influenza Virus Resource. The retrieved sequence served as the reference antigen for all subsequent Immunoinformatics analyses. The physicochemical characteristics of the HA protein, including amino acid composition, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY), were determined using the ExPASy ProtParam server. Secondary structural elements, including α-helices, β-sheets, and random coils, were predicted using the PSIPRED server, whereas disulfide bond formation was analyzed using the SCRATCH Protein Predictor. The three-dimensional (3D) structure of the HA protein was generated using the rosette server. To evaluate its suitability as a vaccine target, antigenicity and allergenicity were predicted using VaxiJen v2.0 and AllerTOP v2.1 [16-19].

Prediction and Screening of Immunogenic Epitopes

Potential cytotoxic T-lymphocyte (CTL), helper T-lymphocyte (HTL), and linear B-cell epitopes were identified using the Immune Epitope Database (IEDB) analysis resource. MHC class I and class II epitopes were selected based on predicted binding affinity and immunogenicity, while B-cell epitopes were identified to maximize humoral immune responses. Predicted epitopes were further screened for antigenicity using VaxiJen v2.0, allergenicity using AllerTOP v2.1, toxicity using ToxinPred2, and peptide solubility using the Innovagen peptide property calculator. Only epitopes predicted to be antigenic, non-allergenic, non-toxic, and highly soluble were considered for vaccine construction. Furthermore, the global distribution of HLA alleles associated with the selected T-cell epitopes was evaluated using the IEDB Population Coverage tool to estimate the worldwide applicability of the proposed vaccine candidate [20].

Multi-Epitope Vaccine Construction and Immunological Evaluation

The selected CTL, HTL, and B-cell epitopes were assembled into a single multi-epitope vaccine construct using suitable peptide linkers to preserve independent immunogenic activity and facilitate antigen processing. An appropriate immunostimulatory adjuvant was incorporated at the N-terminus to enhance the immune response. The final construct was subsequently evaluated for antigenicity (VaxiJen v2.0), allergenicity (AllerTOP v2.1), toxicity (ToxinPred2), solubility (SoluProt v1.0), and physicochemical properties using the ExPASy ProtParam server. Secondary structure prediction of the vaccine construct was performed using PSIPRED, whereas tertiary structure modeling was generated using SCRATCH Protein Predictor. The initial model was refined using the GalaxyRefine server to improve structural quality. Structural validation was carried out through the SAVES v6.1 server, including PROCHECK-based Ramachandran plot analysis to assess stereochemical quality and overall model reliability.

Molecular Docking and Molecular Dynamics Simulation

To investigate the interaction between the designed vaccine and the host innate immune receptor, the crystal structure of Toll-like receptor 3 (TLR3) was retrieved from the Protein Data Bank (PDB). Prior to docking, receptor preparation was performed using Discovery Studio Visualizer. Protein–protein docking was conducted using the

ClusPro server, and intermolecular interactions within the docked complex were analyzed using PDBsum. The structural stability and flexibility of the vaccine–TLR3 complex were further investigated using the iMODS server through Normal Mode Analysis (NMA). Parameters including deformability, B-factor, eigenvalue, covariance matrix, and elastic network were evaluated to estimate the dynamic behavior and stability of the docked complex.

Codon Optimization and In Silico Cloning

The final vaccine amino acid sequence was reverse-translated into a nucleotide sequence using the EMBOSS Backtranseq program. Codon optimization for efficient heterologous expression in Escherichia coli K-12 was performed using the Java Codon Adaptation Tool (JCat) [21], and the codon adaptation index (CAI) and GC content were analyzed to assess expression efficiency. The optimized nucleotide sequence was subsequently inserted into an appropriate expression vector using SnapGene software to simulate recombinant plasmid construction and evaluate the feasibility of in-silico cloning.

Results

3.1 Retrieval and Identification of Hemagglutinin (HA) Protein Sequence

The hemagglutinin (HA) protein sequence of Influenza A virus (A/Wilson-Smith/1933 H1N1) was retrieved from the UniProt/Swiss-Prot database (Accession: P03454). The reviewed HA0 precursor consists of 565 amino acids and is supported by experimental protein evidence, making it a reliable reference sequence for downstream Immunoinformatics analyses, including epitope prediction and multi-epitope vaccine design.

3.2 Physicochemical Characterization of Hemagglutinin (HA) Protein

The physicochemical properties of the HA protein were analyzed using the ExPASy ProtParam tool (Table 3.1). The protein comprises 565 amino acids with a molecular weight of 63,524.79 Da. The theoretical isoelectric point (pI) indicated the protein's charge characteristics under physiological conditions. The instability index (37.18) classified the protein as stable, while the aliphatic index (80.05) suggested moderate thermostability. The negative GRAVY value (−0.379) indicated a hydrophilic nature, supporting surface accessibility and antigenicity. The predicted half-life was 30 h in mammalian reticulocytes, >20 h in yeast, and >10 h in Escherichia coli, indicating good stability across different biological systems. Overall, these properties support the suitability of the HA protein for subsequent epitope prediction and multi-epitope vaccine design.

Table 1: Physicochemical Properties of Hemagglutinin (HA) Protein of Influenza A (H1N1)

Physicochemical Property	Hemagglutinin (HA)
Number of amino acids	565
Molecular weight	63,524.79 Da
Theoretical pI	6.55
Instability index	37.18 (Stable)
Aliphatic index	80.05
GRAVY	−0.379
Half-life (mammalian reticulocytes, in vitro)	30 hours
Half-life (yeast, in vivo)	>20 hours
Half-life (Escherichia coli, in vivo)	>10 hours

3.3 Disulfide Bond Prediction of Hemagglutinin (HA) Protein

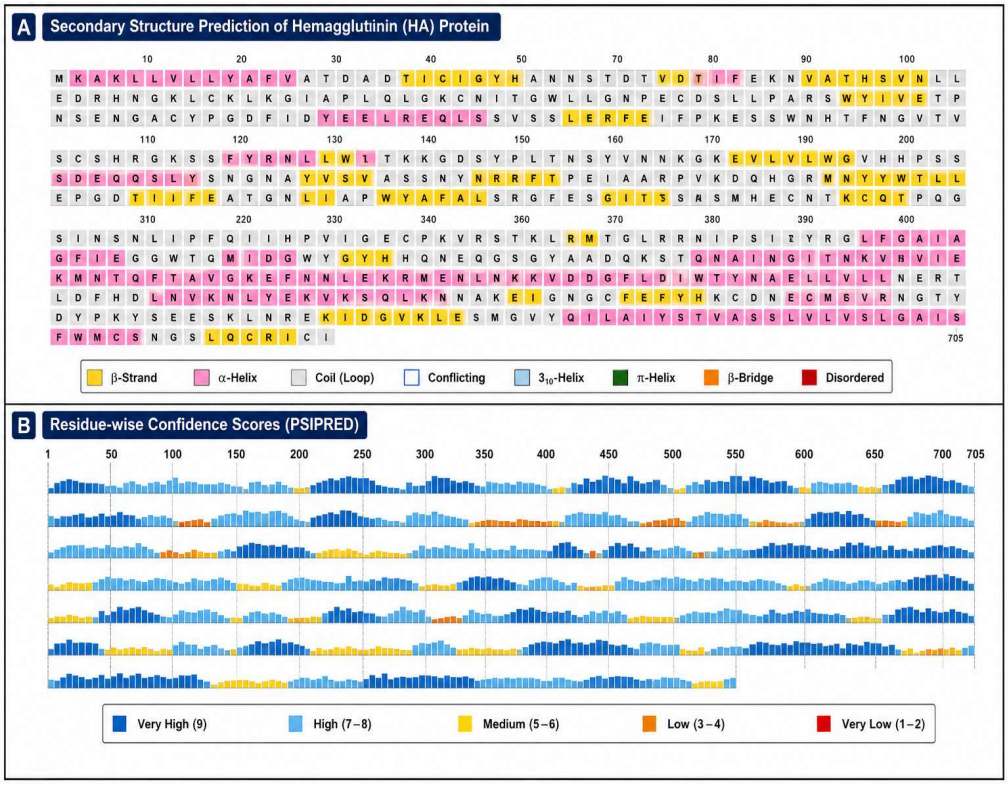
Disulfide bond prediction was performed to assess the structural stability of the hemagglutinin (HA) protein (Table 3.2). A total of six disulfide bonds were predicted, involving cysteine residues at positions 21, 59, 72, 84, 107, 152, 295, 319, 487, 491, 554, and 561. The highest-confidence disulfide bond was predicted between Cys295–Cys319, followed by Cys554–Cys561, Cys487–Cys491, Cys107–Cys152, Cys59–Cys72, and Cys21–Cys84. These predicted disulfide linkages indicate a structurally stable HA protein, supporting its proper folding and suitability for downstream structural analysis and Immunoinformatics-based vaccine design.

