Syrian Arab Republic
Ministry of Higher Education and Scientific Research
Syrian Virtual University (SVU)
Bioinformatics Master Program

IN-SILICO DESIGN OF MULTI-EPITOPE VACCINE FOR H5N1 VIRUS USING IMMUNOINFORMATIC TOOLS

تصميم حاسوبي للقاح متعدد الحواتم ضد فيروس انفلونزا الطيور باستخدام أدوات المعلوماتية المناعية

A Thesis Submitted as a Fulfillment of Requirement for a Master's Degree in Bioinformatics

Submitted by:

Hesham ALKHATEEB

Uni ID: 168269

Supervised by:

Dr. Yanal ALKUDSY

2024

Table of Contents

List of Tables

List of Figures

List of Abbreviation

HA	Hemagglutinin
NA	Neuraminidase
CTL	Cytotoxic T Lymphocyte
HTL	helper T lymphocytes
HLA	Human Leukocyte Antigen
MHC	Major Histocompatibility Complex
TLR	Toll-Like Receptor
IFN-y	Interferon Gamma
IL	Interleukin
SVM	Support Vector Machine
GRAVY	Grand Average of Hydropathicity
CAI	Codon Adaptation Index
IEDB	Immune Epitope Database
VC	Vaccine Construct
RV	Reverse vaccinology
PDB	Protein Data Bank

Abstract

The highly pathogenic avian influenza virus (H5N1) remains a persistent threat to global public health and poultry industries due to its zoonotic potential, rapid mutation rates, and antigenic variability. Traditional vaccine development strategies are often limited by time constraints, high costs, and inadequate cross-strain protection. In this study, an immunoinformatics-driven approach was employed to design a novel multi-epitope vaccine against H5N1. Viral protein sequences for Hemagglutinin (HA) and Neuraminidase (NA) were retrieved from the UniProt database and subjected to comprehensive epitope prediction and screening for T-cell (MHC-I and MHC-II) and B-cell (linear and discontinuous) epitopes. The selected epitopes were filtered based on their antigenicity, non-allergenicity, non-toxicity, and cytokine-inducing potential.

A multi-epitope vaccine construct was designed by assembling these epitopes with appropriate adjuvants (MDA5, H9E) and linkers (AAY, GPGPG, KK) to ensure immunogenicity and structural stability. Physicochemical analysis predicted a stable and hydrophilic vaccine construct, suitable for bacterial expression systems. Tertiary structure modeling using RoseTTAFold and subsequent validation (Ramachandran plot, ERRAT, ProSA) confirmed the reliability and stability of the protein model. Molecular docking studies revealed strong binding interactions with Toll-like receptors TLR7 and TLR8, indicating potential immune activation pathways. Immune simulation analysis predicted robust humoral and cellular immune responses, with increased IgG, IgM, T-helper cell activity, and IFN-γ cytokine production. Codon optimization and in-silico cloning into the pET-26b(+) vector confirmed efficient expression compatibility with *Escherichia coli* systems.

The results suggest that the designed multi-epitope vaccine construct holds significant potential to elicit broad-spectrum immunity against H5N1 and overcome limitations associated with traditional vaccines. Future studies involving in-vitro and in-vivo validation are essential to confirm the vaccine's immunogenicity, safety, and efficacy.

Keywords:

Avian Influenza Vaccine, Bird Flu, H5N1 Virus, Reverse Vaccinology, Immunoinformatics, Computational-Driven Vaccinology, Epitope-Based Vaccine, Multi-Epitope Vaccine.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, **Dr. Yanal Alkudsy**, for his invaluable guidance, insightful feedback, and unwavering support throughout this research journey.

I am profoundly thankful to my wife and family for their endless patience, understanding, and constant encouragement. Thank you all for being a part of this academic milestone.

1. Introduction

1.1 Background of Avian Influenza

Avian influenza, commonly known as bird flu, is an infectious disease caused by influenza A viruses that primarily infect avian species, but it can also cross species barriers to infect humans and other mammals[1]. Among avian influenza viruses, highly pathogenic strains such as H5N1 and H7N9 are of particular concern due to their severe implications for both public health and the global poultry industry[2]. These strains have been responsible for high mortality rates in poultry and sporadic, severe infections in humans, often leading to respiratory complications and death[3]. The World Health Organization categorizes certain avian influenza viruses as having pandemic potential because of their zoonotic capability and genetic diversity, making them a persistent global health threat[4].

Transmission among birds primarily occurs through contact with contaminated surfaces, bodily secretions, or infected individuals, whereas human infection generally arises through direct exposure to infected birds, their droppings, or contaminated environments[5]. The zoonotic potential of these viruses is rooted in their high genetic flexibility; avian influenza viruses undergo rapid mutations and genetic reassortment, which occasionally enable them to infect humans and, in rare cases, spread between humans[6]. The H5N1 strain, which caused human infections in Hong Kong in 1997, demonstrated the virus's ability to cause severe disease, driving global surveillance and containment efforts[7].

Controlling avian influenza is challenging due to the virus's rapid evolution, which allows it to evade immune defenses and develop antiviral resistance[8]. Traditional vaccines against avian influenza face significant limitations, as they need to be reformulated and produced to match emergent strains[2]. Consequently, alternative vaccine development approaches, such as immunoinformatics, are gaining momentum. Immunoinformatics enables rapid vaccine design by identifying conserved epitopes across multiple viral strains, facilitating the creation of multi-epitope vaccines that could offer broad, cross-strain immunity[9].

1.2. Importance of Combating Bird Flu

Combating avian influenza, especially highly pathogenic strains like H5N1, is critical for public health and economic stability. Bird flu not only threatens the poultry industry by causing devastating losses in poultry populations but also poses significant risks to human health due to its zoonotic potential and ability to cause severe, often fatal respiratory illness in humans[6]. Since its first major outbreak in humans in 1997, H5N1 has caused hundreds of cases with high mortality rates, leading to concerns about the virus’s potential to mutate into a form easily transmissible among humans, thus sparking a potential pandemic[10].

The economic consequences of avian influenza outbreaks are substantial, impacting poultry production and trade, food security, and local economies. Governments often resort to culling infected or exposed bird populations, resulting in billions of dollars in losses, particularly in countries where poultry farming is a primary livelihood[11]. According to the Food and Agriculture Organization (FAO), the direct costs of culling, coupled with trade restrictions and decreased consumer demand, lead to extended economic repercussions that strain public health resources and hinder global food supply[12].

From a public health perspective, avian influenza viruses are a significant concern because of their genetic variability and adaptability. High mutation and reassortment rates in these viruses enable them to escape immune responses and develop drug resistance, which complicates containment and treatment efforts[13]. Given the limitations of traditional vaccines in addressing rapidly evolving strains, there is an urgent need for innovative vaccine design approaches that can provide cross-protection against diverse avian influenza strains. Addressing avian influenza is thus critical not only to protect human and animal health but also to safeguard economic stability and global food security[9].

1.3. Limitations of Current Vaccines

Current vaccines against avian influenza, typically inactivated or live-attenuated vaccines, present significant limitations in controlling rapidly mutating strains like H5N1. Traditional vaccines require exact matching to circulating viral strains to be effective, a challenge compounded by the high mutation rate of influenza A viruses, which frequently generate antigenic drift and shift[14]. This necessitates frequent updates to the vaccine composition, leading to time delays and substantial costs in vaccine development and production[15]. Additionally, while vaccines may reduce mortality in poultry, they are less effective at preventing infection or viral shedding, allowing continued transmission among flocks and increasing the risk of cross-species spillover[16].

Human vaccines for avian influenza remain largely experimental, with limited availability and efficacy. Most are developed to target seasonal influenza and lack the broad protective coverage needed to guard against zoonotic avian strains[17]. The reliance on egg-based production methods further slows vaccine availability, as this method takes months to yield sufficient doses and may introduce mutations that compromise vaccine efficacy[18]. These challenges highlight an urgent need for alternative vaccine platforms that can accommodate the genetic variability of avian influenza viruses and provide cross-strain immunity.

Emerging strategies, such as in silico vaccine design, offer promising alternatives by utilizing computational immunoinformatics tools to predict and model conserved epitopes across multiple avian influenza strains[9]. By identifying these conserved regions, multi-epitope

vaccines can be constructed to stimulate immune responses that target common antigens among diverse virus strains, potentially offering broader protection[19]. This approach not only accelerates the design process but also reduces reliance on traditional vaccine production, paving the way for scalable and flexible vaccine solutions against avian influenza.

1.4. Rationale for In Silico Vaccine Design

Traditional vaccine development, while effective, is often hindered by lengthy timelines and substantial costs, particularly in the context of fast-evolving pathogens like avian influenza viruses [20] . Computational tools in vaccine design offer a transformative approach by significantly shortening this process. Unlike conventional methods that rely on laboratory-based pathogen cultivation and testing, in silico approaches utilize vast genomic and proteomic data to screen potential vaccine targets rapidly. By leveraging advanced algorithms, researchers can predict antigenic regions and identify potential epitopes within a fraction of the time required for experimental methods[9]. This acceleration is especially valuable in responding to emergent strains, enabling the development of vaccine candidates that target conserved viral components across multiple strains, thus enhancing preparedness against pandemics[21].

The use of computational tools not only expedites the identification of immunogenic targets but also allows for rapid evaluation of critical parameters such as antigenicity, toxicity, and population coverage. This capability reduces the need for initial wet-lab experiments, thus conserving resources and enabling researchers to prioritize only the most promising candidates for further testing[22].

1.5. Overview of Immunoinformatics

Immunoinformatics, a specialized field within bioinformatics, applies computational techniques to immunology, enabling the virtual exploration and design of vaccines. This approach integrates data from genomics, proteomics, and immunology to model host-pathogen interactions and predict epitopes that can trigger robust immune responses[23]. One of the primary goals of immunoinformatics is to identify epitopes that are likely to be recognized by the immune system's B and T cells. Through tools like the Immune Epitope Database (IEDB) and NetMHC, immunoinformatics platforms can predict epitope binding affinities to MHC molecules, a key step in assessing immunogenic potential[24].

In the case of avian influenza, where viral proteins undergo frequent mutation, immunoinformatics can identify conserved regions that remain stable across strains. Designing multi-epitope vaccines based on these conserved elements allows for the development of broadly protective vaccines, potentially offering immunity against various

viral subtypes[21]. By combining immunoinformatics with other in silico techniques, researchers can model the structural stability and efficacy of vaccine constructs, as well as predict immune responses, paving the way for a faster and more efficient vaccine design pipeline.

1.6. Objective of the Study

This study aims to design a novel multi-epitope vaccine for avian influenza H5N1 using immunoinformatic tools to enhance the speed and precision of vaccine development. Specifically, the study seeks to identify conserved B-cell and T-cell epitopes from highly immunogenic viral proteins, such as Hemagglutinin and Neuraminidase, that can elicit robust immune responses. By leveraging in silico methods, we will construct a vaccine candidate with the potential for broad cross-strain protection against multiple avian influenza subtypes. The objectives are structured as follows:

1. Epitope Prediction and Selection: To use immunoinformatics platforms, including the Immune Epitope Database, to predict and select conserved epitopes based on their binding affinity, immunogenicity, and antigenicity.
2. Vaccine Construct Design: To assemble the selected epitopes into a multi-epitope vaccine construct with appropriate linkers and adjuvants, ensuring structural stability and immunogenic potential.
3. In Silico Evaluation of Vaccine Properties: To assess the construct's antigenicity, allergenicity, and toxicity using computational tools to ensure the vaccine's safety and efficacy.
4. Structural Modeling and Immune Simulation: To model the three-dimensional structure of the vaccine construct and simulate immune responses, thereby predicting its effectiveness in generating an immune response across diverse populations.
5. Codon Optimization and Cloning: To optimize the codon sequence for expression in a suitable host system, preparing the vaccine candidate for future experimental validation.

By achieving these objectives, this study aims to establish a proof of concept for using immunoinformatics in developing safe, effective, and scalable vaccines against avian influenza, potentially contributing to rapid responses against emerging influenza threats.

2. Methods

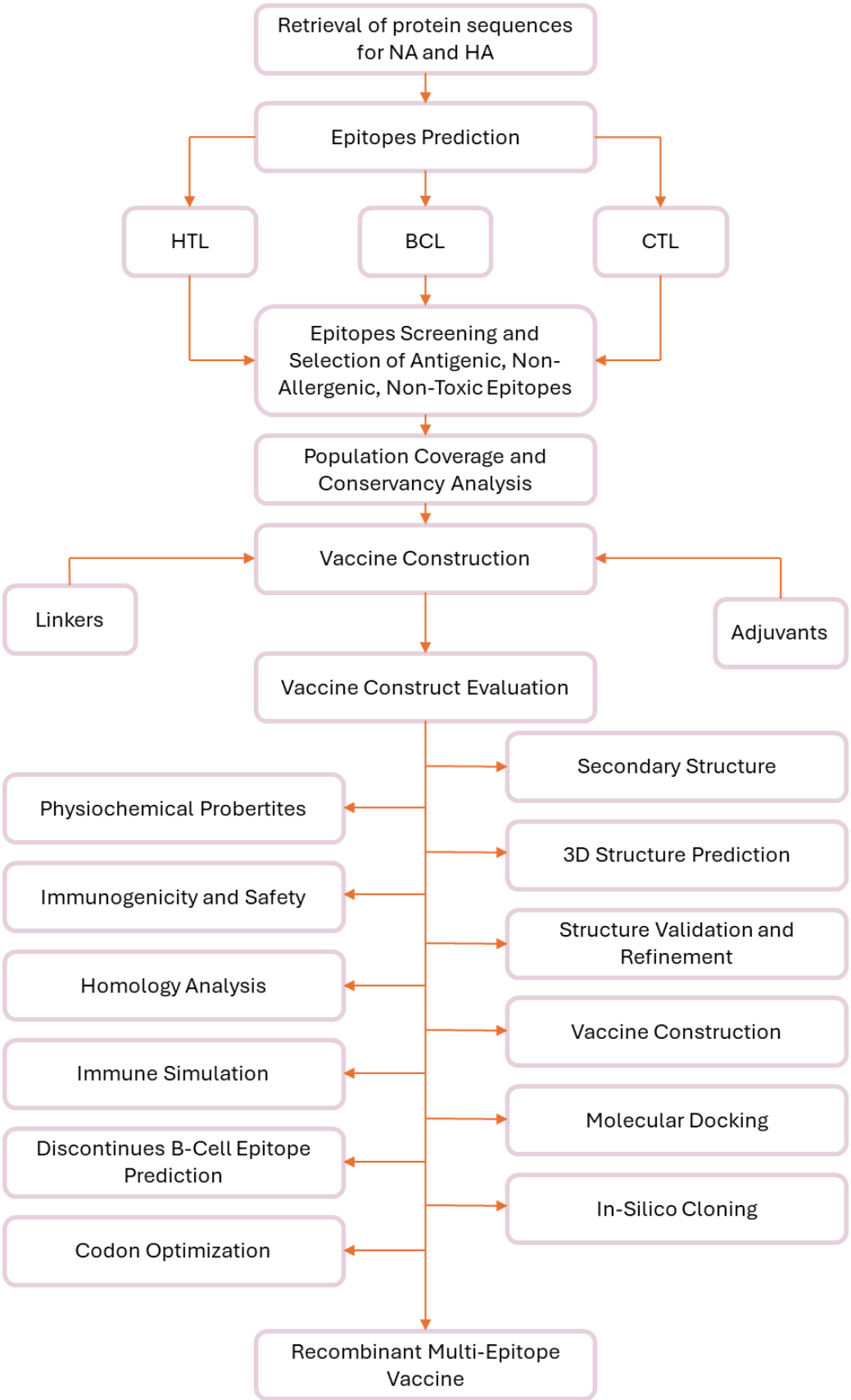
2.1. Data Collection and Selection

For this study, the viral protein sequences for hemagglutinin (HA) and neuraminidase (NA) from the H5N1 avian influenza A virus were retrieved from the UniProt database[25]. These proteins were chosen due to their critical roles in viral entry and release, as well as their immunogenic potential, making them key targets for vaccine development. The sequence retrieval process was carried out using the following keywords: "Hemagglutinin and Neuraminidase H5N1 avian Influenza A virus". To ensure data quality, the search results were filtered by 1) Reviewed Entries, only manually curated entries classified as reviewed (Swiss-Prot) were included to prioritize high-quality annotations. 2) Completeness, sequences annotated as complete were selected to ensure full-length proteins for comprehensive analysis. 3) Relevance, results were limited to proteins specific to the H5N1 strain of avian influenza A virus.

