

In-Silico Vaccine Construction: Targeting HPV Major Capsid L1 Protein

By

Sadia Shahid Shiba
ID: 19146048

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for
the degree of
Bachelor of Pharmacy

School of Pharmacy
Brac University
February 2024

© 2024. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Sadia Shahid Shiba

ID : 19146048

Approval

The project titled “In-Silico Vaccine Construction: Targeting HPV Major Capsid L1 Protein” submitted by Sadia Shahid Shiba (19146048) of Spring, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

Supervisor:
(Member)

Mohammad Kawsar Sharif Siam
Senior Lecturer, School of Pharmacy
Brac University

Approved By:

Assistant Dean and
Program Director:

Dr. Hasina Yasmin
Professor, Assistant Dean and Program Director,
School of Pharmacy
Brac University

Dean:

Dr. Eva Rahman Kabir
Professor and Dean, School of Pharmacy,
Brac University

Ethics Statement

There were no unethical activities engaged in this thesis. No human or animal trials are used in this research. The thesis was conducted maintaining ethical standards at all regard whatsoever.

Abstract

Cervical cancer is a prevalent outcome of human papillomavirus (HPV) infection, which continues to be a major global health concern. Utilizing virus-like particles (VLPs) based on L1 proteins is one of the methods used to battle HPV and has showed promise in the creation of vaccines. Currently, genotype-restricted protection is provided through commercially available vaccines like Gardasil and Cervarix. However, there are significant obstacles to the equitable distribution of these vaccinations to developing countries, principally because of financial limitations. Therefore, the creation of next-generation high-risk HPV vaccines is urgently needed. In this thorough analysis, we have created DNA constructs, mostly based on the L1 genes, that display significant conservation among high-risk HPV strains and have the potential to be immunogenic. The following fundamental components make up our analytical framework: (1) B-cell epitope mapping; (2) CD4⁺ and CD8⁺ T-cell epitope mapping; (3) allergenicity evaluation; (4) biochemical analysis; (5) molecular docking studies; (6) 3D modeling; and (7) data gathering, analysis, and the creation of L1 and L2 DNA constructs. Additionally, our *in vivo* research has shown that L1 DNA constructs can trigger powerful immune responses when given in combination with the right adjuvants or delivery methods. The DNA structures that we have created are excellent candidates for HPV vaccinations that might provide broader protection against high-risk HPV strains. This study emphasizes the significance of ongoing efforts in the development of novel vaccine methods and marks a significant step towards tackling the global burden of HPV-related diseases.

Keywords: Human papillomavirus; T-cell epitope; B-cell epitope, molecular docking, virtual screening, MD simulation.

Acknowledgement

I want to express my gratitude towards my parents in particular for their unwavering guidance and encouragement, as well as my supervisor Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University, for motivating me to exceed my potential boundaries.

Table of Contents

Declaration.....	ii
Approval	iii
Ethics Statement.....	iv
Abstract/ Executive Summary	v
Acknowledgement	vi
Table of Contents	vii
List of Tables	x
List of Figures.....	xi
List of Acronyms	xiii
Chapter 1 Introduction.....	1
1.1 Structure and Genomic feature of Human Papillomavirus	2
1.2 Life cycle and pathogenesis of Human Papillomavirus.....	3
Chapter 2 Methodology	5
2.1 Overview.....	5
2.2 Protein sequences Retrieval and Prioritization	7
2.3 Identifying Cytotoxic T Lymphocyte (CTL) Epitopes	7
2.4 MHC-I binding prediction	8
2.5 Determination of Helper T Lymphocyte (HTL) Epitopes	8
2.6 Prediction of Cytokine Stimulation of Selected HTL epitopes	9
2.7 Linear B Lymphocyte (LBL) Epitope Prediction and Evaluation.....	9