Table 2: Predicted Disulfide Bonds in Hemagglutinin (HA) Protein

Bond Index	Cysteine Position 1	Cysteine Position 2
1	295	319
2	554	561
3	487	491
4	107	152
5	59	72
6	21	84

3.4 Secondary Structure Prediction of Hemagglutinin (HA) Protein

The secondary structure of the HA protein was predicted using the PSIPRED server. The predicted structural elements are shown in Figure 3.1, and the corresponding confidence scores are presented in Figure 3.2. The HA protein exhibited a mixed secondary structure composed of α-helices, β-strands, and predominantly random coils. Random coils constituted the largest proportion of the sequence, indicating high conformational flexibility, whereas α-helices and β-strands were distributed throughout the protein. A prominent α-helical region was observed toward the C-terminal end, while β-strands occurred as short, discontinuous segments. The confidence scores were high for most α-helical and β-strand regions, indicating reliable predictions, whereas lower scores were observed in coil regions, reflecting their inherent flexibility.



3.5 Tertiary Structure Prediction of Hemagglutinin (HA) Protein

The three-dimensional structure of the HA protein was predicted using the Rosetta server. The predicted model is shown in Figure 3.3, while the residue-wise confidence scores (pLDDT) are presented in Figure 3.4. The predicted structure exhibited a compact fold comprising β-sheets, α-helices, and connecting loop regions. The β-sheet-rich core provided structural stability, whereas the loop regions contributed to conformational flexibility, a characteristic feature of viral surface glycoproteins involved in receptor binding and immune recognition. The model achieved an estimated TM-score of ~0.78, indicating a reliable overall fold. The pLDDT profile showed that most residues had confidence scores >80, particularly within the α-helical and β-sheet regions, while lower scores were mainly confined to flexible loop and terminal regions.

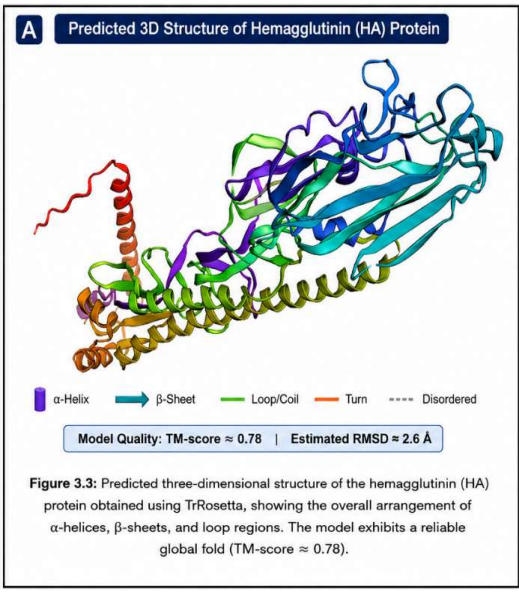


Figure 3.3: Predicted three-dimensional structure of the hemagglutinin (HA) protein obtained using TrRosetta, showing the overall arrangement of α-helices, β-sheets, and loop regions. The model exhibits a reliable global fold (TM-score ≈ 0.78).

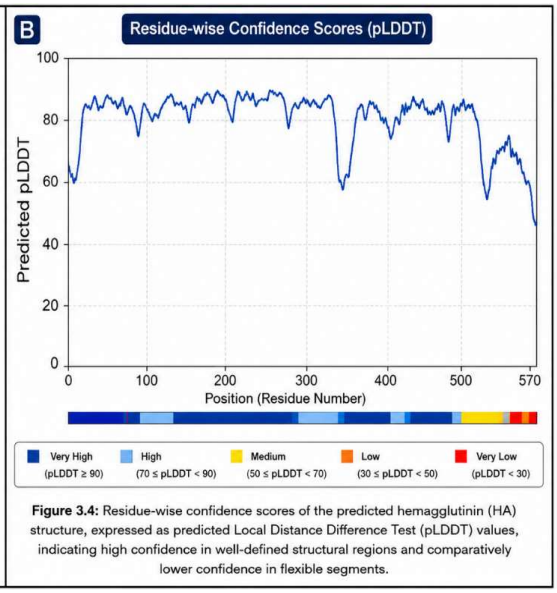


Figure 3.4: Residue-wise confidence scores of the predicted hemagglutinin (HA) structure, expressed as predicted Local Distance Difference Test (pLDDT) values, indicating high confidence in well-defined structural regions and comparatively lower confidence in flexible segments.

3.6 Antigenicity and Allergenicity Assessment of the HA Protein

The antigenicity of the hemagglutinin (HA) protein was evaluated using the VaxiJen v2.0 server with the virus model and a threshold of 0.4. The protein achieved an antigenicity score of 0.5422, classifying it as a probable antigen and indicating its potential to elicit an immune response. This finding is consistent with the established role of HA as the major surface glycoprotein and principal antigenic determinant of Influenza A virus. The allergenicity of the HA protein was predicted using AllerTOP v2.1, which classified the protein as a probable non-allergen, suggesting a low risk of inducing allergic reactions. Collectively, these results demonstrate that the HA protein possesses favorable immunological properties, supporting its suitability for epitope prediction and subsequent vaccine design.

3.7 T-Cell Epitope Prediction, Screening, and Immunological Evaluation

3.7.1 MHC Class I T-Cell Epitope Prediction and Screening

MHC class I (CTL) epitopes were predicted using the IEDB MHC-I binding prediction tool with the reference HLA allele set. A total of 30,051 peptide candidates were identified from the HA protein sequence. Peptides with a percentile rank ≤1 were selected as high-affinity binders, resulting in 256 unique epitopes after removal of redundant sequences. These epitopes were subsequently screened for antigenicity (VaxiJen v2.0), allergenicity (AllerTOP v2.1), toxicity (ToxinPred2), and water solubility (Innovagen peptide calculator). Based on the selection criteria of high antigenicity, non-allergenicity, non-toxicity, and good solubility, six CTL epitopes were identified as the most promising vaccine candidates: DQHGRMNYY, EKNVAVTHSV, KESSWPNHTF, KLTNSYVNNK, MESVRNGTY, and SHRGKSSFY. The immunological properties of the selected epitopes are summarized in Table 3.3.

Table 3: Selected MHC Class I CTL Epitopes with Immunological and Physicochemical Scores.