The antigenicity of the selected HA and NA protein sequences was evaluated using VaxiJen[26] computational tools. Protein sequences were uploaded in FASTA format. VaxiJen[a] is an alignment-free tool that predicts antigenicity based on the physicochemical properties of proteins, with a threshold of ≥0.4 indicating antigenicity. The results were later analyzed to confirm that all selected proteins exhibited strong antigenic potential.

[a] https://www.ddg-pharmfac.net/vaxijen/

Figure 1 Immunoinformatics Approaches Used for Vaccine Design Against H5N1



2.2. Epitope Prediction

The HLA system, located on chromosome 6, plays a pivotal role in the adaptive immune system by encoding molecules that present antigenic peptides to T cells. The HLA system is divided into two main classes: 1). HLA Class I: Includes HLA-A, HLA-B, and HLA-C molecules, which present intracellular peptides to CTLs. 2). HLA Class II: Includes HLA-DR, HLA-DQ, and HLA-DP molecules, which present extracellular peptides to HTLs[27].

2.2.1. MHC-I epitope prediction

For this study, the prediction of T-cell epitopes covered wide demographic groups by choosing the most prevalent HLA alleles in the global population to ensure optimal applicability. Full HLA class I alleles reference set have been selected, which are the following: HLA-A*01:01, HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*03:01, HLA-A*11:01, HLA-A*23:01, HLA-A*24:02, HLA-A*26:01, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:02, HLA-B*07:02, HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*57:01, HLA-B*58:01, HLA-C*04:01, HLA-C*06:02, HLA-C*07:01, HLA-C*12:02.

Epitope prediction for HLA class I molecules was conducted using the NetMHCpan_el algorithm[28], which identifies peptides with strong binding affinities to the selected alleles. This method provides binding affinity scores and percentile ranks, enabling the selection of high-confidence epitopes based on their potential to elicit robust immune responses. Identified epitopes were further filtered based on their antigenicity and safety, ensuring the selection of non-toxic and non-allergenic candidates[29]. The IC50 values for predicted peptides were not explicitly calculated. However, based on the very low percentile ranks ($\leq$2%) and high binding scores obtained from NetMHCpan_el, the selected peptides are strongly indicative of high-affinity binders, with estimated IC50 values of $\leq$50 nM.

2.2.2. MHC-II epitope prediction

To identify potential HTL epitopes, predictions were implemented to the full HLA class II alleles reference set globally. The full reference set were HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*12:01, HLA-DRB1*13:02, HLA-DRB1*15:01. Selecting these alleles ensured broader population coverage and relevance within wider demographic groups.

Epitope prediction was conducted using the NetMHCIIpan algorithm[30], an advanced tool for pan-allelic MHC-II binding affinity prediction. This algorithm identifies peptides capable of binding HLA-DR molecules based on their binding affinity, providing percentile ranks and

predicted IC50 values to indicate binding strength. The process included the following steps: 1). Input Data: Protein sequences for HA and NA were submitted in FASTA format. 2). Selection of Alleles: Predictions were restricted to the full HLA reference set. 3). Prediction Parameters: Peptides of 15 amino acids were generated and analyzed for binding to the selected alleles. 4). Binding affinity thresholds were applied: Peptides with score > 0.9 were classified as high-affinity binders. Percentile ranks ≤ 2% were used to prioritize strong binders. 5). Data Filtering: Predicted epitopes were further screened for antigenicity and safety using additional computational tools to ensure their potential as vaccine candidates. The predicted MHC-II epitopes represent peptides with strong binding affinities to HLA-DR molecules, indicating their capability to stimulate helper T-cell responses effectively.

2.2.3. B-cell Epitopes prediction - Linear

The prediction of linear B-cell epitopes was conducted using the ABCpred server[31], a reliable immunoinformatics tool for identifying linear B-cell epitopes. The amino acid sequences of the vaccine candidate proteins HA and NA were submitted to the server for 16-mer epitope prediction. A threshold of 0.51 was applied to filter epitopes with higher probabilities of inducing B-cell-mediated immune responses. The ABCpred server[b] employs trained recurrent neural networks to predict and rank epitopes. Peptides with higher scores indicate a higher likelihood of functioning as effective epitopes.

[b] https://webs.iiitd.edu.in/raghava/abcpred/index.html

2.3. Epitope Screening and Characterization

To ensure the immunological suitability of the predicted epitopes, a series of screening and characterization steps were performed. The screening workflow included predictions of antigenicity, allergenicity, toxicity, and cytokine inducibility, utilizing various computational tools as outlined below:

1. Antigenicity Prediction: All epitope sequences (MHC-I, MHC-II, and Linear B-cell) were submitted to the VaxiJen v2.0 server to assess their antigenic potential. A threshold of 0.5 was applied to classify epitopes as antigenic. VaxiJen is an alignment-independent tool based on the physicochemical properties of proteins, providing accurate predictions of antigenicity with a reported accuracy exceeding 70%[26.]
2. Allergenicity Assessment: The AllerTOP v2.0 server[c] was used to identify and exclude allergenic epitopes. This tool employs an alignment-free approach, analyzing key physicochemical attributes of proteins to predict their allergenic potential with approximately 94% sensitivity[32].
3. Toxicity Prediction: ToxinPred[d] was employed to evaluate the toxicity of the selected epitopes. The support vector machine (SVM) approach was utilized with default parameters to differentiate between toxic and non-toxic peptides. Only non-toxic epitopes were retained for further analysis[33].
4. Class I Immunogenicity: MHC-I epitopes were assessed using the Class I Immunogenicity tool available on the IEDB Analysis Resource[e]. This tool predicts the ability of peptides to elicit immune responses by binding to MHC-I molecules and activating CTLs[34].
5. IFN-γ Inducibility: MHC-II epitopes were submitted to the IFN-γ Epitope Server[f], which employs an SVM-based method to predict interferon-gamma-inducing peptides. IFN-γ is a critical cytokine regulating adaptive immune responses[35].
6. IL-4 Inducibility: IL4pred[g], an in-silico platform for designing and predicting interleukin-4-inducing peptides, was used to assess the ability of MHC-II epitopes to stimulate IL-4 production[36].

[c] https://www.ddg-pharmfac.net/allertop_test/

[d] https://webs.iiitd.edu.in/raghava/toxinpred/protein.php

[e] http://tools.iedb.org/immunogenicity/

[f] https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php

[g] https://webs.iiitd.edu.in/raghava/il4pred/predict.php

2.4. Population Coverage Calculation:

The Population Coverage Calculation was performed using the IEDB Population Coverage Tool[h]. This tool assesses the predicted population coverage of selected T-cell epitopes based on their binding to HLA alleles. The full HLA reference set was selected to include the most frequent MHC class I (CTL epitopes) and MHC class II (HTL epitopes) alleles globally. The analysis was conducted for the "World" region to assess global coverage, ensuring a comprehensive evaluation.

The aim was to predict the percentage of the global population that can recognize the selected epitopes based on HLA allele distribution. This calculation also includes: 1). Average hits: The mean number of epitope-HLA combinations recognized by individuals in the population. 2). PC90: The minimum number of epitope-HLA combinations recognized by 90% of the population. The selected CTL and HTL epitopes were entered into the Population Coverage Tool. Separate analyses were conducted for MHC Class I, MHC Class II, and their combined coverage.

2.5. Epitope Conservancy Analysis

The degree of conservancy for the selected epitopes of HA and NA proteins was analyzed using the IEDB Conservancy Analysis Tool[i, 37]. This tool allows for determining the level of sequence conservancy of epitopes across multiple related protein sequences. Conservancy was evaluated against the sequences obtained from UniProt to ensure broad strain coverage. The sequence identity threshold was set to determine conservation, ranging from 0% (non-conserved) to 100% (fully conserved). Peptide lengths and exact sequence matching were considered, allowing for the identification of conserved regions across strains or related sequences. Conservancy was calculated as the percentage of sequences in the given database that matched each epitope at a specified identity threshold. Results provided insight into the degree of epitope conservation across various strains, identifying which epitopes were highly conserved. Highly conserved epitopes are crucial in vaccine design as they are less prone to mutation and are likely to elicit immune responses across a broader range of viral strains, facilitating the identification of suitable candidates for vaccine development. Conversely, strain-specific epitopes may have limited utility but could be effective for targeting specific variants.

[hh] http://tools.iedb.org/population/

[i] http://tools.iedb.org/conservancy/

2.6. Construction of the Multi-Epitope Vaccine:

The multi-epitope vaccine was designed to overcome the inherent non-immunogenicity of short peptides, which, when administered as standalone vaccines, fail to elicit strong immune responses [38] . To address this limitation, a potent immune-stimulatory adjuvant was incorporated at the N- and C-terminal of the construct to activate both the innate and adaptive immune systems, ensuring robust immunogenicity. Additionally, linkers were utilized to simulate the vaccine's natural antigen processing, enhance epitope separation, and optimize immune responses[39].

The vaccine construct consisted of carefully screened B-cell, CTL, and HTL epitopes, which were assembled in a sequential manner. Specific linkers were chosen to optimize the immunological processing and presentation of these epitopes, shown in Figure 2. AAY linker was employed to separate CTL epitopes. The AAY sequence facilitates efficient proteasomal cleavage and subsequent presentation by MHC class I molecules, ensuring a strong cytotoxic T-cell response[39]. GPGPG linker was used to connect HTL epitopes. This flexible sequence prevents the formation of junctional epitopes and maintains structural integrity, promoting proper presentation by MHC class II molecules for helper T-cell activation[38]. KK linker was applied between B-cell epitopes to stabilize the peptide structure and enhance the accessibility of linear B-cell epitopes for antibody recognition[40].

To further boost immunogenicity, the adjuvants MDA5 and H9E were incorporated at the N-terminus of the vaccine construct, linked using a rigid EAAAK linker. The MDA5 adjuvant, a cytoplasmic pattern recognition receptor, was selected for its ability to activate innate immunity by recognizing viral RNA and stimulating type I interferons, which are critical for antiviral defense[31]. The H9E epitope, derived from the influenza hemagglutinin protein, was included to stimulate specific humoral and T-helper cell responses, particularly targeting avian influenza viruses[38]. The EAAAK linker was employed to spatially separate the adjuvants from the epitopes, maintaining their functional independence and ensuring proper structural conformation of the vaccine construct[39].

The final vaccine construct was subjected to comprehensive in silico analysis to validate its antigenicity and non-allergenicity, ensuring its suitability as a safe and effective immunogen for further development.

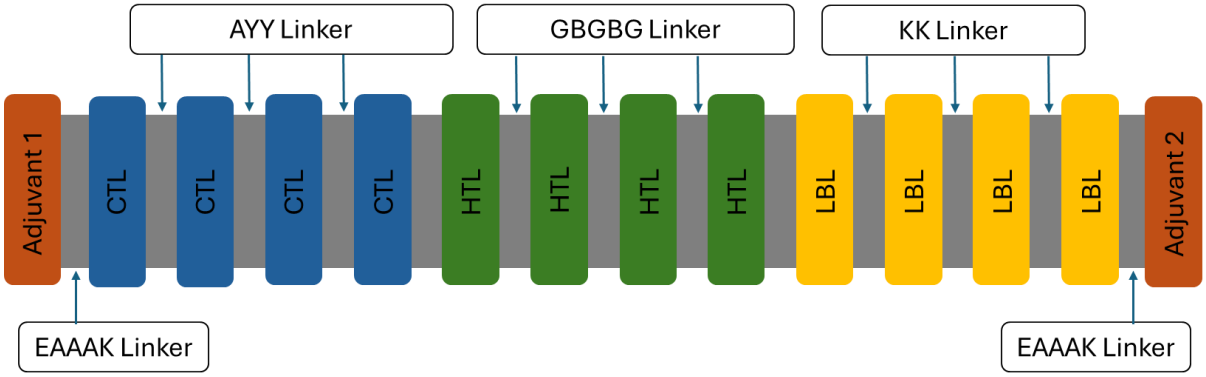


Figure 2 The structural arrangement of B and T cell epitopes along with linkers and adjuvant for the final

2.7. Specifications Prediction of the Vaccine Construct:

2.7.1. Physicochemical Property Analysis

The physical and chemical properties of the vaccine construct were analyzed using the ProtParam tool [j, 41]. This analysis included assessments of amino acid composition, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, extinction coefficient, atomic composition, and the grand average of hydropathicity (GRAVY). The instability index, an essential metric for protein stability, was calculated, with protein candidates exhibiting an instability index greater than 40 being excluded as unstable. This selection criterion aimed to identify stable vaccine candidates suitable for subsequent experimental studies[41].

2.7.2. Secondary Structure Prediction

The secondary structural attributes of the vaccine construct, such as α-helices, β-strands, and random coils, were predicted using SOPMA[k], and NetSurfP-2.0[l]. SOPMA predicts secondary structure based on probability matrices derived from amino acid sequences[42], while NetSurfP-2.0 incorporates sequence-based learning to predict secondary structures and surface accessibility, enhancing the understanding of the vaccine construct's structural properties[43].

2.7.3. Solubility Prediction

The SOLpro server[m] was utilized to forecast the solubility of the vaccine construct when expressed in *Escherichia coli*. SOLpro employs a two-stage support vector machine (SVM) architecture, where classifiers in the initial layer assess sequence features, and the final classifier integrates the outputs to predict solubility with associated probabilities[44]. Additional solubility predictions were performed using Protein-Sol, a web-based suite for theoretical solubility calculations, to corroborate the findings[45].

2.7.4. Antigenicity, Allergenicity, and Toxicity Assessments

The antigenicity of the vaccine construct was assessed using the VaxiJen server (threshold set at 0.5), which predicts antigenicity independent of sequence alignment, facilitating the identification of potential vaccine candidates[26]. Allergenicity was evaluated using AllerTOP v.

[j] https://web.expasy.org/protparam/

[k] https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html

[l] https://services.healthtech.dtu.dk/services/NetSurfP-2.0/

[m] https://scratch.proteomics.ics.uci.edu/

2.0, which predicts allergens based on physicochemical properties and machine learning techniques[46]. The toxicity of selected sequences was analyzed using ToxDL[n], a deep learning-based approach for in silico prediction of protein toxicity from sequence alone.

2.7.5. Homology Analysis

To ensure minimal cross-reactivity with human proteins and reduce the risk of autoimmune reactions, the vaccine construct was subjected to homology analysis against the Homo sapiens proteome. BLASTp (Basic Local Alignment Search Tool for proteins, blast.ncbi.nlm.nih.gov) was employed for this purpose, and sequences with significant homology were excluded to enhance the vaccine's specificity to the target pathogen[47].

2.8. Tertiary Structure Prediction and Validation

The tertiary structure prediction and validation of the designed multi-epitope vaccine were carried out through a systematic workflow involving computational tools to ensure accurate structural modeling, refinement, and validation.