2.8 Vaccine Construct Design with Adjuvants and Specific Linkers	10
2.9 Biochemical Analysis and Evaluation of Constructed Vaccine	10
2.10 Establishing, Improving, and Evaluating 3D Structures Using Homology Modeling	11
2.11 Vaccine Generating Ramachandran Plots and Z-score.....	12
2.12 Molecular Docking of Vaccine with Relevant Human Receptors.....	13
2.13 Immune Response Simulations.....	13
2.14 Significance of the Methodology.....	14
Chapter 3 Result	15
3.1 Antigenicity of L1 major capsid protein	15
3.2 The characterization of MHC Class I Alleles Distinctive to Cytotoxic T Lymphocyte (CTL) Epitopes	18
3.3 Antigenicity, Allergenicity, and Toxicity Determination of CTL Epitopes	20
3.4 Strong Binding HTL Epitopes Prediction via NetMHCpan 4.0	21
3.5 Cytokine Stimulating Ability of Obtained HTL Epitopes	22
3.6 Determination of B-cell Epitopes	25
3.7 In-silico Biochemical Analysis of Candidate Vaccines.....	28
3.8 Homology modeling	29
3.9 Z-Score and Ramachandran Plots Prediction	32
3.10 Analysis of the Constructed Vaccine Candidate by Molecular Docking with the Relevant Human Receptor	35
3.11 Immune Simulations	39

Chapter 4 Discussion	43
Chapter 5 Conclusion	46
References.....	47

List of Tables

Table 1: CTL epitopes of Protein no: 1 & 2 with their respective combined scores on the NetCTL1.2 server	17
Table 2: Sequence Number, Length, Score_EL, Percentile Rank, and Binding Level of Major Histocompatibility Complex Class I (MHC I) Alleles Particular to Epitopes.....	19
Table 3: Selected CTL epitopes with their antigenicity, allergenicity an toxicity result.....	20
Table 4: HTL epitopes of protein candidate no: 1 with their IFN, IL4 and IL10 inducing capability.....	23
Table 5: HTL epitopes of protein candidate no: 1 with their IFN, IL4 and IL10 inducing capability.....	23
Table 6: Selected HTL epitopes with their antigenicity, allergenicity and toxicity for Protein no: 1 & 2.	25
Table 7: Antigenic B-cell epitopes with their ToxinPred and AllerTOP v2.0 server prediction for both Protein 1 & 2	27
Table 8: Antigenicity, Allergenicity and Toxin Presence Profile of Final Vaccines.....	29
Table 9: Summary of Biochemical analysis and 3D Structure Analysis	35

List of Figure

Figure 1: Stages of in-silico epitope-based vaccine design process	6
Figure 2: Antigenicity Scores of The Two L1 proteins obtained from VaxiJen v2.0 server... 16	
Figure 3: CTL epitopes obtained from NetCTL1.2 output	18
Figure 4: ToxinPred Server Output of Selected CTL epitopes of Two Candidate Proteins....21	
Figure 5: Preview of NetMHCpan 4.0 server output showing strong and weak binding peptides for different alleles	22
Figure 6: B-Cell epitope prediction	26
Figure 7: Diagram showing the connection between a score and position of B-cell epitope for protein no: 1(Jespersen et al., 2017)	26
Figure 8: Diagram showing the connection between a score and position for protein no: 2(Jespersen et al., 2017).....	27
Figure 9: ProtParam Tool Result for Vaccine (i).....	31
Figure 10:ProtParam Tool Result for Vaccine (ii).....	31
Figure 11: Phyre2 result showing homology modeling of vaccine (Kelley et al., 2015)	32
Figure 12: Vaccine (i): (A) The graph of the Z-score against the number of residues and the Z-score on the ProSAweb server.	33
Figure 13: Vaccine (ii): (A) The graph of the Z-score against the number of residues and the Z-score on the ProSAweb server.	33
Figure 14: Ramachandran Plot of Vaccine (i) & (ii)	34
Figure 15: <i>Creating a docked complex in three dimensions between the final vaccine (i) and TLR4 using Discovery Studio.....</i>	<i>37</i>
Figure 16: <i>Creating a docked complex in three dimensions between the final vaccine (ii) and TLR4 using Discovery Studio.</i>	<i>37</i>
Figure 17: Protein-Protein docking ClusProv2.0 Model Score for Vaccine (i) & (ii).....	38

Figure 18: CimmSim Output showing Immune Simulation of Vaccine (i).....41

Figure 19: CimmSim server output for Vaccine (i).....42

List of Acronyms

MHC Major Histocompatibility Complex.