Predicted Epitope	Antigenicity (VaxiJen Score)	Allergenicity	Toxicity	Solubility
DQHGRMNYY	0.6524(Antigen)	Non-allergen	Non-toxin	Good
EKNVAVTHSV	0.6186(Antigen)	Non-allergen	Non-toxin	Good
KESSWPNHTF	0.9507(Anitgen)	Non-allergen	Non-toxin	Good
KLTNSYVNNK	0.5304(Antigen)	Non-allergen	Non-toxin	Good
MESVRNGTY	0.8219(Antigen)	Non-allergen	Non-toxin	Good
SHRGKSSFY	1.1764(Antigen)	Non-allergen	Non-toxin	Good

3.7.2 MHC Class II T-Cell Epitope Prediction and Screening

Helper T-lymphocyte (HTL) epitopes were predicted using the IEDB MHC class II binding prediction tool with the reference HLA allele set. A total of 14,877 peptide candidates were generated from the HA protein sequence. Pep-

tides with a percentile rank ≤10 were selected as high-affinity MHC class II binders, resulting in 90 unique epitopes after removal of redundant sequences. These epitopes were subsequently evaluated for antigenicity (VaxiJen v2.0), allergenicity (AllerTOP v2.1), toxicity (ToxinPred2), and water solubility (Innovagen peptide calculator). Following multi-parameter screening, four HTL epitopes met all selection criteria, including high antigenicity, non-allergenicity, non-toxicity, and good solubility. The selected epitopes were ASSNYNRRFTPEIAA, DTVDTIFEKNVAVTH, EQGSGYAADQKSTQN, and FTAVGKEFNNLEKRM. The physicochemical and immunological characteristics of the selected HTL epitopes are summarized in Table 3.4.

Table 4: Selected MHC Class II HTL Epitopes with Immunological and Physicochemical Scores

Predicted Epitope	Antigenicity VaxiJen Score)	Allergenicity	Toxicity	Solubility
ASSNYNRRFTPEIAA	0.7049(Antigen)	Non-allergen	Non-toxin	Good
DTVDTIFEKNVAVTH	0.4788(Antigen)	Non-allergen	Non-toxin	Good
EQGSGYAADQKSTQN	0.5466(Antigen)	Non-allergen	Non-toxin	Good
FTAVGKEFNNLEKRM	0.9447(Antigen)	Non-allergen	Non-toxin	Good

3.8 B-Cell Epitope Prediction and Screening

Linear B-cell epitopes of the hemagglutinin (HA) protein were predicted to identify potential antibody-binding regions. The HA protein sequence was analyzed using the IEDB linear B-cell epitope prediction tool, and the residue-wise epitope propensity profile is shown in Figure 3.5. A threshold score of 0.500 was applied, with residues scoring above the cutoff considered potential antigenic regions. The analysis identified 26 linear B-cell epitopes distributed across the HA sequence. All predicted epitopes were subsequently evaluated for antigenicity (VaxiJen v2.0), allergenicity (AllerTOP v2.1), toxicity (ToxinPred2), and water solubility (Innovagen peptide calculator). Following multi-parameter screening, 12 epitopes satisfied all selection criteria, including high antigenicity, non-allergenicity, non-toxicity, and good solubility, and were selected for vaccine construction.

3.9 Population Coverage of Selected T-Cell Epitopes

The global population coverage of the selected six MHC class I and four MHC class II epitopes was evaluated using the IEDB Population Coverage tool. The combined epitope set achieved an estimated global population coverage of 81.68%, with an average of 1.93 epitope–HLA hits per individual and a pc90 value of 0.55 (Figure 3.6), indicating broad HLA representation across different populations. The distribution of epitope–HLA recognition (Figure 3.7) showed that 18.32% of individuals were predicted not to recognize any selected epitope, whereas the majority recognized one to four epitope–HLA combinations. The cumulative population distribution (Figure 3.8) demonstrated a progressive decline in the proportion of individuals with increasing numbers of recognized epitope–HLA combinations, indicating that most individuals are likely to respond to at least one selected epitope.

3.10 Multi-Epitope Vaccine Construct Design

The multi-epitope vaccine was constructed by sequentially assembling the selected B-cell, MHC class I, and MHC class II epitopes into a single chimeric sequence. To enhance immunogenicity, human β-defensin-2 (UniProt accession: O15263) was incorporated as an N-terminal adjuvant. The adjuvant was linked to the epitope region using an EAAAK linker to ensure structural separation and minimize functional interference. Selected B-cell epitopes were arranged immediately downstream of the adjuvant, followed by six MHC class I epitopes linked with AAY linkers to facilitate antigen processing and presentation through the MHC class I pathway. The four MHC class II epitopes were connected using GPGPG linkers to promote efficient presentation through the MHC class II pathway. The final construct comprised the β-defensin-2 adjuvant, linker sequences, and the selected B-cell, MHC class I, and MHC class II epitopes in a continuous sequence for subsequent structural and immunological analyses.

3.11 Evaluation of the Multi-Epitope Vaccine Construct

The designed vaccine construct was evaluated for its immunological and physicochemical properties. VaxiJen v2.0 predicted the construct to be antigenic with a score of 0.7295, while AllerTOP v2.1 and ToxinPred2 classified it as non-allergenic and non-toxic, respectively. SoluProt v1.0 predicted a solubility score of 0.516, indicating acceptable recombinant protein solubility (Table 3.5). Physicochemical analysis using ExPASy ProtParam showed that the construct consisted of 235 amino acids with a molecular weight of 25,962.61 Da and a theoretical pI of 8.16. The instability

index (28.48) classified the construct as stable, while the aliphatic index (48.13) indicated moderate thermostability. The negative GRAVY value (−0.803) suggested a hydrophilic nature, favorable for solubility and molecular interactions. The predicted half-life was 30 h in mammalian reticulocytes, >20 h in yeast, and >10 h in *Escherichia coli*, indicating good stability across expression systems (Table 6).

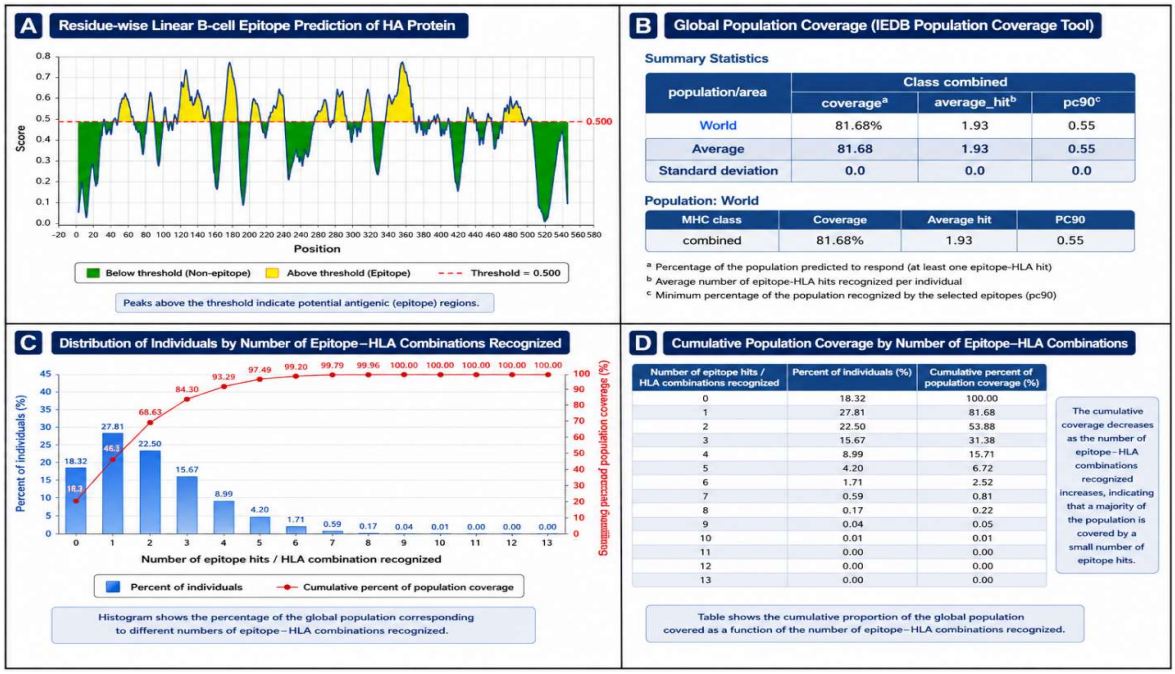


Table 5: Immunological Properties of the Designed Multi-Epitope Vaccine Construct