2.8.1. 3D Structure Prediction

The 3D structure of the multi-epitope vaccine was predicted using a combination of PHYRE2 (Protein Homology/Analogy Recognition Engine V2.0) server and RoseTTAFold, leveraging the strengths of both tools to enhance model accuracy and reliability.

PHYRE2[o] employs advanced remote homology detection techniques by combining profile-profile alignment and fold recognition algorithms to generate accurate structural models based on homologous protein templates from PDB[48]. It also predicts ligand-binding sites and evaluates the functional and structural implications of amino acid substitutions within the modeled protein sequence. The submitted protein sequence was analyzed using the intensive modeling mode of PHYRE2, which provides a refined 3D model optimized for structural integrity.

In parallel, RoseTTAFold[p], a deep-learning-based protein structure prediction tool, was employed to complement the PHYRE2-derived model. RoseTTAFold integrates end-to-end three-track neural network architecture to simultaneously consider protein sequence, distance constraints, and structural features for accurate structure modeling[49]. This approach significantly improves structural prediction for proteins with limited homologous templates

[n] http://www.csbio.sjtu.edu.cn/bioinf/ToxDL/

[o] https://www.sbg.bio.ic.ac.uk/phyre2/

[p] https://robetta.bakerlab.org/

in the PDB database. The outputs from both servers were carefully compared and evaluated to ensure structural consistency and reliability, and the superior model was selected for subsequent refinement and validation.

2.8.2. Model Validation

The structural model obtained from RoseTTAFold was subjected to validation using the SAVES v6.1 server[q], which integrates multiple structural evaluation tools, including PROCHECK and ERRAT, to assess the stereochemical quality and structural reliability of protein models. PROCHECK server, which generates a Ramachandran plot to assess the stereochemical quality of the protein model. The Ramachandran plot illustrates the distribution of the backbone dihedral angles (phi (ϕ) and psi (ψ)) of amino acid residues, categorizing them into favored, allowed, and outlier regions[50]. The percentage of residues in these regions was analyzed to evaluate the quality of the predicted tertiary structure and to identify potential irregularities or structural deviations.

Additionally, ERRAT was used to calculate the Overall Quality Factor, which reflects the model's structural reliability based on non-bonded atomic interactions. Higher ERRAT scores indicate better structural quality, with values above 90% generally considered excellent.

2.8.3. Model Refinement

To enhance the stability and structural accuracy of the predicted model, GalaxyRefine[r] was employed. This tool applies iterative structure perturbation and subsequent relaxation through molecular dynamics simulation, resulting in improved structural geometry and side-chain packing[51]. Both pre- and post-refinement models were analyzed to ensure significant structural improvements. Following refinement, the protein structure underwent validation using the ProSA[s] (Protein Structure Analysis), which is a widely accepted tool for identifying potential errors in 3D protein models by calculating the Z-score of the structure[52]. The Z-score indicates the overall quality of the protein structure by comparing its total energy deviation with a database of native protein conformations. Z-scores that fall outside the range characteristic of experimentally derived protein structures suggest potential structural anomalies.

[q] https://saves.mbi.ucla.edu/

[r] https://galaxy.seoklab.org/

[s] https://prosa.services.came.sbg.ac.at/prosa.php

2.9. Defining Discontinuous B-cell Epitopes (Conformational)

Conformational B-cell epitopes, essential for eliciting a humoral immune response, were identified from the refined 3D protein vaccine model using the Ellipro server[t]. Ellipro predicts discontinuous epitopes based on the protein's tertiary structure by combining geometric properties and protrusion indices to rank potential epitopic regions, enabling the selection of immunologically relevant conformational epitopes[53].

2.10. Immune Response Simulation

The C-ImmSim server[u] was employed to simulate the immune response of the multi-epitope vaccine construct. C-ImmSim models immune interactions across three anatomical compartments: the bone marrow, where B-cell production occurs; the thymus, responsible for T-cell maturation; and the lymph node, where antigen presentation and immune activation take place. The simulation integrates immunological parameters, including antigen concentration, cytokine release, and immune cell proliferation, to predict the vaccine's immunogenic profile over time (Rapin et al., 2010).

[t] http://tools.iedb.org/ellipro/

[u] https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php

2.11. Molecular Docking with TLR7&8

2.11.1. Rationale for Docking with TLR7 and TLR8

TLRs are critical components of the innate immune system, responsible for recognizing pathogen-associated molecular patterns (PAMPs) and initiating downstream signaling cascades that result in immune activation. Among these, TLR7 and TLR8 are intracellular receptors located primarily in endosomal compartments and are known to detect single-stranded RNA (ssRNA) molecules. Their activation triggers the production of pro-inflammatory cytokines and type I interferons, which are essential for orchestrating antiviral responses[54].

In the context of vaccine design, targeting TLR7 and TLR8 can enhance the immunogenicity of the vaccine construct, thereby improving the activation of both innate and adaptive immunity. Therefore, molecular docking was performed between the designed vaccine construct and TLR7 and TLR8 to predict and evaluate their potential interaction patterns, binding affinity, and structural stability[55].

2.11.2. Structure Retrieval of TLR7 and TLR8

The 3D structures of TLR7 and TLR8 were retrieved from the PDB as they provide experimentally resolved high-resolution structures. TLR7 (PDB ID: 7CYN, CryoEM of TLR7 complexed with ligand), TLR8 (PDB ID: 3W3G, Crystal structure of TLR8 unliganded form). The structures were downloaded in PDB format, and any non-essential ligands, water molecules, and ions were removed using Chimera[v] to prepare the receptors for docking.

The 3D structure of the vaccine construct was previously modeled and refined, the top-ranked model was selected based on structural validation parameters.

2.11.3. Docking Software Selection

Molecular docking was performed using HDOCK[w], a widely recognized docking tool for protein-protein interactions. HDOCK uses a hybrid algorithm that combines template-based modeling with ab initio docking, making it highly suitable for predicting receptor-ligand interactions.

---

[v] https://www.cgl.ucsf.edu/chimera/

[w] http://hdock.phys.hust.edu.cn/

2.11.4. Docking Procedure

The prepared TLR7 and TLR8 receptor structures and the vaccine construct were uploaded to the HDOCK server. TLR7 and TLR8 were designated as receptors, while the vaccine construct was set as the ligand. Default parameters were selected, with a focus on global docking followed by local optimization. The docking process generated multiple poses ranked based on binding energy scores.

2.11.5 Docking Results Analysis

The docking results were analyzed based on the binding energy scores provided by HDOCK, where lower energy scores indicated more stable interactions. The top-ranked docking poses for TLR7-vaccine construct and TLR8-vaccine construct complexes were visualized and analyzed using PyMOL[x]. Key interaction features such as hydrogen bonds, and salt bridges, at the binding interface were identified and highlighted.

2.12. Codon Optimization and In Silico Cloning

The gene sequence encoding the recombinant vaccine construct was optimized using the GeneScript[y] codon optimization tool to ensure efficient expression in the chosen host, Escherichia coli K12. Since the expression host differs from the natural host of the influenza virus, codon optimization was essential to address discrepancies in codon usage preference between these organisms. The optimization process adjusted the gene sequence based on the codon usage table for E. coli K12, enhancing translational efficiency and reducing the likelihood of translational errors or ribosomal stalling.

Two primary metrics were considered during the codon optimization process. First, Codon Adaptation Index (CAI), which measures the similarity of codon usage between the optimized gene and the preferred codons of the expression host. A CAI value close to 1.0 indicates optimal adaptation, predicting higher levels of protein expression in the selected host[56]. Second, GC Content was adjusted to fall within the 30–70% range, an ideal window that ensures both efficient transcription and stable mRNA secondary structure, reducing the risk of translation inhibition[57].

To facilitate subsequent in-silico cloning, restriction enzyme sites for NdeI and XhoI were manually introduced at the 5’ and 3’ ends of the optimized gene sequence, respectively. These

[x] https://pymol.org/

[y] https://www.genscript.com/

restriction sites are compatible with the multiple cloning site (MCS) of the pET-26b(+) vector, enabling seamless directional cloning.

The optimized gene sequence was cloned into the pET-26b(+) vector using SnapGene software[z]. The vector and gene insert were digested with NdeI and XhoI restriction enzymes, followed by ligation to ensure correct orientation and preservation of the open reading frame (ORF). The final construct was validated in silico by analyzing restriction digestion patterns and confirming sequence integrity, ensuring suitability for downstream expression and purification experiments.

[z] https://www.snapgene.com/

3. Results

3.1. Chosen Dataset

By applying the criteria mentioned in the methods, the search results were narrowed down to 11 proteins that met the inclusion standards. These proteins are detailed in Table 1, including their UniProt accession numbers, and associated metadata.

Table 1 Chosen proteins from UniProt with their references and accession number

Accession Nr	Protein Names	Organism	PubMed ID
Q6DPL2	**Neuraminidase**	Influenza A virus (strain A/Vietnam/1203/2004 H5N1)	15241415 16318689 15681421 16915235 18480754
Q710U6	**Neuraminidase**	Influenza A virus (strain A/Chicken/Scotland/1959 H5N1)	16439620 15567494 16192481 15744059
Q809V0	**Neuraminidase**	Influenza A virus (strain A/Silky Chicken/Hong Kong/SF189/2001 H5N1 genotype A)	12077307
Q809V2	**Neuraminidase**	Influenza A virus (strain A/Chicken/Hong Kong/FY150/2001 H5N1 genotype D)	12077307
Q809U7	**Neuraminidase**	Influenza A virus (strain A/Chicken/Hong Kong/715.5/2001 H5N1 genotype E)	12077307
Q9W7Y7	**Neuraminidase**	Influenza A virus (strain A/Hong Kong/156/1997 H5N1 genotype Gs/Gd)	9482438 9430591 9658115 15567494 16192481 15744059
P09345	**Hemagglutinin**	Influenza A virus (strain A/Chicken/Scotland/1959 H5N1)	3375087 16439620
Q8QPL1	**Hemagglutinin**	Influenza A virus (strain A/Duck/Hong Kong/2986.1/2000 H5N1 genotype C)	11878904

O56140	**Hemagglutinin**	Influenza A virus (strain A/Hong Kong/156/1997 H5N1 genotype Gs/Gd)	9482438 9658115 9430591
Q9Q0U6	**Hemagglutinin**	Influenza A virus (strain A/Goose/Guangdong/1/1996 H5N1 genotype Gs/Gd)	10484749
O89746	**Hemagglutinin**	Influenza A virus (strain A/Chicken/Hong Kong/220/1997 H5N1 genotype Gs/Gd)	9658115

The antigenicity of all selected proteins was evaluated using ANTIGENpro and VaxiJen, and all proteins scored above 0.92 in ANTIGENpro and above 0.48 in VaxiJen, indicating strong antigenic potential as antigenic candidates for vaccine design.

3.2. Epitope Prediction

3.2.1. MHC-I Epitope Prediction

High-affinity MHC-I binding peptides were identified for both HA and NA proteins using the NetMHCpan_el prediction method. For hemagglutinin and neuraminidase proteins, many peptides were identified with high binding scores >0.9 across multiple HLA alleles, as summarized in Table 2 and 3. These peptides exhibited strong binding affinities and low percentile ranks (≤ 0.05), indicating their potential to elicit robust CTL responses. Notably, several peptides were conserved across multiple protein sequences, underscoring their relevance as cross-strain vaccine targets. The antigenicity and toxicity of all selected epitopes was evaluated using VaxiJen and ToxinPred, and all results shown in Tables 2&3.

Table 2 High-Affinity MHC-I Binding Peptides Predicted for Hemagglutinin Proteins with their Antigenicity and Toxicity prediction. Seq Num refers to the corresponding HA proteins: (Seq 1: O56140, Seq 2: O89746, Seq 3: P09345, Seq 4: Q8QPL1, Seq 5: Q9Q0U6)

Peptide	Score	Percentile Rank	Seq Num	Antigenicity	Probability	Toxicity
RSIPEIATR	0.953	0.01	3	0.7685	Antigen	Non-Toxin
NTQFEAVGR	0.918	0.06	1,2,4,5	1.2894	Antigen	Non-Toxin
TLNQRLVPK	0.897	0.03	4	0.7616	Antigen	Non-Toxin
REEISGVKL	0.979	0.01	1,2,3,4,5	0.6846	Antigen	Non-Toxin
LYDKVRLQL	0.906	0.01	1,2,3,4,5	0.6437	Antigen	Non-Toxin
CPYHGRSSF	0.924	0.04	5	1.8587	Antigen	Non-Toxin
CPYIGRSSF	0.898	0.04	3	1.9088	Antigen	Non-Toxin
CPYLGRSSF	0.921	0.04	1	1.921	Antigen	Non-Toxin
CPYQGKSSF	0.954	0.02	4	1.7085	Antigen	Non-Toxin
MPFHNIHPL	0.946	0.02	1	1.2633	Antigen	Non-Toxin
QSGRMEFFW	0.966	0.05	1	1.0528	Antigen	Non-Toxin
RLVPKIATR	0.963	0.01	4	1.0086	Antigen	Non-Toxin
TIMEKNVTV	0.942	0.02	1	0.716	Antigen	Non-Toxin
VLATGLRNV	0.943	0.02	3	0.8085	Antigen	Non-Toxin
AIMKSGLAY	0.925	0.01	3	-0.1689	Non-Antigen	Non-Toxin
SSFFRNVVW	0.980	0.03	1	-0.0903	Non-Antigen	Non-Toxin
WLIKKNSTY	0.905	0.02	2	0.362	Non-Antigen	Non-Toxin

Table 3 High-Affinity MHC-I Binding Peptides for Neuraminidase Proteins with their Antigenicity and Toxicity prediction. Seq Num refer to the corresponding HA proteins: (Seq1: Q6DPL2, Seq2: Q710U6, Seq3: Q809U7, Seq4: Q809V0, Seq5: Q809V2, Seq6: Q9W7Y7)

Peptide	Score	Percentile Rank	Seq Num	Antigenicity	Probability	Toxicity
LQIGNIISV	0.969	0.01	6	1.2507	Antigen	Non-Toxin
NSDTVGWSW	0.930	0.04	1,2,3,4,5,6	0.6144	Antigen	Non-Toxin
VQHPELTGV	0.908	0.03	3	1.305	Antigen	Non-Toxin
MVIGIVSLM	0.897	0.02	3,4,5	1.1294	Antigen	Non-Toxin
AYGIKGFSF	0.928	0.02	5	1.2968	Antigen	Non-Toxin
AYGVKGFSF	0.932	0.02	1,2,4,6	1.2975	Antigen	Non-Toxin
SPSPYNSRF	0.907	0.04	2	0.9266	Antigen	Non-Toxin
TETDSSFSL	0.983	0.01	6	1.275	Antigen	Non-Toxin
EAPSPYNSR	0.911	0.07	1,3,4,5,6	0.5467	Antigen	Non-Toxin
ESPSPYNSR	0.908	0.07	2	0.6920	Antigen	Non-Toxin
CPINGWAVY	0.922	0.03	1	-0.24	Non-Antigen	Non-Toxin
CPISGWAVY	0.945	0.02	3	0.2021	Non-Antigen	Non-Toxin
ELDAPNYHY	0.955	0.02	1	0.2911	Non-Antigen	Non-Toxin
IITDTIKSW	0.936	0.08	1	-0.432	Non-Antigen	Non-Toxin
NPNQKIITI	0.916	0.02	1	0.3888	Non-Antigen	Non-Toxin
SACHDGISW	0.944	0.04	6	-0.2518	Non-Antigen	Non-Toxin
SSLCPINGW	0.942	0.07	1	0.1678	Non-Antigen	Non-Toxin

3.2.2. MHC-II epitope prediction

The MHC-II binding prediction revealed several high-affinity peptides across the selected HA and NA proteins. Each peptide predicted for MHC-II binding is characterized by its unique sequence, a core binding region responsible for interaction with HLA molecules, and a binding score, where higher values >0.9 indicate stronger binding. The percentile rank reflects the relative affinity of the peptide, with ranks ≤2% indicating high-affinity binders. Peptides were associated with specific proteins and HLA-DR alleles, emphasizing their relevance for vaccine design. These key details are summarized in Tables 4 and 5. Also, the antigenicity and toxicity of all selected epitopes were evaluated using VaxiJen and ToxinPred.