CTL Cytotoxic T lymphocyte.

HTL Helper T lymphocyte

DNA Deoxy-ribonucleic acid.

RNA Ribonucleic acid.

IFN Interferon

Chapter 1

Introduction

Human papillomavirus (HPV)-related sexually transmitted diseases (STIs) provide a serious threat to world health. The most common STI, HPV, has more than 225 varieties divided into five divisions. It might present as oropharyngeal and cutaneous malignancies (high-risk HPV) or cutaneous and anogenital warts (low-risk HPV). While HPV can usually be cleared or suppressed by the immune system, high-risk HPV infections in women, particularly those linked to HPV-16 and HPV-18, can lead to cervical cancer within three to five years. With a focus on their association with cervical cancer, this report offers a concise review of the HPV infections' global effects. Multiple proteins are encoded by the non-encapsulated, compact, double-stranded DNA virus known as HPV, including L1, the major capsid protein. L1 is the main target for recent treatments and vaccines and plays a significant role in host receptor interactions and surface antigenicity.

Vaccination programs are the only preventive measure against HPV-related cancer, but the high cost of commercial vaccines limits their accessibility in low-income countries. Three HPV vaccines that are now on the market have shown to slow the growth and epidermal and anogenital warts and cancers progressing. Bivalent, tetravalent, and nine-valent vaccinations have been approved. Although they target HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58, the available HPV vaccines (such as Gardasil, Gardasil-9, and Cervarix) lack any therapeutic benefits. Due to the high prevalence and fatality rates associated with HPV-linked cervical cancer, the urgent need for an efficient therapeutic or dual-action vaccine is obvious. This study underscores the need to improve HPV vaccines in order to address this serious public health issue. As a result, the existing licensed vaccinations still need to be improved in order to

increase their efficacy, despite having been shown to be effective in lowering the incidence of HPV-positive patients.

The HPV L1 protein, being able to self-assemble into virus-like particles (VLPs), triggers the development of antibodies that neutralize HPV and stop infections. Although effective, VLP-based HPV vaccines are still out of many developing countries' financial reach. Studies have shown that HPV vaccines can induce immune responses that are more effective than the virus itself, highlighting the significance of reducing potential side effects.

Epitope-based vaccines have become more popular as a result of their stability, specificity, safety, and simplicity of production. The selection of effective T-cell and B-cell immunogenic epitopes in antigenic peptides depends heavily on immunoinformatic, a subfield of bioinformatics. Utilizing immunoinformatic software tools speeds up research and lowers the price of synthesized peptides. With an emphasis on epitope-based techniques as a possible route for future vaccine development, this research emphasizes the need to improve HPV vaccination tactics for increased immunogenicity and worldwide accessibility.

1.1 Structure and Genomic feature of Human Papillomavirus

Papillomaviruses are icosahedral DNA viruses, small and non-enveloped, with a diameter of 52 to 55 nm. These viruses are distinguished by their single, roughly 8000 base pair double-stranded DNA molecule, which is firmly attached to cellular histones and encased in a protein capsid made up of 72 pentameric capsomers. Two important structural proteins, late (L)1 (55 kDa, which makes about 80% of the total viral protein) and L2 (70 kDa), are both encoded by the virus and make up the capsid. Viral-like particles (VLPs) can be produced in mammalian or non-mammalian expression systems by expressing L1, either by itself or in combination with L2. The intact virion has a sedimentation coefficient (S_{20, W}) of 300 and a cesium chloride density of 1.34 g/mL. The viral genome is divided into two groups of genes: early