Property	Result
Antigenicity (VaxiJen v2.0)	0.7295 (Antigen)
Allergenicity (AllerTOP v2.1)	Non-allergen
Toxicity (ToxinPred2)	Non-toxic
Solubility (SoluProt 1.0)	0.516

Table 6: Physicochemical Properties of the Designed Multi-Epitope Vaccine Construct

Physicochemical Properties	Vaccine Construct
Number of amino acids	235
Molecular weight	25,962.61 Da
Theoretical pI	8.16
Instability index	28.48 (Stable)
Aliphatic index	48.13
GRAVY	-0.803
Half-life (mammalian reticulocytes, in vitro)	30 hours
Half-life (yeast, in vivo)	>20 hours
Half-life (Escherichia coli, in vivo)	>10 hours

3.12 Secondary Structure Analysis of the Multi-Epitope Vaccine Construct

The secondary structure of the multi-epitope vaccine construct was predicted using the PSIPRED server. The residue-wise secondary structure prediction is presented in Figure 3.9, while the detailed PSIPRED output is shown in Figure 3.10. The predicted structure was predominantly composed of coil (C) regions, interspersed with α-helices (H) and a limited number of β-strands (E). The abundance of coil regions indicates high conformational flexibility, whereas the α-helices and β-strands provide localized structural stability. This structural organization is characteristic of multi-epitope vaccine constructs and is favorable for epitope accessibility and immune recognition. The detailed PSIPRED output (Figure 3.10) includes the amino acid sequence (AA), predicted secondary structure (Pred), confidence scores (Conf), and graphical representation of the structural elements. High confidence scores were observed across most residues, indicating reliable secondary structure prediction, while lower confidence values were primarily associated with transitions between structural elements.

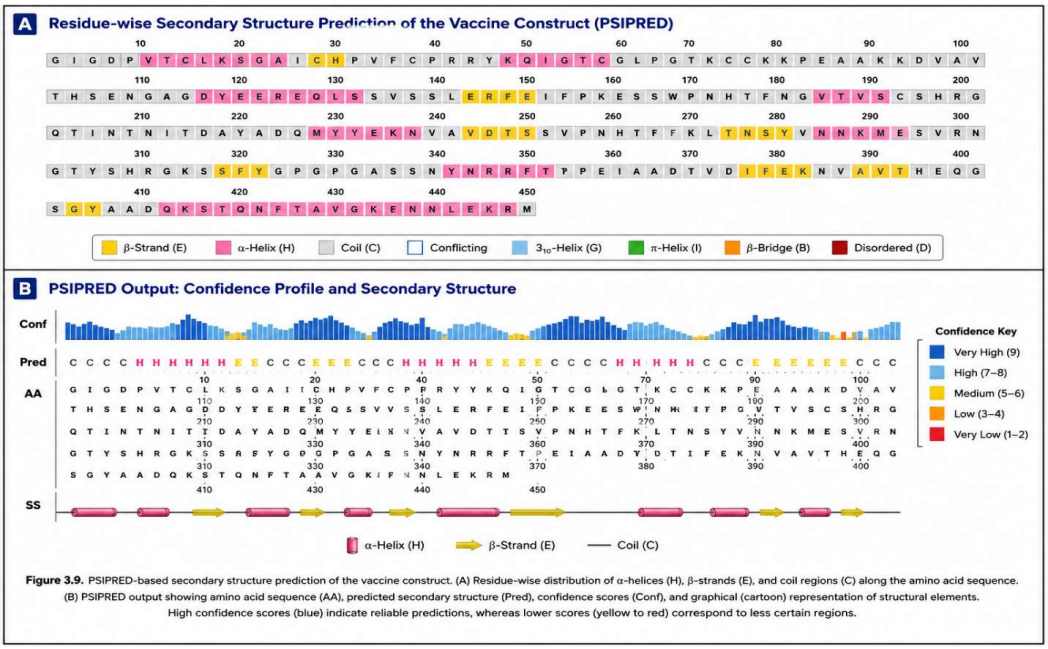


Figure 3.9. PSIPRED-based secondary structure prediction of the vaccine construct. (A) Residue-wise distribution of α-helices (H), β-strands (E), and coil regions (C) along the amino acid sequence. (B) PSIPRED output showing amino acid sequence (AA), predicted secondary structure (Pred), confidence scores (Conf), and graphical (cartoon) representation of structural elements. High confidence scores (blue) indicate reliable predictions, whereas lower scores (yellow to red) correspond to less certain regions.

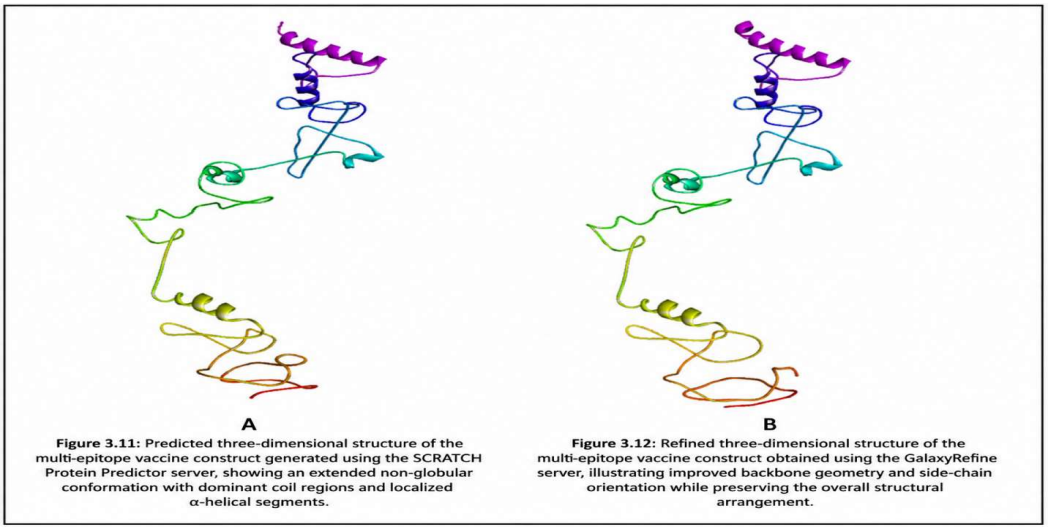


Figure 3.11: Predicted three-dimensional structure of the multi-epitope vaccine construct generated using the SCRATCH Protein Predictor server, showing an extended non-globular conformation with dominant coil regions and localized α-helical segments.

Figure 3.12: Refined three-dimensional structure of the multi-epitope vaccine construct obtained using the GalaxyRefine server, illustrating improved backbone geometry and side-chain orientation while preserving the overall structural arrangement.