Table 4 High-Affinity, Antigenic, and Non-Toxic MHC-II Binding Peptides Predicted for Hemagglutinin Proteins with their Antigenicity score. Seq Num refers to the corresponding HA proteins: (Seq1: O56140, Seq2: O89746, Seq3: P09345, Seq4: Q8QPL1, Seq5: Q9Q0U6)

Peptide	Seq Num	Score	rank	Antigenicity	Probability	Toxicity
SSMPFHNIHPLTIGE	1	0.9381	0.08	1.3426	Antigen	Non-Toxin

APEYAYKIVKKGDST	1,2	0.9793	0.01	0.6221	Antigen	Non-Toxin
APEYAYKIVKKGDST	1,2	0.966	0.01	0.6221	Antigen	Non-Toxin
SSMPFHNIHPHTIGE	3	0.944	0.07	0.7154	Antigen	Non-Toxin
IAPEYAYKIVKKGDS	1,2,3,4,5	0.9292	0.04	0.7989	Antigen	Non-Toxin
NSSMPFHNIHPHTIG	3	0.9024	0.15	0.8068	Antigen	Non-Toxin
NTQFEAVGREFNNLE	1,2,4,5	0.916	0.31	1.1300	Antigen	Non-Toxin
NTQFKAVGKEFNNLE	3	0.9481	0.13	1.0860	Antigen	Non-Toxin
GKEFNNLERRVENLN	3	0.9184	0.29	0.8247	Antigen	Non-Toxin
EWSYIVEKASPANDL	1,2,4,5	0.9875	0.01	0.5948	Antigen	Non-Toxin
GREFNNLERRIENLN	1,2,4,5	0.897	0.47	0.8032	Antigen	Non-Toxin
APEYAYKIVKKGDSA	3	0.9812	0.01	0.6399	Antigen	Non-Toxin
EYAYKIVKKGDSAIM	3	0.9851	0.01	0.6036	Antigen	Non-Toxin
MNTQFKAVGKEFNNL	3	0.9144	0.32	0.9841	Antigen	Non-Toxin
PEWSYIVEKASPAND	1	0.9894	0.01	0.6112	Antigen	Non-Toxin
PEYAYKIVKKGDSAI	3	0.9861	0.01	0.7066	Antigen	Non-Toxin
PTTYVSVGTSTLNQR	3	0.9077	0.15	1.0272	Antigen	Non-Toxin

Table 5 High-Affinity, Antigenic, and Non-Toxic MHC-II Binding Peptides Predicted for Neuraminidase Proteins with their Antigenicity score. Seq Num refer to the HA proteins: (Seq1: Q6DPL2, Seq2: Q710U6, Seq3: Q809U7, Seq4: Q809V0, Seq5: Q809V2, Seq6: Q9W7Y7)

Peptide	Seq Num	Score	Rank	Antigenicity	Probability	Toxicity
SNTNFLTEKAVASVK	1	0.9504	0.21	0.5086	Antigen	Non-Toxin
NTNFLTEKAVASVKL	1	0.9369	0.29	0.5184	Antigen	Non-Toxin
WAVYSKDNGIRIGSK	3,5	0.891	0.11	0.8890	Antigen	Non-Toxin
WAIYSKDNSIRIGSK	6	0.9115	0.1	1.1726	Antigen	Non-Toxin
WAVYSKDNSIRIGSK	1	0.9214	0.08	1.0797	Antigen	Non-Toxin

3.2.3. Linear B-cell Epitopes Prediction

The ABCpred server identified several 16-mer linear B-cell epitopes from the HA and NA proteins of the vaccine candidates. These epitopes were ranked based on their scores, with higher scores >0.9 reflecting a greater probability of being immunogenic. The predicted epitopes are summarized in Table 6, including their protein source, sequence, associated scores, and antigenicity and toxicity of all selected epitopes was evaluated using VaxiJen and ToxinPred.

Table 6 High-Affinity Linear B Lymphocyte (LBL) Epitopes Predicted for HA and NA Proteins, with antigenicity and toxicity prediction.

Accession	Protein	Sequence	Score	Antigenicity	Probability	Toxicity
Q6DPL2	NA	AGEITCVCRDNWHGSN	0.90	0.84	Antigen	Non-Toxin
Q809U7	NA	SHSIQTGNQHQAEPCN	0.93	0.6402	Antigen	Non-Toxin
Q809U7	NA	YHYEECSCYPDAGEIT	0.91	0.5496	Antigen	Non-Toxin
Q710U6	NA	SCPIGESPSPYNSRFE	0.91	0.5699	Antigen	Non-Toxin
O89746	HA	DSTIMKSELEYGNCNT	0.93	0.6012	Antigen	Non-Toxin
P09345	HA	ICIGYHANKSTKQVDT	0.91	0.9065	Antigen	Non-Toxin
O56140	HA	VDTIMEKNVTVTHAQD	0.9	0.6865	Antigen	Non-Toxin
O56140	HA	VLWGIHHPNDAAEQTK	0.86	0.5440	Antigen	Non-Toxin
P09345	HA	CPYIGRSSFFRNVVWL	0.92	0.6270	Antigen	Non-Toxin
Q710U6	NA	AGEIMCVCRDNWHGSN	0.92	0.4084	Non-Antigen	Non-Toxin
Q6DPL2	NA	IGYICSGVFGDNPRPN	0.92	0.4733	Non-Antigen	Non-Toxin
Q6DPL2	NA	NQHQSEPISNTNFLTE	0.9	0.4560	Non-Antigen	Non-Toxin
Q710U6	NA	SHSIQTGNQNQPEICN	0.91	0.4604	Non-Antigen	Non-Toxin
Q710U6	NA	CFTIMTDGPSNGQASY	0.92	0.0736	Non-Antigen	Non-Toxin
Q6DPL2	NA	DGTGSCGPVSSNGAYG	0.91	0.3121	Non-Antigen	Non-Toxin
Q6DPL2	NA	FEMIWDPNGWTETDSS	0.92	0.1037	Non-Antigen	Non-Toxin
P09345	HA	PHTIGECPKYVKSDRL	0.9	0.2492	Non-Antigen	Non-Toxin
O56140	HA	TKLYQNPTTYISVGTS	0.97	0.2942	Non-Antigen	Non-Toxin
P09345	HA	VPEWSYIVEKDNPINS	0.91	0.3740	Non-Antigen	Non-Toxin
Q9Q0U6	HA	ESTQKAIDGVTNKVNS	0.88	0.2418	Non-Antigen	Non-Toxin

3.3. Epitope Screening and Characterization

The screening and characterization of epitopes yielded promising candidates, detailed in the corresponding tables below. Antigenicity and toxicity analysis of all epitopes are included in the previous tables 1-6. These epitopes were selected for allergenicity test too, table 7-9. Non-allergenic epitopes selected for further cytokine inducibility analysis. IFN-γ Inducibility for MHC-II epitopes was evaluated using the IFN-γ Epitope Server, while IL-4 Inducibility of the MHC-II epitopes were determined using IL4pred, the results are presented in Table 8. The immunogenicity analysis for MHC-I epitopes using the Class I Immunogenicity tool indicated several high-potential CTL-activating peptides, score > 0, as detailed in Table 7. The multi-step screening and characterization process ensured the selection of epitopes with high antigenic potential, immunogenicity, and safety profiles, laying the foundation for further development steps. The final selected epitopes after filtration and screening were put in table 10.

Table 7 Antigenic, and non-toxic MHC-I Epitopes with Allergenicity and Class I immunogenicity.

Protein	peptide	Antigenicity	Allergenicity	Class I Immunogenicity
HA	RSIPEIATR	0.7685	Non-Allergen	0.31052
HA	NTQFEAVGR	1.2894	Non-Allergen	0.26917
HA	TLNQRLVPK	0.7616	Non-Allergen	-0.05034
HA	REEISGVKL	0.6846	Non-Allergen	-0.05394
HA	LYDKVRLQL	0.6437	Non-Allergen	-0.19792
HA	CPYHGRSSF	1.8587	Allergen	-0.12321
HA	CPYIGRSSF	1.9088	Allergen	-0.02184
HA	CPYLGRSSF	1.921	Allergen	-0.16692
HA	CPYQGKSSF	1.7085	Allergen	-0.52404
HA	MPFHNIHPL	1.2633	Allergen	0.21035
HA	QSGRMEFFW	1.0528	Allergen	0.15353
HA	RLVPKIATR	1.0086	Allergen	-0.02678
HA	TIMEKNVTV	0.716	Allergen	-0.11482
HA	VLATGLRNV	0.8085	Allergen	0.11422
NA	LQIGNIISV	1.2507	Non-Allergen	0.21194
NA	NSDTVGWSW	0.6144	Non-Allergen	0.20864
NA	VQHPELTGV	1.305	Allergen	0.13896
NA	MVIGIVSLM	1.1294	Non-Allergen	0.09966
NA	AYGIKGFSF	1.2968	Non-Allergen	-0.03104

NA	AYGVKGFSF	1.2975	Non-Allergen	-0.12342
NA	SPSPYNSRF	0.9266	Non-Allergen	-0.18393
NA	TETDSSFSL	1.275	Allergen	-0.27977
NA	EAPSPYNSR	0.5467	Non-Allergen	-0.28647
NA	ESPSPYNSR	0.6920	Non-Allergen	-0.28647

Table 8 Antigenic, non-allergenic, and non-toxic MHC-I Epitopes, with IFN- and IL4-inducibility Prediction for Na and HA Proteins.

Protein	peptide	Allergenicity	IFN Score	IFN Result	SVM Score	IL4 inducer
HA	APEYAYKIVKKGDST	Non-Allergen	0.4336381	Positive	-0.13	Non inducer
HA	APEYAYKIVKKGDST	Non-Allergen	0.344	Positive	-0.13	Non inducer
HA	EWSYIVEKASPANDL	Non-Allergen	-0.6118637	Negative	1.07	IL4 inducer
HA	GKEFNNLERRVENLN	Non-Allergen	-0.4815333	Negative	0.61	IL4 inducer
HA	GREFNNLERRIENLN	Non-Allergen	-0.6482728	Negative	0.33	IL4 inducer
HA	IAPEYAYKIVKKGDS	Non-Allergen	0.196069	Positive	-0.18	Non inducer
HA	NSSMPFHNIHPHTIG	Non-Allergen	0.07632	Positive	0.24	IL4 inducer
NA	NTNFLTEKAVASVKL	Non-Allergen	0.5315287	Positive	-0.12	Non inducer
HA	NTQFEAVGREFNNLE	Non-Allergen	-0.2251253	Negative	0.38	IL4 inducer
HA	NTQFKAVGKEFNNLE	Non-Allergen	-0.3013475	Negative	1.28	IL4 inducer
NA	SNTNFLTEKAVASVK	Non-Allergen	0.5896483	Positive	-0.08	Non inducer
HA	SSMPFHNIHPHTIGE	Non-Allergen	0.3417935	Positive	0.24	IL4 inducer
HA	SSMPFHNIHPLTIGE	Non-Allergen	0.4984038	Positive	-0.03	Non inducer
NA	WAIYSKDNSIRIGSK	Non-Allergen	-0.1124776	Negative	0.28	IL4 inducer
NA	WAVYSKDNGIRIGSK	Non-Allergen	-0.2008197	Negative	0.24	IL4 inducer
HA	APEYAYKIVKKGDSA	Allergen				
HA	EYAYKIVKKGDSAIM	Allergen				
HA	MNTQFKAVGKEFNNL	Allergen				
HA	PEWSYIVEKASPAND	Allergen				
HA	PEYAYKIVKKGDSAI	Allergen				
HA	PTTYVSVGTSTLNQR	Allergen				
NA	WAVYSKDNSIRIGSK	Allergen				

Table 9 Antigenic, non-toxic, non-allergenic, and high-affinity Linear B Lymphocyte Binding Peptides Predicted for NA and HA Proteins

Accession	Protein	Sequence	Probablity	Allergenicity
O89746	HA	DSTIMKSELEYGNCNT	Probable Antigen	NON-ALLERGEN
P09345	HA	ICIGYHANKSTKQVDT	Probable Antigen	NON-ALLERGEN
O56140	HA	VDTIMEKNVTVTHAQD	Probable Antigen	NON-ALLERGEN
O56140	HA	VLWGIHHPNDAAEQTK	Probable Antigen	NON-ALLERGEN
Q6DPL2	NA	AGEITCVCRDNWHGSN	Probable Antigen	NON-ALLERGEN
Q809U7	NA	SHSIQTGNQHQAEPCN	Probable Antigen	NON-ALLERGEN
Q809U7	NA	YHYEECSCYPDAGEIT	Probable Antigen	NON-ALLERGEN

Table 10 Final selected epitopes after filtration, considering high-affinity, safety, ability to inducing strong immunity and the production of cytokines (epitopes highlighted in red can induce both IFN and IL4)

Protein	Seq Nr	Peptide	Eptiope Type
HA	P09345	RSIPEIATR	MHC-I
HA	O56140 O89746 Q8QPL1 Q9Q0U6	NTQFEAVGR	MHC-I
NA	Q9W7Y7	LQIGNIISV	MHC-I
NA	All	NSDTVGWSW	MHC-I
NA	Q809V2 Q809V0 Q809U7	MVIGIVSLM	MHC-I
HA	O56140 O89746	APEYAYKIVKKGDST	MHC-II
HA	All	IAPEYAYKIVKKGDS	MHC-II
HA	P09345	NSSMPFHNIHPHTIG	MHC-II
NA	Q6DPL2	NTNFLTEKAVASVKL	MHC-II
NA	Q6DPL2	SNTNFLTEKAVASVK	MHC-II
HA	P09345	SSMPFHNIHPHTIGE	MHC-II
HA	O56140	SSMPFHNIHPLTIGE	MHC-II
HA	O89746	DSTIMKSELEYGNCNT	Linear B-Cell
HA	P09345	ICIGYHANKSTKQVDT	Linear B-Cell
HA	O56140	VDTIMEKNVTVTHAQD	Linear B-Cell
HA	O56140	VLWGIHHPNDAAEQTK	Linear B-Cell
NA	Q6DPL2	AGEITCVCRDNWHGSN	Linear B-Cell
NA	Q809U7	SHSIQTGNQHQAEPCN	Linear B-Cell
NA	Q809U7	YHYEECSCYPDAGEIT	Linear B-Cell

3.4. Population Coverage Calculation

The global population coverage for MHC class I epitopes was 99.3%. The average hit was 14.42, indicating that individuals recognize, on average, over 14 epitope-HLA combinations. The PC90 value was 10.12, showing the minimum number of combinations recognized by 90% of the population. The population coverage for MHC class II epitopes was 81.81%. The average hit was 7.75, with a PC90 value of 3.85. When both MHC class I and MHC class II epitopes were analyzed together, the global combined coverage reached 99.87%, indicating nearly universal coverage. The average hit for combined coverage was 22.17, and the PC90 was 13.31.

These results, as summarized in table 11, demonstrate that the designed multi-epitope vaccine has exceptionally high population coverage, confirming its potential to elicit immune responses across a broad spectrum of individuals worldwide. The graphs presented in figures 3,4, and 5 illustrate the cumulative distribution of epitope hits/HLA combinations recognized, further validating the broad coverage achieved for MHC class I, MHC class II, and the combined classes.