3.13 Tertiary Structure Modeling and Refinement of the Multi-Epitope Vaccine Construct

The three-dimensional structure of the multi-epitope vaccine construct was predicted using the SCRATCH Protein Predictor server. The predicted model (Figure 3.11) exhibited an extended non-globular conformation dominated by coil regions with a few localized α-helical segments. The absence of extensive β-sheet structures and the elongated architecture are consistent with the presence of multiple epitopes connected by flexible linkers, resulting in increased

surface accessibility. The initial model was refined using the GalaxyRefine server to improve its structural quality. The refined structure (Figure 3.12) showed improved backbone geometry and side-chain conformations while preserving the overall structural arrangement. The refined model retained its extended architecture, making it suitable for downstream analyses, including structural validation, molecular docking, and immune simulation.

3.14 Structural Validation of the Refined Vaccine Construct Using Ramachandran Plot Analysis

The stereochemical quality of the refined vaccine construct was evaluated using the PROCHECK module of the SAVES v6.1 server. The Ramachandran plot (Figure 3.13) showed that, of 204 non-glycine and non-proline residues, 192 (94.1%) were located in the most favored regions, while 11 (5.4%) were in the additionally allowed regions. No residues were observed in the generously allowed regions, and only one residue (0.5%) was present in the disallowed region. The construct contained 19 glycine and 11 proline residues, which were evaluated separately due to their distinct conformational properties. Overall, the high percentage of residues in favored regions and the minimal number of outliers indicate that the refined model possesses good stereochemical quality and is suitable for subsequent molecular docking and other structural analyses.

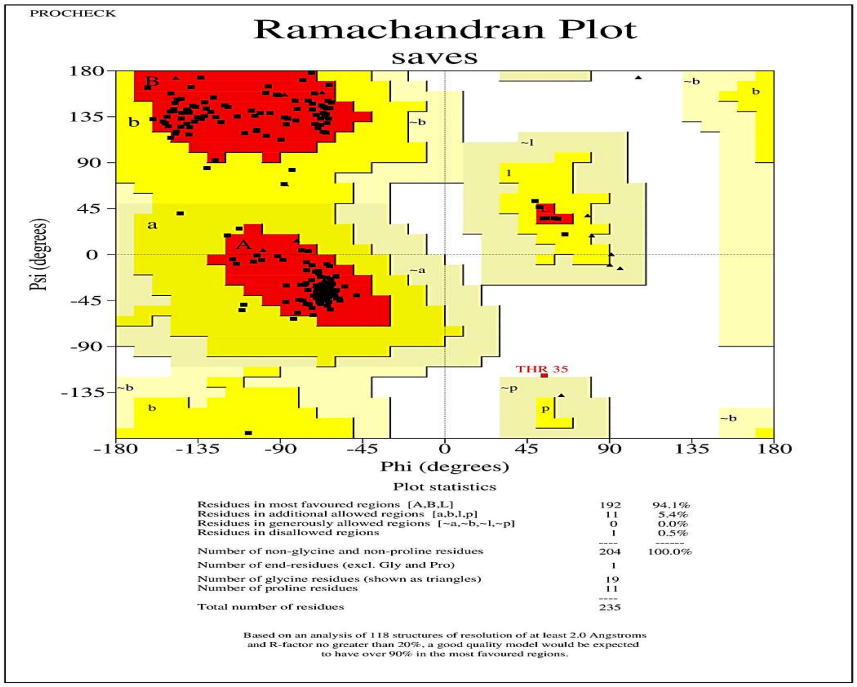


Figure 313: The PROCHECK server was used to generate a Ramachandran plot of a multi epitope vaccine construct with apparent success. Regions that are allowed (most favored regions A, B, L) are shown in red, additional allowed regions in dark yellow, generously allowed regions in light yellow and disallowed regions in white. Triangles are used for glycine residues. The plot reveals that this model is structurally reliable with a high percentage (94.1%) of residues in the most favored regions.

3.15 Molecular Docking of the Vaccine Construct with TLR3

The crystal structure of Toll-like receptor 3 (TLR3) (PDB ID: 2A0Z) was retrieved from the Protein Data Bank and prepared for docking by removing water molecules, heteroatoms, and bound ligands using Discovery Studio (Figure 3.14). Protein–protein docking between the multi-epitope vaccine construct and TLR3 was performed using the ClusPro server. Among the generated models, cluster 0 (67 members) was selected as the best docking solution, with a weighted score of −819.3 and a lowest energy score of −1002.4. The docked complex (Figure 3.15) showed that the vaccine construct binds to the concave surface of the TLR3 ectodomain, forming multiple interaction sites while maintaining its extended conformation. The favorable docking scores and binding orientation indicate a stable interaction, suggesting the potential of the vaccine construct to engage TLR3 and stimulate innate immune responses.

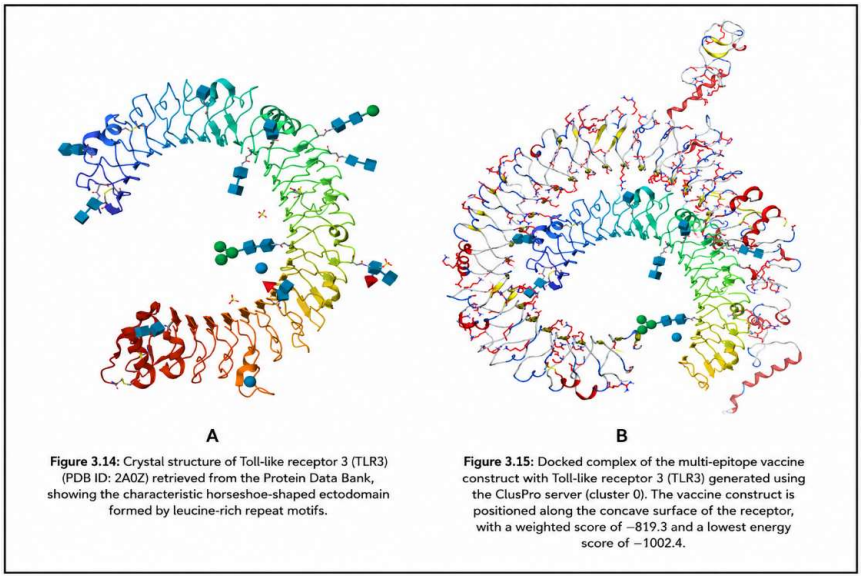


Figure 3.14: Crystal structure of Toll-like receptor 3 (TLR3) (PDB ID: 2A0Z) retrieved from the Protein Data Bank, showing the characteristic horseshoe-shaped ectodomain formed by leucine-rich repeat motifs.

Figure 3.15: Docked complex of the multi-epitope vaccine construct with Toll-like receptor 3 (TLR3) generated using the ClusPro server (cluster 0). The vaccine construct is positioned along the concave surface of the receptor, with a weighted score of −819.3 and a lowest energy score of −1002.4.

3.16 Protein–Protein Interaction Analysis of the Docked Complex

The docked TLR3–vaccine complex was analyzed using PDBsum to characterize the protein–protein interaction interface. The complex contained 26 interface residues on TLR3 (chain A) and 33 interface residues on the vaccine construct (chain B), with interface areas of 1555 Å² and 1503 Å², respectively. A total of 14 hydrogen bonds and 202 non-bonded contacts were identified at the binding interface, whereas no salt bridges or disulfide bonds were observed. These results indicate that the complex is primarily stabilized through hydrogen bonding and van der Waals interactions. Residue-level analysis (Figure 3.17) identified key interacting residues, including Asn597, Lys619, Ser673, Asn678, and His682 of TLR3, interacting with Glu76, Arg99, Tyr113, Asp114, and Thr129 of the vaccine construct. Overall, the extensive interaction interface and favorable intermolecular contacts support the formation of a stable TLR3–vaccine complex, consistent with the docking results.