Table 11 Population Coverage Calculation Result

population/area	Class I			Class II			Class combined		
	coverage	Average hit	pc90	coverage	Average hit	pc90	coverage	Average hit	pc90
World	**99.3%**	**14.42**	**10.12**	**81.81%**	**7.75**	**3.85**	**99.87%**	**22.17**	**13.31**

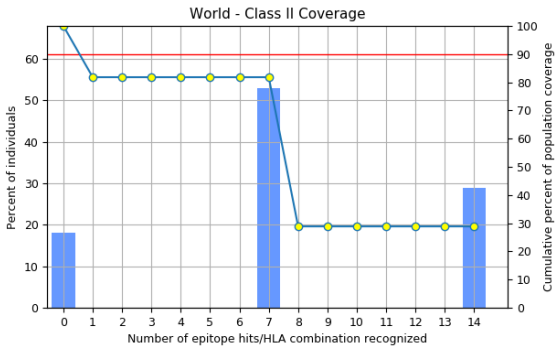


Figure 3 Class II coverage of the selected epitopes

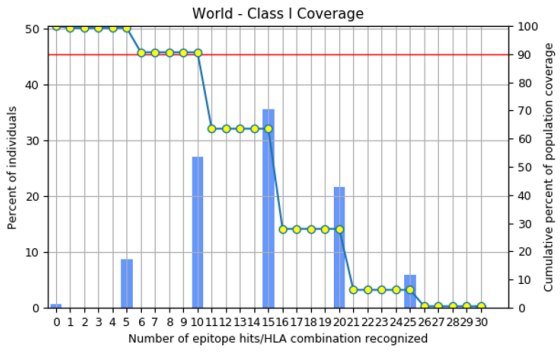


Figure 4 Class I coverage of the selected epitopes

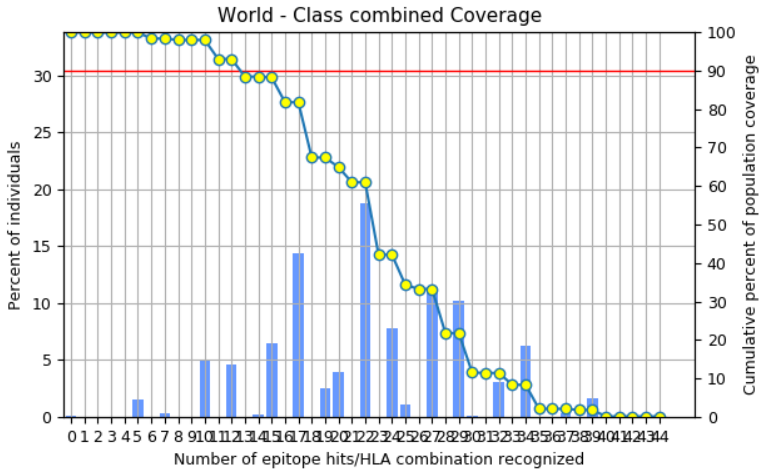


Figure 5 Class combined coverage of the selected epitopes

3.5. Epitope Conservancy Analysis

The conservancy of predicted epitopes for the HA and NA proteins was assessed using the IEDB Conservancy Analysis Tool. The analysis provided the percentage of sequence matches across related sequences at a specified identity threshold, along with the minimum identity level.

A total of 11 epitopes were analyzed for their degree of conservancy as shown in Table 12. Two epitopes demonstrated 100% conservation across all sequences. The remaining epitopes displayed varying levels of conservation, but one epitope RSIPEIATR exhibits low identity and conservation compared to the others, while SSMPFHNIHPHTIGE excluded due it's high similarity with NSSMPFHNIHPHTIG, to reduce the redundancy of epitopes.

Table 12 Epitope Conservancy Analysis results for HA proteins epitopes.

Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity
RSIPEIATR	9	20.00% (1/5)	66.67%
NTQFEAVGR	9	80.00% (4/5)	77.78%
APEYAYKIVKKGDST	15	40.00% (2/5)	93.33%
IAPEYAYKIVKKGDS	15	100.00% (5/5)	100.00%
NSSMPFHNIHPHTIG	15	20.00% (1/5)	93.33%
SSMPFHNIHPHTIGE	15	20.00% (1/5)	93.33%
SSMPFHNIHPLTIGE	15	80.00% (4/5)	93.33%
DSTIMKSELEYGNCNT	16	40.00% (2/5)	75.00%
ICIGYHANKSTKQVDT	16	20.00% (1/5)	87.50%

VDTIMEKNVTVTHAQD	16	100.00% (5/5)	100.00%
VLWGIHHPNDAAEQTK	16	60.00% (3/5)	93.75%

For the NA protein, 8 epitopes were evaluated as shown in Table 13, One epitope showed 100% conservation across all sequences (6/6) with a minimum identity of 100%, Two epitopes displayed relatively high conservation levels with 93.75% minimum identity. Other epitopes demonstrated moderate to low conservancy, while SNTNFLTEKAVASVK excluded due its high similarity with NTNFLTEKAVASVKL and low Min. identity, to reduce the redundancy of epitopes.

Table 13 Epitope Conservancy Analysis results for NA proteins epitopes.

Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity
LQIGNIISV	9	16.67% (1/6)	77.78%
NSDTVGWSW	9	100.00% (6/6)	100.00%
MVIGIVSLM	9	50.00% (3/6)	55.56%
NTNFLTEKAVASVKL	15	16.67% (1/6)	60.00%
SNTNFLTEKAVASVK	15	16.67% (1/6)	53.33%
AGEITCVCRDNWHGSN	16	83.33% (5/6)	93.75%
SHSIQTGNQHQAEPCN	16	33.33% (2/6)	62.50%
YHYEECSCYPDAGEIT	16	66.67% (4/6)	93.75%

3.6. Construction of the Multi-Epitope Vaccine:

The designed multi-epitope vaccine construct integrates 17 epitopes comprising 5 CTLs, 5 HTL, and 7 LBL epitopes, strategically connected using specific linkers and adjuvants to ensure immunogenicity, structural stability, and efficient antigen presentation.

Two potent immune-stimulatory adjuvants were incorporated in the vaccine construct: MDA5 (MGPGQGPAKGLVLQEKYLGRL) at the N-terminus, H9E (FIEGGWTGMIDGWYG) at the C-terminus. The two adjuvants were linked using a rigid EAAAK linker to maintain structural independence, ensuring functional separation and proper folding of the vaccine construct.

To optimize the immunological processing and presentation of epitopes, specific linkers were employed between different epitope categories: AAY Linker used to connect the CTL epitopes, GPGPG Linker applied between the HTL epitopes. KK Linker used to connect the LBL epitopes.

The resulting primary sequence of the vaccine is as follows:

MGPGQGPAKGLVLQEKYLGRLEAAAKRSIPEIATRAAYNTQFEAVGRAAYLQIGNIISVAAYNSDTVGWS
WAAYMVIGIVSLMGPGPGAPEYAYKIVKKGDSTGPGPGNTNFLTEKAVASVKLGPGPGNSSMPFHNIH
PHTIGGPGPGSSMPFHNIHPLTIGEGPGPGIAPEYAYKIVKKGDSKKICIGYHANKSTKQVDTKKDSTIMKS
ELEYGNCNTKKYHYEECSCYPDAGEITKKVLWGIHHPNDAAEQTKKKAGEITCVCRDNWHGSNKKSHSI
QTGNQHQAEPCNKKVDTIMEKNVTVTHAQDEAAAKFIEGGWTGMIDGWYGGGGGSHHHHHH

The designed multi-epitope vaccine construct integrates adjuvants and linkers to optimize innate and adaptive immune responses. The inclusion of MDA5 and H9E, in conjunction with the specific linkers as shown in figure 2, ensures efficient epitope processing, presentation, and immune activation, and a 6 Histidine tag was added at the C-terminal part with linker GGGGS to improve protein purification and identification.

3.7. Specifications Prediction of the Vaccine Construct:

3.7.1. Physicochemical Properties

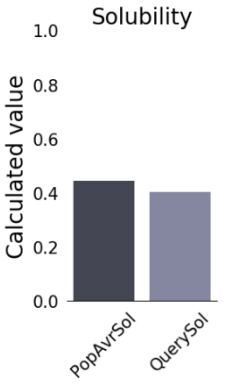
The vaccine construct comprised 340 amino acids with a molecular weight of 36480 Da and a theoretical isoelectric point (pI) of 8.76. The extinction coefficient was calculated to be 55350 $M^{-1}cm^{-1}$, and the chemical formula was determined as:

$$C_{1612}H_{2489}N_{457}O_{483}S_{15}$$

The construct exhibited a computed instability index of 35.67, classifying the protein as stable. The aliphatic index, a measure of thermal stability, was 62.62, indicating moderate thermostability. Additionally, the grand average of hydropathicity (GRAVY) was calculated as -0.570, suggesting that the protein has a hydrophilic nature.

3.7.2. Solubility Prediction

The solubility of the vaccine construct was evaluated using multiple tools. The Protein-Sol server predicted a scaled solubility value of 0.406, which is below the average population threshold, indicating moderate solubility. However, the SOLpro tool predicted the vaccine construct to be soluble with a high probability of 0.930, demonstrating its suitability for overexpression in Escherichia coli.



3.7.3. Secondary Structure Analysis

The secondary structure of the construct was analyzed using SOPMA and visualized using NetSurfP-2.0 shown in figure 6, revealing the following distribution: Alpha helices 63 residues (18.53%), Extended strands 94 residues (27.65%), Beta turns 23 residues (6.76%), Random coils: 160 residues (47.06%). This suggests that the protein predominantly consists of random coils, with moderate proportions of alpha helices and extended strands contributing to its structural integrity.

Structure	Residues	Percentage
Alpha helices	63	18.53%
Extended strands	94	27.65%
Beta turns	23	6.76%
Random coils	160	47.06%

The distribution of secondary structure elements indicates that the vaccine construct possesses a flexible yet stable conformation, with sufficient surface accessibility for antibody recognition and antigenic presentation. The balance between structural stability (alpha helices and extended strands) and flexibility (random coils and beta turns) supports the construct's suitability as an effective vaccine candidate.

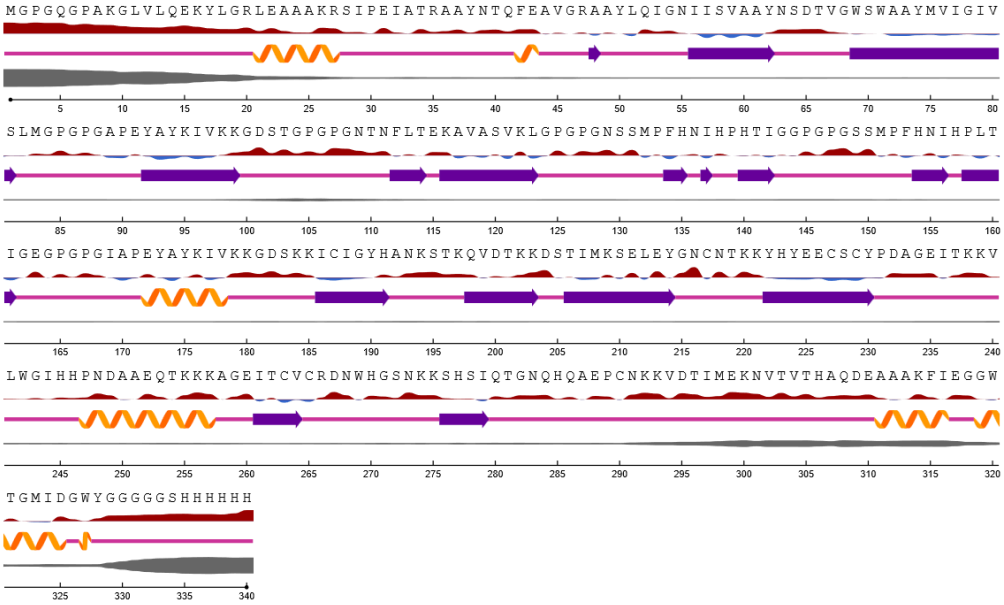


Figure 6 Secondary Structure from NetSurfP-2.0 shows Helix, Strand, Coil.

3.7.4. Antigenicity, Allergenicity, and Toxicity Predictions

The antigenicity of the construct was predicted using Vaxijen, yielding a score of 0.5737, classifying it as a probable antigen. The construct was also predicted to be a non-allergen according to AllerTOP and non-toxic as determined by the ToxDL server, which reported a low toxicity score with no toxic domains detected.

3.7.5. Homology Analysis

The homology analysis of the vaccine construct was performed using the BLASTp tool against the Homo sapiens protein database to evaluate the potential for cross-reactivity and unintended adverse immune responses. Two regions of alignment were detected, as follows:

Table 14 Homology analysis results

Range	Residues	Score	Expect Value	Identity	Positives	Gaps
1	221–314	47.8 bits	9e-05	32% (38/119)	42% (50/119)	21% (26/119)
2	256–366	39.7 bits	0.038	28% (49/175)	33% (59/175)	36% (64/175)

The sequence identities in both alignment ranges (32% and 28%) are well below the 35% threshold, which is generally considered the cutoff for predicting significant structural or

functional similarity that could result in autoimmunity or adverse reactions[58]. Sequence homology below this threshold typically indicates a low risk of molecular mimicry, reducing the likelihood of immune cross-reactivity with human proteins[59].

Furthermore, the observed sequence gaps and relatively low positive match percentages suggest poor structural conservation, supporting the safety profile of the vaccine construct[60]. This low homology reduces concerns regarding unintended host protein targeting or autoimmune reactions, which are critical considerations in vaccine safety evaluation[61].

Range 1: 221 to 314 GenPept Graphics ▼ Next Match ▲ P

Score	Expect	Method	Identities	Positives	Gaps
47.8 bits(112)	9e-05	Compositional matrix adjust.	38/119(32%)	50/119(42%)	26/119(21%)

```
Query  26   KRSIPEIATRAAYNTQFEAVGRAAYLQIGNIISVAAYNSDTVGWSWAAYMVIGIVSLMGP  85
            KRS P+IATR   N Q    GR  +                   SW    +   ++
Sbjct  221  KRSTPDIATRPKVNGQ---GGRMEF------------------SWTLLDMWDTINFEST  258

Query  86   GPG-APEYAYKIVKKGDSTGPGPGNTNFLTEKAVASVKLGPGPGNSSMPFHNIHPHTIG  143
            G   APEY +KI K+G S   G   T    E      +   G  N+++PFHN+HP TIG
Sbjct  259  GNLIAPEYGFKISKRGSS---GIMKTEGTLENCETKCQTPLGAINTTLPFHNVHPLTIG  314
```

Range 2: 256 to 366 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
39.7 bits(91)	0.038	Compositional matrix adjust.	49/175(28%)	59/175(33%)	64/175(36%)

```
Query  163  EGPGPGIAPEYAYKIVKKGDSKKICIGYHANKSTKQVDTKKDSTIMKSELEYGNCNTKKY  222
            E  G  IAPEY +KI K+G S                      IMK+E    NC TK
Sbjct  256  ESTGNLIAPEYGFKISKRGSSG---------------------IMKTEGTLENCETK--  291

Query  223  HYEECSCYPDAGEITKKVLWGIHHPNDAAEQTKKKAGEITCVCRDNWHGSNKKSHSIQTG  282
                  C    G I   + +   HP    E  K             +  S K   +
Sbjct  292  ------CQTPLGAINTTLPFHNVHPLTIGECPK-------------YVKSEKLVLATGLR  332

Query  283  NQHQAEPCNKKVDTIMEKNVTVTHAQDEAAAKFIEGGWTGMIDGWYGGGGGSHHH  337
            N  Q E                +     A A FIEGGW GM+DGWYG     +HH
Sbjct  333  NVPQIE----------------SRGLFGAIAGFIEGGWQGMVDGWYG-----YHH  366
```

Figure 7 The homologous sequences form BLASTp

3.8. Tertiary Structure Prediction and Validation

3.8.1. 3D Structure Prediction

The tertiary structure of the multi-epitope vaccine was initially predicted using PHYRE2, yielding a model with 99.5% confidence, shown in figure 8. However, the predicted structure exhibited a low sequence coverage of only 30%, corresponding to 102 amino acids out of the total 340 residues. This limited coverage indicates that a substantial portion of the protein sequence was not accurately modeled, potentially compromising the structural and functional representation of the full vaccine construct. Consequently, the PHYRE2-derived model was excluded from further analysis to ensure structural reliability and downstream accuracy.