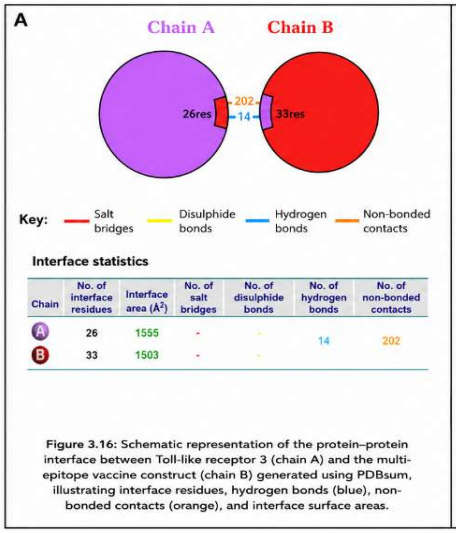


Figure 3.16: Schematic representation of the protein–protein interface between Toll-like receptor 3 (chain A) and the multi-epitope vaccine construct (chain B) generated using PDBsum, illustrating interface residues, hydrogen bonds (blue), non-bonded contacts (orange), and interface surface areas.

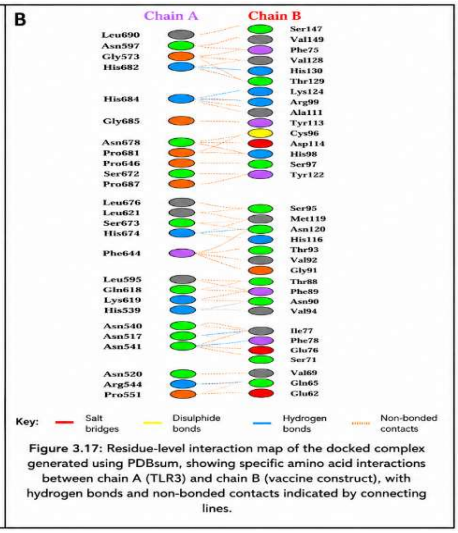


Figure 3.17: Residue-level interaction map of the docked complex generated using PDBsum, showing specific amino acid interactions between chain A (TLR3) and chain B (vaccine construct), with hydrogen bonds and non-bonded contacts indicated by connecting lines.

3.17 Normal Mode Analysis of the Vaccine–TLR3 Complex

The structural dynamics of the vaccine–TLR3 complex were evaluated using normal mode analysis (NMA) implemented in iMODS. The deformability profile (Figure 3.18) showed low flexibility across most residues, with only a few localized peaks corresponding to flexible loop or hinge regions. Similarly, the B-factor profile (Figure 3.19) indicated limited atomic fluctuations, supporting the overall rigidity of the complex. The first normal mode exhibited a low eigenvalue (2.575700×10^{-6}) (Figure 3.20), indicating that the complex can undergo collective conformational motions with minimal energy while maintaining structural integrity. The variance plot (Figure 3.21) demonstrated that the initial low-frequency modes accounted for most of the overall motion. The covariance matrix (Figure 3.22) revealed

coordinated residue motions, with correlated (red) and anti-correlated (blue) regions indicating concerted dynamic behavior. The elastic network model (Figure 3.23) showed a dense interaction network, reflecting strong residue connectivity and structural stability.

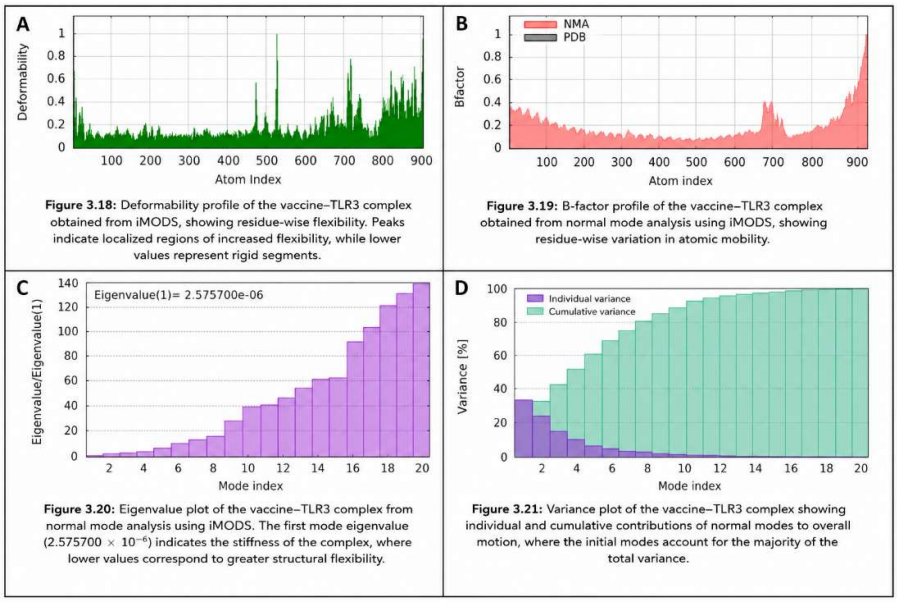


Figure 3.18: Deformability profile of the vaccine–TLR3 complex obtained from iMODS, showing residue-wise flexibility. Peaks indicate localized regions of increased flexibility, while lower values represent rigid segments.

Figure 3.19: B-factor profile of the vaccine–TLR3 complex obtained from normal mode analysis using iMODS, showing residue-wise variation in atomic mobility.

Figure 3.20: Eigenvalue plot of the vaccine–TLR3 complex from normal mode analysis using iMODS. The first mode eigenvalue (2.575700×10^{-6}) indicates the stiffness of the complex, where lower values correspond to greater structural flexibility.

Figure 3.21: Variance plot of the vaccine–TLR3 complex showing individual and cumulative contributions of normal modes to overall motion, where the initial modes account for the majority of the total variance.

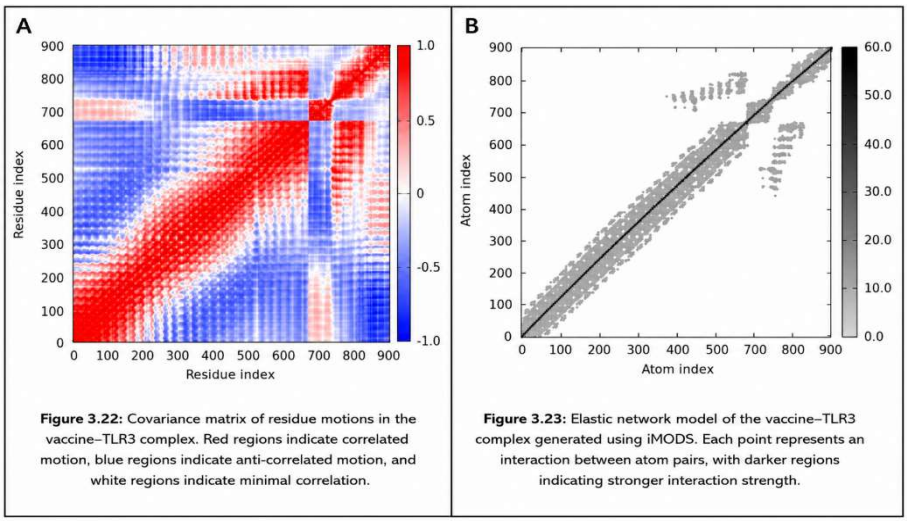


Figure 3.22: Covariance matrix of residue motions in the vaccine–TLR3 complex. Red regions indicate correlated motion, blue regions indicate anti-correlated motion, and white regions indicate minimal correlation.