In contrast, RoseTTAFold successfully generated five distinct predicted structural models for the multi-epitope vaccine, one of them shown in figure 9. These models were selected for further validation and refinement steps to assess their stereochemical quality, structural consistency, and overall reliability. The use of multiple models from RoseTTAFold allowed a comparative evaluation, increasing the likelihood of identifying the most accurate and stable structural representation of the vaccine candidate.

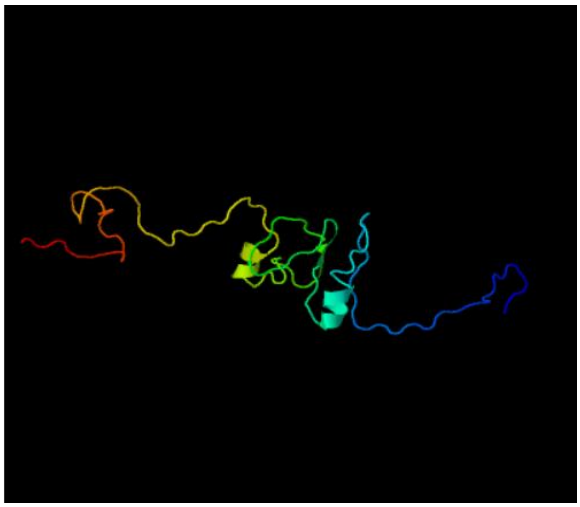


Figure 9 3D model generated with PHYRE2 with 30% coverage

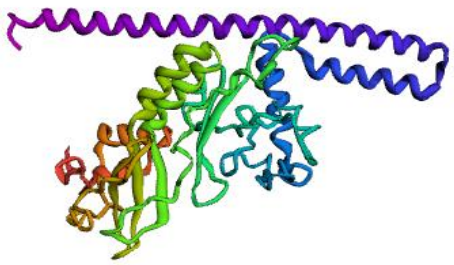


Figure 8 3D model generated with RoseTTAFold with 100% coverage

3.8.2. Model Validation

The five predicted 3D models generated by RoseTTAFold were subjected to validation using the SAVES v6.1 server, which integrates multiple structural evaluation tools, including PROCHECK and ERRAT, to assess the stereochemical quality and structural reliability of protein models.

PROCHECK was employed to generate Ramachandran plots, and ERRAT was used to calculate the Overall Quality Factor. Among the five models, two exhibited superior structural characteristics and were selected for further refinement:

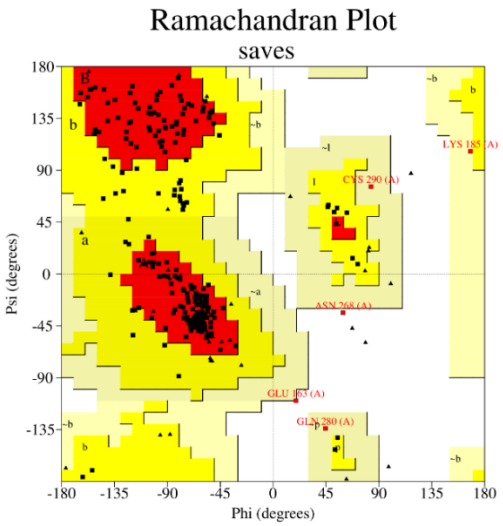


Figure 10 Model 1 Ramachandran Plot

Model 1: ERRAT Score: 93.67. Ramachandran Plot: Residues in most favored regions: 83.8%. Residues in additional allowed regions: 14.3%. Residues in generously allowed regions: 1.5%. Residues in disallowed regions: 0.4%, see figure 10.

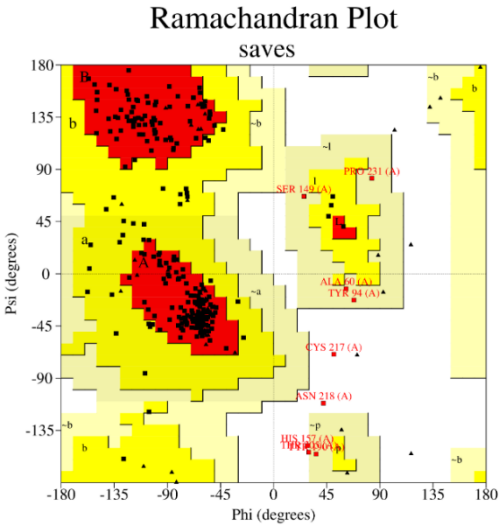


Figure 11 Model 2 Ramachandran Plot

Model 2: ERRAT Score: 88.92. Ramachandran Plot: Residues in most favored regions: 85.7%. Residues in additional allowed regions: 11.4%. Residues in generously allowed regions: 1.5%. Residues in disallowed regions: 1.5%, see figure 11.

The results from PROCHECK and ERRAT collectively indicate that Model 1 demonstrated slightly better stereochemical quality and atomic interaction reliability compared to Model 2, despite Model 2 showing a higher percentage of residues in favored regions. Both models were deemed suitable for the subsequent refinement step to further optimize their structural properties.

3.8.3. Model Refinement

The initial validation results from PROCHECK and ERRAT indicated that while the selected models exhibited satisfactory stereochemical quality and structural reliability, certain geometric inconsistencies and atomic interactions required refinement to enhance their stability and overall quality. To address these issues, the selected structures were submitted to the GalaxyRefine server, a widely used tool for protein structure refinement.

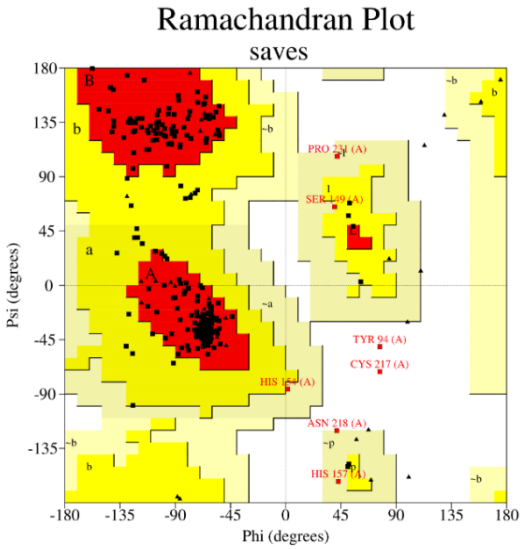


Figure 13 Ramachandran Plot for the validated and refined model. Favored: 89.7, Allowed: 8.1, Generosly allowed: 1.5, Disallowed: 0.7

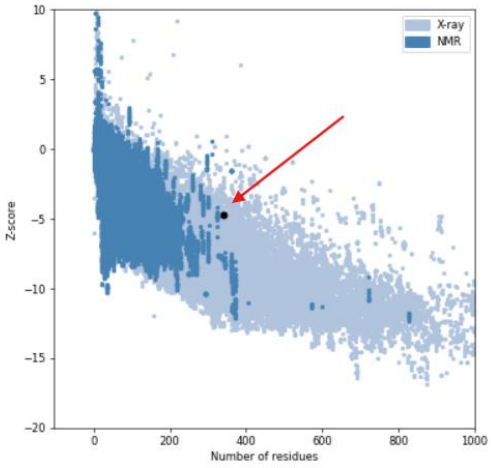


Figure 12 the Overall model quality calculated by ProSA server, showing z-Score -4.72

The GalaxyRefine server generated five refined versions of each selected model structure, providing 10 improved models ranked based on structural quality and refinement scores. Each refined model underwent further validation using PROCHECK and ERRAT to ensure structural improvements were effectively achieved.

The refined models exhibited enhancements in stereochemical geometry, with increased percentages of residues in favored regions of the Ramachandran plot 89.7, figure 12, and higher ERRAT score 89.13, reflecting improved structural consistency and reduced atomic energy conflicts. The best-refined structure demonstrated significant improvements in stability and quality metrics, making it suitable for subsequent docking analysis.

The z-score of the vaccine candidate after refinement - 4.72 is slightly in range of native protein conformation. It is depicted in figure 13 in a large black spot. z-Score plot consists of z-scores of all experimental protein chains in PDB defined by NMR spectroscopy (dark blue) and X-ray crystallography (light blue). The validated and refined tertiary structure was visualized using PyMOL (The PyMOL Molecular Graphics System, Version 2.0) to facilitate

further interpretation of the protein's structural properties, including secondary structure composition, ligand-binding sites, and spatial arrangement of epitopes, figure 14.

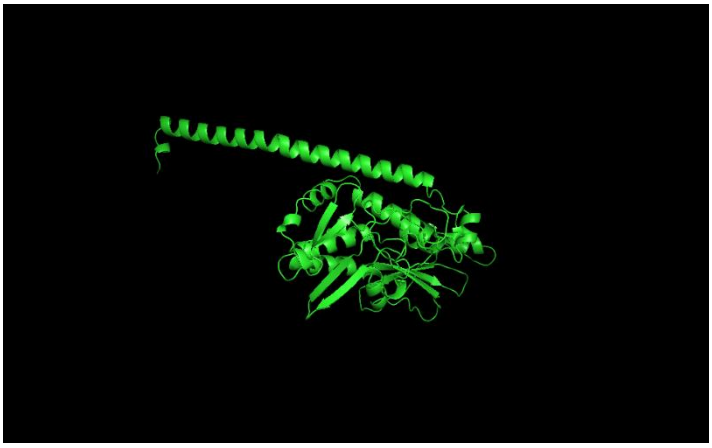


Figure 14 The 3d structure of validated and refined model visualized with PyMOL

3.9. Defining Discontinuous B-cell Epitopes (Conformational)

The refined 3D vaccine model was analyzed using the Ellipro server to predict discontinuous B-cell epitopes based on the protein's tertiary structure, shown in figure 15. The server identified several conformational B-cell epitopes with scores ranging from 0.869 to 0.533. These scores reflect the relative protrusion and spatial accessibility of the predicted epitopes on the protein surface, indicating their potential immunogenicity. The amino acid residues, sequence locations, and Ellipro scores of the identified epitopes are summarized in table 15.

Table 15 Predicted Discontinuous Epitopes

No.	Residues	Number of residues	Score
1	A:M1, A:G2, A:P3, A:G4, A:Q5, A:G6, A:P7, A:A8, A:K9, A:G10, A:L11, A:V12, A:L13, A:Q14, A:E15, A:K16, A:Y17, A:L18, A:G19, A:R20, A:L21, A:E22, A:A23, A:A24, A:A25, A:K26, A:R27, A:S28, A:I29, A:P30, A:E31, A:I32, A:A33, A:T34, A:A36, A:A37, A:T40	37	0.874
2	A:V178, A:K179, A:K180, A:G181, A:D182, A:S183, A:K184, A:K185	8	0.751
3	A:Y214, A:G215, A:N216, A:C217, A:N218, A:T219, A:K220, A:K221, A:Y222, A:H223, A:Y224, A:G259, A:E260, A:D267, A:N268, A:W269, A:H270, A:G271, A:S272, A:N273, A:Q280, A:T281, A:G282, A:N283, A:Q284, A:H285, A:Q286, A:A287, A:E288, A:P289, A:C290, A:N291, A:K292, A:K293, A:V294, A:D295, A:T296, A:I297, A:M298, A:E299, A:K300, A:N301, A:V302, A:T303, A:V304, A:T305, A:H306, A:A307, A:Q308, A:D309, A:E310, A:A311, A:A312, A:A313, A:K314, A:I316, A:E317	57	0.735
4	A:K99, A:G100, A:S102, A:T103, A:G104, A:P105, A:G106, A:P107, A:G108, A:N109, A:T110, A:N111, A:A117, A:V118, A:S120, A:K122, A:L123, A:G124, A:P125, A:G126, A:P127, A:G128, A:N129, A:S130	24	0.684
5	A:S334, A:H335, A:H338	3	0.664
6	A:G54, A:N55, A:I57, A:S58, A:V59, A:A60, A:A61, A:Y62, A:N63, A:S64, A:D65, A:T66, A:I142, A:G143, A:G144, A:P145, A:G146, A:P147, A:G148, A:S149, A:S150, A:M151, A:A170, A:P171, A:E172, A:Y173, A:H191, A:A192, A:N193, A:K194, A:S195, A:T196, A:K197, A:Q198, A:V199, A:D200, A:K202, A:K203	38	0.6

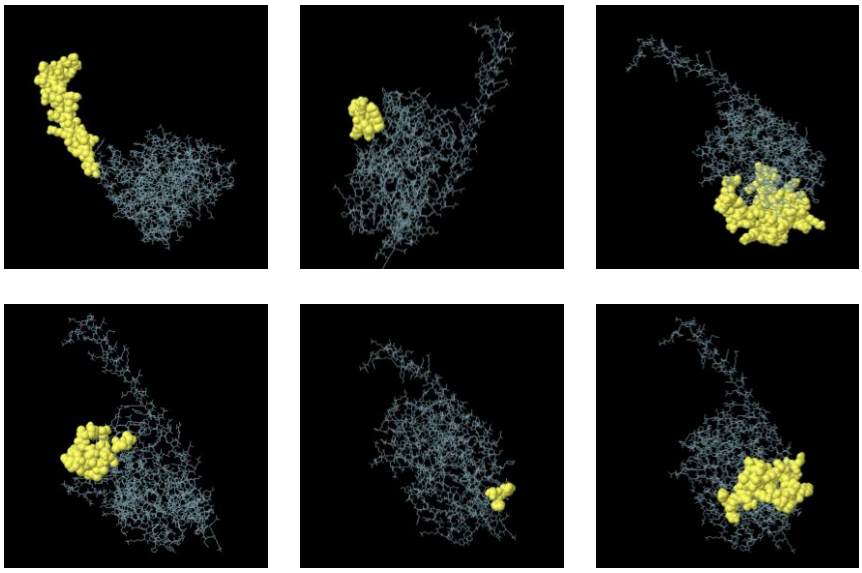


Figure 15 3D visualization of discontinuous B cell epitopes in H5N1 construct.

3.10. Immune Response Simulation

The immune response simulation of the multi-epitope vaccine construct was performed using the C-ImmSim server, which models immune dynamics across three anatomical compartments: the bone marrow, thymus, and lymph node. The simulation parameters included a volume of 10, HLA types (A0101, A0301, B3501, B4403, DRB1_0701, DRB1_0405), a random seed of 12345, 100 simulation steps, and a single injection set to 1. All other parameters were kept at their default values.

The simulation results revealed a robust humoral and cellular immune response. High levels of IgG and IgM antibodies were observed, with a subsequent increase in IgG1 levels correlating with antigen clearance. A strong interleukin and cytokine response was detected, including a notable rise in IFN-gamma concentration, indicating a potent cell-mediated immune response. Additionally, Populations of B-cells and T-helper cells increased significantly following the antigen exposure, reinforcing the vaccine's ability to activate both humoral and cellular immunity, see figure 16.

These results suggest that the vaccine candidate successfully induced a protective immune response capable of antigen clearance and immunological memory formation, supporting its potential as an effective immunogen.