Figure 3.23: Elastic network model of the vaccine–TLR3 complex generated using iMODS. Each point represents an interaction between atom pairs, with darker regions indicating stronger interaction strength.

3.18 Codon Optimization and In Silico Cloning

The amino acid sequence of the multi-epitope vaccine construct was reverse-translated into a nucleotide sequence using EMBOSS Backtranseq and subsequently optimized for expression in *Escherichia coli* K-12 using the JCat server. The optimized sequence achieved a codon adaptation index (CAI) of 0.97 and a GC content of 49.21%, indicating efficient translational potential and compatibility with the host expression system. Additionally, no internal EcoRI restriction sites were detected, facilitating downstream cloning procedures (Figure 3.24). To evaluate cloning feasibility, the optimized gene (~705 bp) was inserted *in silico* into the pET-28a(+) expression vector using the Eco53kI and BstZ17I restriction sites (Figures 3.25 and 3.26). The pET-28a(+) vector provides a T7 promoter for high-level expression, a 6×His tag for affinity purification, a lac operator for inducible expression, and a kanamycin resistance gene for selection. The recombinant construct showed correct gene insertion and orientation while preserving all essential vector elements, confirming its suitability for recombinant expression. Overall, the codon optimization and *in silico* cloning results support the experimental feasibility and efficient heterologous expression of the proposed vaccine construct in *E. coli*.

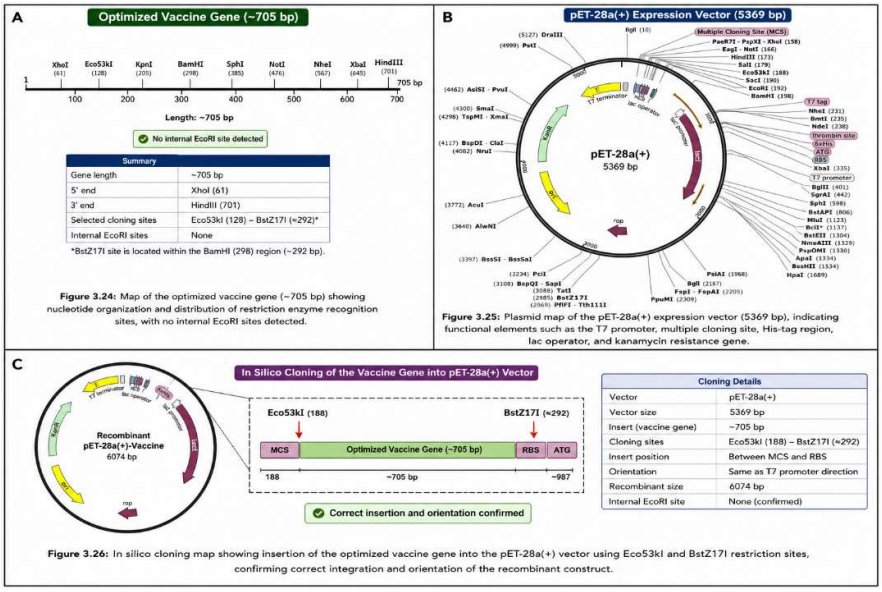


Figure 3.24: Map of the optimized vaccine gene (~705 bp) showing nucleotide organization and distribution of restriction enzyme recognition sites, with no internal EcoRI sites detected.

Figure 3.25: Plasmid map of the pET-28a(+) expression vector (5369 bp), indicating functional elements such as the T7 promoter, multiple cloning site, His-tag region, lac operator, and kanamycin resistance gene.

Figure 3.26: In silico cloning map showing insertion of the optimized vaccine gene into the pET-28a(+) vector using Eco53kI and BstZ17I restriction sites, confirming correct integration and orientation of the recombinant construct.

Discussion

The HA glycoprotein of Influenza A virus is a well-known immunogenic molecule, essential for viral entry and is an important target for rational vaccine design. Analyses have been facilitated by having a high-fidelity sequence base for the reviewed A/Wilson-Smith/1933 H1N1 sequence from UniProtKB, as is increasingly being used in recent Immunoinformatics workflows to greatly improve the reliability of subsequent epitope predictions and structural modeling [60]. The physicochemical characterization showed the instability index to be 37.18 and the GRAVY value was −0.379, which meant that the protein architecture is stable and hydrophilic. These parameters have been correlated with characteristics of viable recombinant vaccine antigens in the literature over the past five years, and a negative GRAVY score and stability index below 40 have consistently been correlated with increased solubility as well as structural integrity and suitability for heterologous expression systems [61]. Moreover, the conserved disulfide connectivity suggested by the prediction of 6 conserved disulfide bonds confirms the robustness of the HA structure; recently, the importance of retaining the disulfide connectivity in antigen design for preserving conformational epitopes and ensuring antigen structure is correct has been shown to correlate with better docking stability and immune recognition [22].

The secondary and tertiary structure analysis showed the composition of α-helices, β-sheets, and extensive coil regions to be roughly equal, while the coil regions were mostly found in surface-exposed loops. Recent epitope mapping studies have demonstrated that flexible coil regions are indeed common binding sites of B-cell epitopes and are easily accessible to B-cell receptors and the proteases involved in their processing [23] and therefore immunologically beneficial. The high confidence tertiary model was validated by good TM-scores and pLDDT values and was used as a reliable framework to identify epitopes. Six MHC class I, four MHC class II, and twelve B-cell epitopes were obtained from the screening pipeline employed here, which included filtering epitopes based on MHC binding affinity, antigenicity, non-allergenicity, non-toxicity, and solubility. The stringent filtration process used reflects the protocols used in current literature, in which sequential testing of the immunological and physicochemical properties was shown to reduce the number of false positives, and to prioritize the epitopes that have the highest probability of producing a safe and strong immune response [24]. Selected epitopes showed high antigenicity scores, and favorable safety features, in line with previous multi-epitope designs, which have reported comparable screening results for influenza and other viral pathogens.

Analysis of the population coverage of the selected epitope repertoire estimated 81.68% global coverage. This coverage is like high-performing vaccine candidates from around the world which were reported over the last five years, generally achieve >75% coverage when including epitope targeting to prevalent HLA alleles, thus minimizing the risk of vaccine failure due to genetic diversity [25]. The distribution of epitope hits further indicated that most people would be able to recognize more than one epitope, a property known to be linked to the durability of immune responses and ability to deal with viral escape variants. The vaccine construct was designed using human β-defensin-2

as an N-terminal adjuvating agent to enhance immunogenicity of the vaccine. Recent *in-silico* and experimental studies have shown that β-defensin-2 is a functional endogenous adjuvant capable of recruiting dendritic cells to enhance adaptive immunity via Toll-like receptor, which is why its incorporation is supported [65]. Moreover, strategic incorporation of EAAAK, AAY and GPGPG linkers ensures the lack of junctional neo-epitopes and promotes efficient processing by the proteasomal system, which has been demonstrated using recent multi-epitope constructs where the epitope presentation and immunogenicity are enhanced [26].