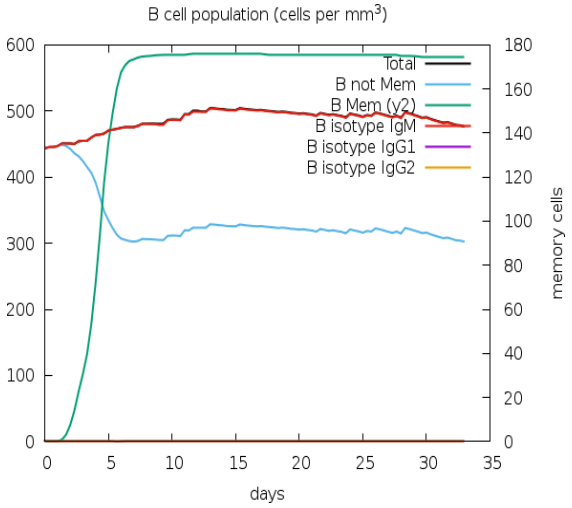


Figure 16.A

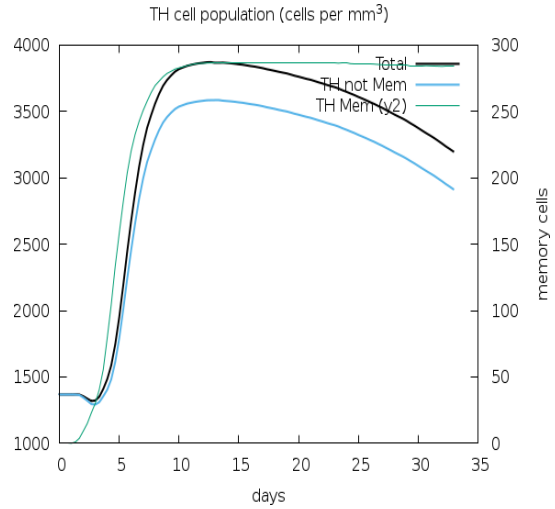


Figure 16.B

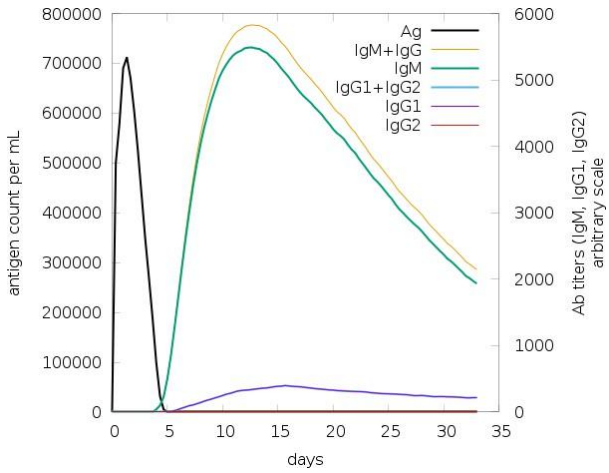


Figure 16.C

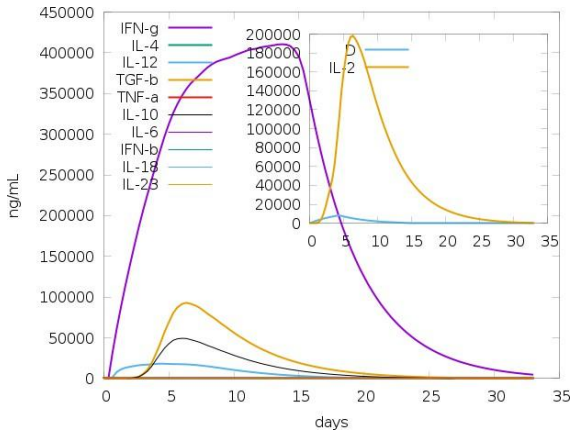


Figure 16.D

Figure 16 Immune Simulation Results
A: B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM, IgG1 and IgG2.
B: CD4 T-helper lymphocytes count. The plot shows total and memory counts.
C: The immunoglobulins and the immunocomplexes
D: Concentration of cytokines and interleukins. Inset plot shows danger signal together with leukocyte growth factor IL-2.

3.11. Molecular Docking with TLR7 and TLR8

Molecular docking was performed to predict the interaction between the vaccine construct and the TLR7 and TLR8, figure 18, using the HDOCK server. The docking simulation generated ten possible binding poses for each receptor, ranked based on binding energy scores, shown in tables 16-17. The pose with the lowest binding energy and the highest docking confidence score was selected for further analysis.

Table 16 Summary of the Top 10 Models for TLR7 with VC

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-357.18	-340.79	-330.98	-330.85	-328.95	-326.66	-325.60	-321.28	-319.38	-319.11
Confidence Score	0.9844	0.9785	0.9739	0.9738	0.9729	0.9716	0.9710	0.9685	0.9673	0.9671
Ligand rmsd (Å)	171.10	180.25	232.38	195.45	190.56	181.92	174.52	202.87	170.51	227.16

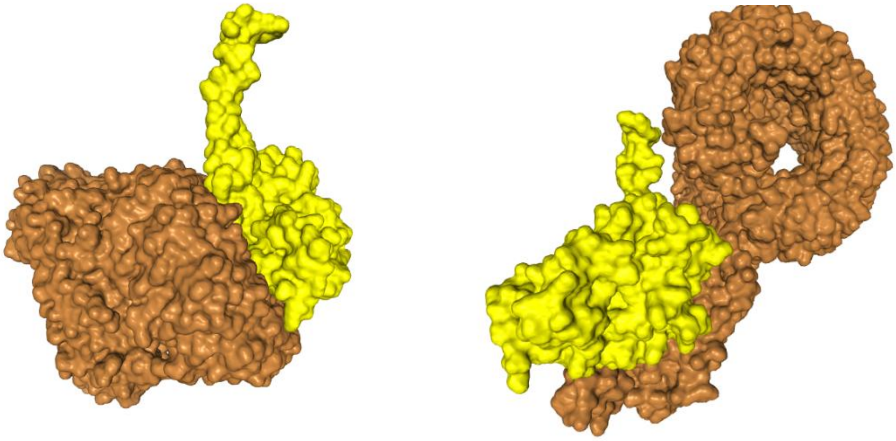
Table 17 Summary of the Top 10 Models for TLR8 with VC

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-307.39	-291.07	-281.95	-278.26	-274.11	-267.80	-263.35	-262.21	-259.38	-259.36
Confidence Score	0.9588	0.9438	0.9333	0.9286	0.9229	0.9134	0.9061	0.9041	0.8991	0.8991
Ligand rmsd (Å)	45.08	72.89	101.13	62.88	49.73	83.33	88.72	64.92	78.12	84.18

The vaccine construct demonstrated a strong binding affinity with TLR8, figure 17, with a docking score of -307.39. A detailed interaction analysis revealed the formation of 17 hydrogen bonds and 4 salt bridges at the binding interface. These interactions suggest significant stability and specificity of the complex, highlighting the potential of the vaccine construct to effectively engage with TLR8.

The docking results for the vaccine construct with TLR7, figure 20, showed a slightly higher binding affinity, with a docking score of -357.18. The analysis revealed 4 hydrogen bonds contributing to the stability of the receptor-ligand complex. Despite a lower number of hydrogen bonds compared to the TLR8 interaction, the stronger docking score suggests a favorable interaction driven by other stabilizing forces, potentially including hydrophobic interactions.

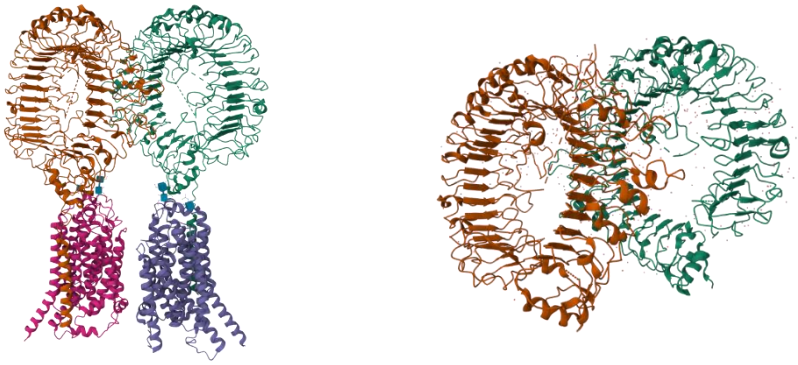
The top-ranked docking poses for both TLR7 and TLR8 complexes were visualized using PyMOL. Key residues involved in hydrogen bonding and salt bridge formation were identified and highlighted at the interaction interface, figure 19. The differences in hydrogen bonding patterns and salt bridge interactions between TLR7 and TLR8 complexes reflect distinct binding mechanisms, which may influence the downstream immune signaling pathways triggered by each receptor.



Predicted Interaction for TLR8 and VC

Predicted Interaction for TLR8 and VC

Figure 17 3D Image of Predicted interaction between TLR7/8 with VC

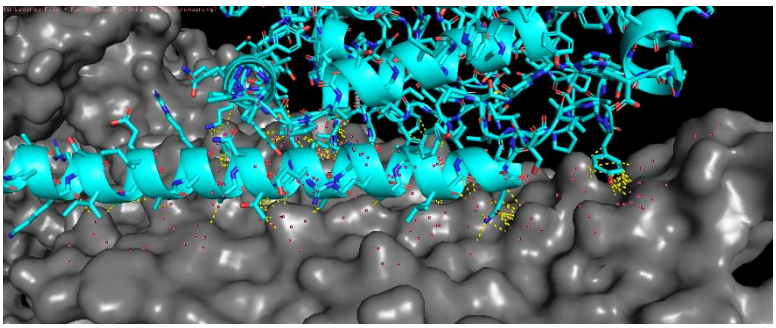


TLR7

TLR8 before removing water molecules

Figure 18 Structures of TLR7/8 retrieved from PDB

Receptor-Ligand interactions between VC and TLR7



Receptor-Ligand interactions between VC and TLR8

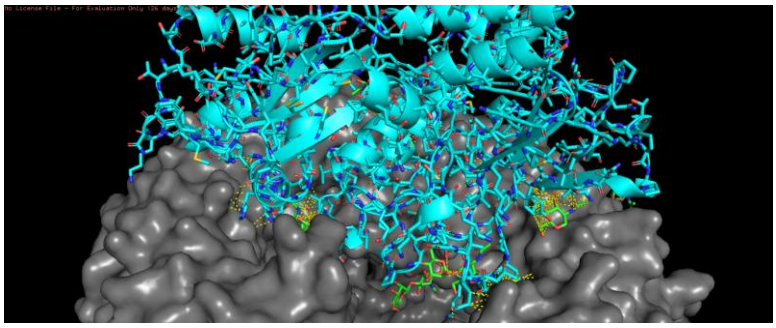


Figure 19 Receptor-Ligand interactions between VC and TLR7/8, yellow lines refer to the different bonds exist, visualized with PyMOL

3.12. Codon Optimization and in-silico cloning

The gene sequence encoding the recombinant vaccine construct, table 18, was successfully optimized for expression in Escherichia coli K12. The optimization process resulted in a Codon Adaptation Index (CAI) score of 1.0, indicating a highly favorable adaptation to the host's codon usage preferences. Additionally, the GC content of the optimized sequence was adjusted to 53.63%, falling within the optimal range for efficient transcription and stable mRNA structure. These metrics suggest a high potential for efficient and reliable protein expression in the selected host.

Table 18 The gene sequence encoding the recombinant vaccine construct

ATGGGTCCGGGTCAGGGTCCGGCTAAAGGTCTGGTTCTGCAGGAAAAATA	50
CCTGGGTCGTCTGGAAGCTGCTGCTAAACGTTCTATCCCGGAAATCGCTA	100
CCCGTGCTGCTTACAACACCCAGTTCGAAGCTGTTGGTCGTGCTGCTTAC	150
CTGCAGATCGGTAACATCATCTCTGTTGCTGCTTACAACTCTGACACCGT	200
TGGTTGGTCTTGGGCTGCTTACATGGTTATCGGTATCGTTTCTCTGATGG	250
GTCCGGGTCCGGGTGCTCCGGAATACGCTTACAAAATCGTTAAAAAAGGT	300
GACTCTACCGGTCCGGGTCCGGGTAACACCAACTTCCTGACCGAAAAAGC	350
TGTTGCTTCTGTTAAACTGGGTCCGGGTCCGGGTAACTCTTCTATGCCGT	400
TCCACAACATCCACCCGCACACCATCGGTGGTCCGGGTCCGGGTTCTTCT	450
ATGCCGTTCCACAACATCCACCCGCTGACCATCGGTGAAGGTCCGGGTCC	500
GGGTATCGCTCCGGAATACGCTTACAAAATCGTTAAAAAAGGTGACTCTA	550
AAAAAATCTGCATCGGTTACCACGCTAACAAATCTACCAAACAGGTTGAC	600
ACCAAAAAAGACTCTACCATCATGAAATCTGAACTGGAATACGGTAACTG	650
CAACACCAAAAAATACCACTACGAAGAATGCTCTTGCTACCCGGACGCTG	700
GTGAAATCACCAAAAAAGTTCTGTGGGGTATCCACCACCCGAACGACGCT	750
GCTGAACAGACCAAAAAAAAAGCTGGTGAAATCACCTGCGTTTGCCGTGA	800
CAACTGGCACGGTTCTAACAAAAAATCTCACTCTATCCAGACCGGTAACC	850
AGCACCAGGCTGAACCGTGCAACAAAAAAGTTGACACCATCATGGAAAAA	900
AACGTTACCGTTACCCACGCTCAGGACGAAGCTGCTGCTAAATTCATCGA	950
AGGTGGTTGGACCGGTATGATCGACGGTTGGTACGGTGGTGGTGGTGGTT	1000
CTCACCACCACCACCACCAC	1050

To facilitate cloning into the pET-26b(+) vector, restriction enzyme sites for NdeI and XhoI were manually introduced at the 5’ and 3’ ends of the optimized gene sequence, respectively. The recombinant construct, shown in red color in figure 25, was successfully inserted into the vector using SnapGene software, with proper alignment and orientation confirmed in silico.

The final construct was validated through virtual restriction digestion and sequence analysis, confirming the integrity and correctness of the gene insertion. These results indicate that the optimized construct is ready for downstream expression and protein production experiments.