The refined vaccine model was validated structurally using the Ramachandran plot analysis, which showed that more than 94% of the residues were in favorable regions thus demonstrating good stereochemical quality. The docking simulation to Toll-like receptor 3 (TLR3) resulted in a stable complex with a weighted score of −819.3 and good number of interface contacts that were supported by hydrogen bond and non-bonded interactions. These energy and interface similarities to recent docking studies involving vaccine-receptors and TLRs, which also showed good binding to and activation of innate immune signaling pathways, suggest that the complexes formed here will have good stability and activation of innate immune signaling pathways [22]. Additionally, normal mode analysis revealed low deformability and a stable eigenvalue for the vaccine-receptor complex, suggesting that the vaccine-receptor complex would retain its structure even in the presence of dynamic fluctuations, a result that was also confirmed with recent dynamic simulations of vaccine complexes [27] where restricted flexibility and low eigenvalues were found to correlate with complex stability and effective receptor engagement. Last, the codon optimization used for Escherichia coli K-12 yielded a codon adaptation index (CAI) of 0.97 and a GC content of 49.21%, which are in the optimum range for recombinant protein production in E. coli K-12. The recent cloning simulations and expression experiments show that the high CAI value (close to 1.0) and GC contents (40% to 60%) are predictive of efficient transcription and translation of the designed construct, proving the feasibility of the experiments [21]. The similarities in immunological, structural, and expression properties of the designed multi-epitope vaccine with the criteria defined in the recent literature indicate that it may be a good candidate for pursuing further preclinical studies for vaccine development against Influenza A H1N1.

Conclusion

A multi-epitope vaccine candidate against Influenza A (H1N1) HA protein was successfully designed using an Immunoinformatics-based approach. The vaccine exhibited favorable antigenicity, stability, non-allergenicity, non-toxicity, broad global population coverage (81.68%), reliable structural quality, stable interaction with TLR3, and efficient expression potential in Escherichia coli. These findings support the proposed vaccine as a promising candidate for further in vitro and in vivo validation to confirm its immunogenicity, safety, and protective efficacy.

Future Perspective

Although the proposed multi-epitope vaccine demonstrated promising computational results, experimental validation is essential. The predicted epitopes should be evaluated through in vitro studies to confirm their immunogenicity, antigen expression, and safety, followed by in vivo studies to assess protective efficacy and immune responses. Further optimization of adjuvants, epitope arrangement, and linker sequences may improve vaccine performance. Additionally, population-specific coverage analyses and comprehensive toxicological and stability studies are required before advancing the vaccine candidate toward clinical development.

References

1. Hay, B. Lomniczi, A. Bellamy, J. Skehel, *Transcription of the influenza virus genome*, *Virology*. 355 (1977) 337–355.
2. Dias, D. Bouvier, T. Crépin, A.A. McCarthy, D.J. Hart, F. Baudin, et al., *The cap snatching endonuclease of influenza virus polymerase resides in the PA subunit*., Nature. 458 (2009) 914–8.
3. Portela, P. Digard, *The influenza virus nucleoprotein: a multifunctional RNA binding protein pivotal to virus replication*., J. Gen. Virol. 83 (2002) 723–34.
4. M. Chua, S. Schmid, J. Perez, *Influenza A virus utilizes suboptimal splicing to coordinate the timing of infection*, Cell Rep. 3 (2013) 23–29.
5. Mohr, *Phosphorylation and dephosphorylation events that regulate viral mRNA translation*, Virus Res. 119 (2006) 89–99.
6. W.J. Wurzer, O. Planz, C. Ehrhardt, M. Giner, T. Silberzahn, S. Pleschka, et al., *Caspase 3 activation is essential for efficient influenza virus propagation*., EMBO J. 22 (2003) 2717–28.
7. S. V Avilov, D. Moisy, N. Naffakh, S. Cusack, *Influenza A virus progeny vRNP trafficking in live infected cells studied with the virus-encoded fluorescently tagged PB2 protein*., Vaccine. 30 (2012) 7411–7.
8. A.J. Eisfeld, E. Kawakami, T. Watanabe, G. Neumann, Y. Kawaoka, *RAB11A is essential for transport of the influenza virus genome to the plasma membrane*., J. Virol. 85 (2011) 6117–26.
9. M. Veit, B. Thaa, *Association of influenza virus proteins with membrane rafts*., Adv. Virol. 2011 (2011) 370606.
10. P. Gómez-Puertas, C. Albo, E. Pérez-Pastrana, A. Vivo, A. Portela, *Influenza virus matrix protein is the major driving force in virus budding*., J. Virol. 74 (2000) 11538–47.
11. R. Ruigrok, F. Baudin, I. Petit, W. Weissenhorn, *Role of influenza virus M1 protein in the viral budding process*, Int. Congr. Ser. 1219 (2001) 397–404.
12. M. Hilsch, B. Goldenbogen, C. Sieben, C.T. Höfer, J.P. Rabe, E. Klipp, et al., *Influenza A Matrix Protein M1 Multimerizes upon Binding to Lipid Membranes*., Biophys. J. 107 (2014) 912–23.
13. Y. Liang, Y. Hong, T. Parslow, *cis-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments*, J. Virol. 79 (2005) 10348–10355.
14. E. Fournier, V. Moules, B. Essere, J.-C. Paillart, J.-D. Sirbat, C. Isel, et al., *A supramolecular assembly formed by influenza A virus genomic RNA segments*., Nucleic Acids Res. 40 (2012) 2197–209.
15. R.W. Ruigrok, A. Aitken, L.J. Calder, S.R. Martin, J.J. Skehel, S.A. Wharton, et al., *Studies on the structure of the influenza virus haemagglutinin at the pH of membrane fusion*., J. Gen. Virol. 69 (Pt 11 (1988) 2785–95.
16. C.S. Copeland, R.W. Doms, E.M. Bolzau, R.G. Webster, A. Helenius, *Assembly of influenza hemagglutinin trimers and its role in intracellular transport*., J. Cell Biol. 103 (1986) 1179–91.
17. C.S. Copeland, K.P. Zimmer, K.R. Wagner, G.A. Healey, I. Mellman, A. Helenius, *Folding, trimerization, and transport are sequential events in the biogenesis of influenza virus hemagglutinin*., Cell. 53 (1988) 197–209.
18. M.S. Segal, J.M. Bye, J.F. Sambrook, M.J. Gething, *Disulfide bond formation during the folding of influenza virus hemagglutinin*., J. Cell Biol. 118 (1992) 227–44.
19. H. Klenk, R. Rott, M. Orlich, J. Blödorn, *Activation of influenza A viruses by trypsin treatment, Virology*. 439 (1975) 426–439.
20. J.M. Nicholls, M.C.W. Chan, W.Y. Chan, H.K. Wong, C.Y. Cheung, D.L.W. Kwong, et al., *Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract*., Nat. Med. 13 (2007) 147–9.
21. J.M. Nicholls, A.J. Bourne, H. Chen, Y. Guan, J.S.M. Peiris, *Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses*., Respir. Res. 8 (2007) 73.
22. Pappas, K. Viswanathan, A. Chandrasekaran, R. Raman, J.M. Katz, R. Sasisekharan, et al., *Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957*., PLoS One. 5 (2010) e11158.
23. M. Imai, T. Watanabe, M. Hatta, S.C. Das, M. Ozawa, K. Shinya, et al., *Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets*., Nature. 486 (2012) 420–8.
24. UniProt Consortium. UniProt: *The universal protein knowledgebase in 2023. Nucleic Acids Res*. 2023 Jan 6;51(D1):D523-D531.
25. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. *Database resources of the National Center for Biotechnology Information*. Nucleic Acids Res. 2023 Jan 6;51(D1):D29-D38.
26. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. *Protein identification and analysis tools on the ExPASy server*. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa (NJ): Humana Press; 2005. p. 571-607.
27. Buchan DWA, Jones DT. The PSIPRED *Protein Analysis Workbench*: 20 years on. Nucleic Acids Res. 2019 Jul 2;47(W1):W402-W407.