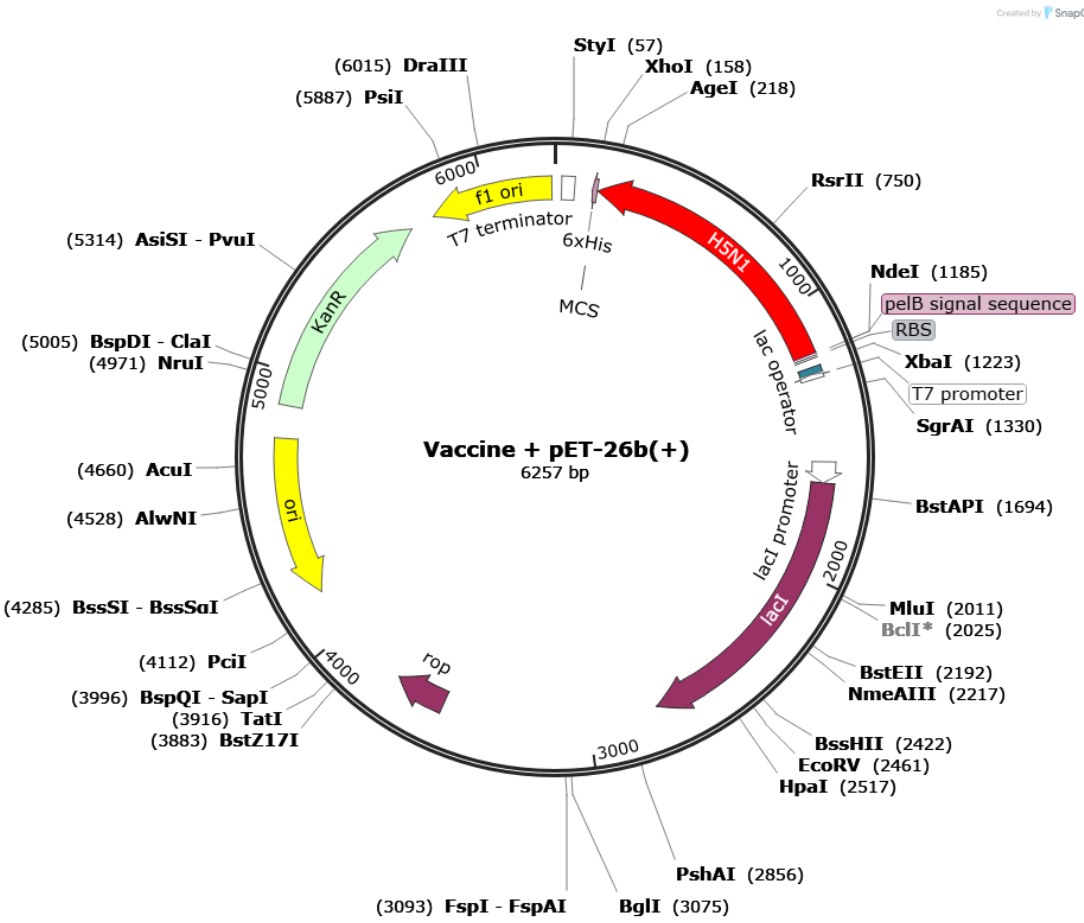


Figure 20 In-silico cloning of the vaccine candidate into pET23d (+) expression vector using Snapgene software. The red arrow represents the coding sequence of the vaccine candidate, and the black circle represents the vector backbone.

4. Discussion

The study aimed to design a multi-epitope vaccine targeting the H5N1 avian influenza virus using immunoinformatics tools. The results highlight the successful identification, characterization, and evaluation of conserved epitopes capable of eliciting robust humoral and cellular immune responses. This approach represents an efficient and cost-effective method to address the limitations of traditional vaccines, particularly in combating highly variable pathogens like avian influenza viruses.

The initial epitope prediction for MHC-I, MHC-II, and linear B-cell epitopes resulted in the identification of several highly antigenic, non-toxic, and non-allergenic candidates. The selected epitopes demonstrated strong binding affinities with prevalent HLA alleles across global populations, as evidenced by high percentile ranks and IC50 values. Importantly, both CTL and HTL epitopes were identified to ensure comprehensive immune activation. Linear B-cell epitopes further contributed to antibody-mediated immune responses, providing an additional layer of protection.

The population coverage analysis revealed exceptional global coverage for the selected epitopes, with combined MHC-I and MHC-II epitopes covering nearly 99.87% of the global population. This indicates the vaccine's potential for universal applicability, an essential feature for pandemic preparedness. These results align with previous immunoinformatics-driven vaccine studies, which emphasize the importance of broad HLA allele coverage in multi-epitope vaccine design[9, 22].

Epitope conservancy analysis demonstrated that a significant proportion of the selected epitopes were conserved across multiple H5N1 viral strains. Conserved epitopes are crucial for providing cross-strain immunity, especially for highly mutable viruses like H5N1. For instance, epitopes such as IAPEYAYKIVKKGDS and VDTIMEKNVTVTHAQD exhibited complete conservancy, underscoring their potential as universal vaccine targets. Similar findings have been reported in computational studies targeting conserved epitopes in other influenza subtypes[37].

However, certain epitopes displayed lower levels of conservancy, which may limit their effectiveness across different viral strains. Nevertheless, the overall inclusion of conserved epitopes enhances the robustness of the designed vaccine construct, offering broad-spectrum protection.

The multi-epitope vaccine construct was designed with carefully selected adjuvants (MDA5, H9E) and immunogenic linkers (AAY, GPGPG, KK) to enhance structural stability and facilitate efficient antigen presentation. Physicochemical analysis revealed favorable properties, including high stability, moderate hydrophilicity, and predicted solubility in bacterial

expression systems. These parameters are consistent with established criteria for protein-based vaccines[41, 44].

Secondary structure analysis indicated a balance between α-helices, β-strands, and random coils, supporting the vaccine's structural integrity and immunogenic potential. The tertiary structure was successfully modeled and refined using RoseTTAFold and validated using multiple tools, including Ramachandran plot, ERRAT, and ProSA. The refined model exhibited high stereochemical quality and a Z-score within the range of native proteins, indicating structural reliability.

Molecular docking studies with TLR7 and TLR8 revealed stable and specific interactions between the vaccine construct and the receptors. Both complexes demonstrated favorable docking scores, with TLR7 showing slightly higher binding affinity. Hydrogen bonding and salt bridge interactions were observed at the receptor-ligand interface, suggesting robust immunological signaling pathways. The differential binding patterns between TLR7 and TLR8 may indicate distinct downstream immune responses, which warrant further experimental validation[55].

Immune response simulation using C-ImmSim provided critical insights into the vaccine's immunogenicity. The simulation predicted strong activation of both humoral (elevated IgG and IgM levels) and cellular immunity (increased T-helper and B-cell populations). High interferon-gamma (IFN-γ) concentrations further emphasized the construct's potential to elicit Th1-mediated immune responses, which are critical for combating viral infections. These observations align with previous immunoinformatics studies that have highlighted the importance of IFN-γ induction in vaccine efficacy[35].

Codon optimization was performed to ensure efficient expression of the vaccine construct in *Escherichia coli*. The optimized construct achieved a Codon Adaptation Index (CAI) of 1.0 and a GC content of 53.63%, both of which fall within the optimal range for bacterial expression systems[56, 57]. Virtual cloning into the pET-26b(+) expression vector confirmed proper orientation and integrity of the gene sequence, paving the way for future experimental validation.

Implications and Limitations

This study demonstrates the power of immunoinformatics in accelerating vaccine design, offering a cost-effective and time-efficient alternative to traditional vaccine development pipelines. The incorporation of both CTL and HTL epitopes, along with linear and discontinuous B-cell epitopes, ensures a multi-faceted immune response, addressing the limitations of previous single-epitope vaccines.

However, several limitations remain. The in-silico predictions, while highly informative, require experimental validation through in-vitro and in-vivo studies to confirm immunogenicity, safety, and efficacy. Additionally, the potential for post-translational modifications and glycosylation in the vaccine construct, which cannot be fully accounted for in silico, may influence its final immunogenic profile.

Future Directions

Future studies should focus on:

1. Experimental validation of vaccine immunogenicity and safety in preclinical animal models.
2. Expression and purification of the vaccine construct in bacterial systems.
3. Evaluation of long-term immune memory responses post-vaccination.
4. Assessment of cross-strain immunity against diverse H5N1 viral strains.

Conclusion

The in-silico designed multi-epitope vaccine against H5N1 demonstrates promising immunogenic potential, broad population coverage, and structural stability. The integration of computational tools enabled efficient prediction and validation of critical vaccine parameters. With further experimental validation, this vaccine candidate could represent a significant advancement in avian influenza vaccine development, addressing current challenges in antigenic variability and rapid outbreak response.

References

[1] Alexander, D. J. (2007). An overview of the epidemiology of avian influenza. *Vaccine, 25*(30), 5637-5644.

[2] Swayne, D. E. (2008). Avian influenza. Wiley-Blackwell.

[3] World Health Organization. (2015). Influenza at the human-animal interface.

[4] World Health Organization. (2015). Influenza at the human-animal interface.

[5] Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D., & Fouchier, R. A. (2006). Global patterns of influenza A virus in wild birds. Science, 312(5772), 384-388.

[6] Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., & Kawaoka, Y. (2006). Evolution and ecology of influenza A viruses. Microbiological Reviews, 56(1), 152-179.

[7] Claas, E. C., Osterhaus, A. D., Van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., & Webster, R. G. (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Proceedings of the National Academy of Sciences, 95(11), 6349-6354.

[8] de Jong, M. D., & Hien, T. T. (2006). Avian influenza A (H5N1). Journal of Clinical Virology, 35(1), 2-13.

[9] Naz, A., Awan, F. M., Obaid, A., Muhammad, S. A., Paracha, R. Z., & Ahmad, J. (2015). Immunoinformatics approaches to explore potential epitope-based peptide vaccine targets for human rhinoviruses. International Journal of Peptide Research and Therapeutics, 21(2), 161-173.

[10] Centers for Disease Control and Prevention. (2020). Highly Pathogenic Asian Avian Influenza A(H5N1) in People. Retrieved from https://www.cdc.gov/flu/avianflu/h5n1-people.htm

[11] World Bank. (2006). Enhancing control of highly pathogenic avian influenza in developing countries through compensation: Issues and good practice.

[12] Food and Agriculture Organization. (2011). Approaches to controlling, preventing, and eliminating H5N1 highly pathogenic avian influenza in endemic countries. Retrieved from http://www.fao.org/docrep/014/i2150e/i2150e.pdf

[13] Liu, J., Shi, W., & Gao, G. F. (2014). Poultry viruses: Intermingling of influenza virus and paramyxovirus. Trends in Microbiology, 22(3), 151-159.

[14] Dlugolenski, D., & Baines, J. D. (2011). Avian influenza vaccines: Recent developments in vaccine technology. Expert Review of Vaccines, 10(3), 277-288.

[15] Paules, C. I., & Subbarao, K. (2017). Influenza. The Lancet, 390(10095), 697-708.

[16] Swayne, D. E. (2012). Impact of vaccines and vaccination on global control of avian influenza. Avian Diseases, 56(4s1), 818-828.

[17] Sparrow, E., Friede, M., Sheikh, M., Torvaldsen, S., & Newall, A. T. (2017). Passive immunization for influenza through antibody therapies, a review. Journal of Infectious Diseases, 216(1), S108-S115.

[18] Dormitzer, P. R. (2011). Rapid production of influenza vaccine. Nature Reviews Microbiology, 9(3), 226-231.

[19] Hasan, M., Ghosh, P. P., Azim, K. F., Mukta, S., Abir, R. A., & Nahar, J. (2019). Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. Microbial Pathogenesis, 130, 19-37.

[20] Sette, A., & Rappuoli, R. (2010). Reverse vaccinology: developing vaccines in the era of genomics. Immunity, 33(4), 530-541.

[21] Hasan, M., Ghosh, P. P., Azim, K. F., Mukta, S., Abir, R. A., & Nahar, J. (2019). Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. Microbial Pathogenesis, 130, 19-37.

[22] Yadav, R., Kumar, S., Mahendran, M., Kumar, A., Choudhury, A., & Misra, P. (2020). Immunoinformatics: In silico approaches to design peptide-based epitope vaccine. Biological Procedures Online, 22(1), 1-13.

[23] Tomar, N., & De, R. K. (2014). Immunoinformatics: A brief review. Methods in Molecular Biology, 1184, 23-55.

[24] Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... & Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Research, 47(D1), D339-D343.

[25] UniProt Consortium. (2023). UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Research, 51(D1), D523–D533. https://doi.org/10.1093/nar/gkac1052

[26] Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics, 8(1), 4.

[27] Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., & Marsh, S. G. E. (2015). The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Research, 43(D1), D423–D431. https://doi.org/10.1093/nar/gku1161

[28] Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., & Nielsen, M. (2008). NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. Nucleic Acids Research, 36(suppl_2), W509-W512.

[29] Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., & Buus, S. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics, 61(1), 1-13.

[30] Nilsson JB, Kaabinejadian S, Yari H, Kester MGD, van Balen P, Hildebrand WH, Nielsen M. 2023. Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. Sci Adv. 9(47). Doi: 10.1126/sciadv. adj6367.

[31] Saha, S and Raghava G.P.S. (2006) Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. Proteins. 1; 65(1):40-8.

[32] Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I. AllerTOP v.2 - a server for in silico prediction of allergens. J. Mol. Model., 20, 2278, 2014.

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

[34] Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. 2013. Properties of MHC class I presented peptides that enhance immunogenicity. PloS Comp. Biol. 8(1):361.
PMID: 24204222

[35] Dhanda et. al 2013: Designing of interferon-gamma inducing MHC class-II binders. Biology Direct 2013

[36] Dhanda S.K. et. al 2013: Prediction of Interleukin-4 inducing peptides Clinical and developmental immuonology

[37] Bui H. H,Sidney J, Li W, Fusseder N, Sette A., 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics 8(1):361. PMID: 17897458

[38] Pandey, R. K., et al. (2018). Designing multi-epitope peptide-based vaccine against SARS-CoV-2. Scientific Reports, 10(1), 1-13.

[39] Lafuente, E. M., & Reche, P. A. (2009). Prediction of MHC-peptide binding: A systematic and comprehensive evaluation of binding affinity methods. PLOS Computational Biology, 5(4), e1000435.

[40] Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Letters, 276(1-2), 172-174.

[41] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook (pp. 571–607). Springer. https://doi.org/10.1385/1-59259-890-0:571

[42] Geourjon, C., & Deleage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer Applications in the Biosciences, 11(6), 681–684. https://doi.org/10.1093/bioinformatics/11.6.681

[43] Klausen, M. S., Anderson, M. V., Jespersen, M. C., Nielsen, M., Marcatili, P., & Lund, O. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Nucleic Acids Research, 47(W1), W265–W269. https://doi.org/10.1093/nar/gkz398

[44] Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: Accurate sequence-based prediction of protein solubility. Bioinformatics, 25(17), 2200–2207. https://doi.org/10.1093/bioinformatics/btp386

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

[46] Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP v. 2—A server for in silico prediction of allergens. BMC Bioinformatics, 14(Suppl 6), S4. https://doi.org/10.1186/1471-2105-14-S6-S4

[47] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403-410.

[48] Kelley LA et al. Nature Protocols 10, 845-858 (2015)

[49] Minkyung Baek et al., Accurate prediction of protein structures and interactions using a three-track neural network. Science373,871-876(2021). DOI:10.1126/science. abj8754

[50] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2), 283–291. https://doi.org/10.1107/S0021889892009944

[51] Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking and overall structure optimization. Nucleic Acids Research, 41(W1), W384–W388. https://doi.org/10.1093/nar/gkt458

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

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

[54] Diebold, S. S., et al. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. Proceedings of the National Academy of Sciences, 101(15), 5598–5603.

[55] Dashti, F., et al. A computational approach to design a multiepitope vaccine against H5N1 virus. Virol J 21, 67 (2024). https://doi.org/10.1186/s12985-024-02337-7

[56] Sharp, P. M., & Li, W. H. (1987). The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research, 15(3), 1281–1295. https://doi.org/10.1093/nar/15.3.1281

[57] Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science, 324(5924), 255–258. https://doi.org/10.1126/science.1170160

[58] Moise, L., Terry, F., Gutierrez, A., Tassone, R., Losikoff, P. T., Gregory, S. H., & De Groot, A. S. (2011). H7N9 influenza vaccine safety: Immunoinformatics-guided design. Frontiers in Immunology, 2, 26. https://doi.org/10.3389/fimmu.2011.00026

[59] De Groot, A. S., Moise, L., & Terry, F. (2005). Immunoinformatics: Bioinformatics to guide vaccine design. Methods in Molecular Biology, 409, 335–368. https://doi.org/10.1007/978-1-59745-547-3_24

[60] Khan, F. H., Sayed, S. M., & Minhas, F. U. (2020). Immunoinformatics approaches to vaccine design. Frontiers in Immunology, 11, 589685. https://doi.org/10.3389/fimmu.2020.589685

[61] De Groot, A. S., & Martin, W. (2009). Reducing risk, improving safety: Vaccine design using immunoinformatics. Expert Review of Vaccines, 8(5), 595-605. https://doi.org/10.1586/erv.09.19