genes (E1, E2, E3, E4, E5, E6, and E7), that regulate the replication of genomes and viral transcription, and late genes (L1 and L2), which encode key structural proteins required for capsid construction. L1 protein is a significant capsid protein of human papillomavirus (HPV) and it stays connected to the viral DNA throughout entrance. It preserves conformation-dependent epitopes identified by neutralizing antibodies and is largely full length. Most likely, it is set up as capsomeres, which are pentamers made up of many L1 molecules. It binds directly using viral DNA in the capsid and exhibits DNA binding activity. The components of the capsid are the major protein of the capsid L1 and the minor capsid protein L2. Notably, the capsid is made up of 360 L1 monomers, which combine to create 72 capsomeres, the pentameric units that make up the capsid shell. Importantly, a single L2 copy can fit into each capsomere, and research indicates that a capsid can hold up to 72 L2 molecules. This thorough analysis explains the papillomaviruses' intricate structural makeup, genomic layout, and host preferences, providing important new information about their biology and posing prospective treatment targets. The HPV virus's encapsidated DNA interacts with the L1 protein, which is enclosed within the capsid. DNA encapsidation depends on the L1 protein's C terminus, which has a propensity to bind DNA. The L1 protein interacts with HPV DNA that has been encapsulated thanks to its C-terminus. Pseudovirus L1, L2, and viral DNA move into the nucleus during infection.

1.2 Life cycle and pathogenesis of Human Papillomavirus

Approximately eight ORFs are present in the genomes of papillomaviruses, which are transcribed from a single DNA strand. The three functional segments of these ORFs are as follows: the late (L) region, which codes for structural proteins (L1–L2) necessary for virion assembly; the early (E) region, which codes for proteins (E1–E7) necessary for viral replication; and the long control region (LCR), a non-coding region that contains cis elements essential for viral DNA replication and transcription. The E1 and E2 proteins function as

markers for the replication origin, and E2 is also the main transcriptional regulator of viral genes. E5 may act in both early and late phases of the virus life cycle, while E4, despite its name, is thought to be engaged in the later stages. The E6 and E7 proteins specifically target p105Rb and p53, two negative cell cycle regulators, to help maintain viral episomes in a stable state and encourage differentiated cells to reenter the S phase. During the formation of progeny virion, the L1 and L2 proteins assemble in capsomers to create icosahedral capsids around the viral DNA. Because they only cause productive infections in the stratified epithelia of the epidermis, anogenital tract, and mouth cavity, papillomaviruses are extremely epitheliotropic. The differentiation of infected epithelial cells, which usually begins in basal epithelial cells and may occur at areas of damage, is associated with the viral life cycle. The E1, E2, E6, and, occasionally, E7 genes are necessary for HPVs to establish their genome in basal cells. Suprabasal cells in HPV-positive cells boost DNA synthesis and show markers for cell proliferation during cell division by failing to exit the cell cycle. In order to initiate suprabasal DNA synthesis, HPV 16 E7 is essential. The assembly of offspring viruses, capsid gene expression, and viral genome amplification are all facilitated by the suprabasal compartment. The cottontail rabbit papillomavirus (CRPV) L1 capsid gene expression and viral DNA amplification depend on the E4 gene. In the terminally differentiated cell compartment, progeny virus encapsidation within capsids is quantitatively dependent on L2, the minor capsid protein. Additionally, L2 is necessary for the infectivity of HPV 16 and HPV 31 virions, which may have an impact on the viral DNA's nuclear localization, intracellular transit, and cell-surface binding.

This study uses immunoinformatic databases from L1 structural proteins to develop a multi-epitope HPV16 vaccination that is secure and efficient. To evaluate the immunological reaction and stability, *in silico* studies, dynamics, and molecular docking were used. This study is a potential step toward developing an enhanced HPV16 vaccination.

Methodology**2.1 Overview**

The basis for creating vaccines with the use of computational and bioinformatics tools has been supplied by molecular biology and information technology advancements, as well as the expansion of genetic data banks (Parvizpour et al., 2020). These methods are used to generate hypothetical proteins and map the best immunogenic components in silico (Parvizpour et al., 2020). Simulating and assessing the planned vaccine before undertaking any experimental validation offers a route for effective research and development (R&D).

By recognizing B-Cells and T-Cells, epitope vaccines seek to elicit immune responses (Parvizpour et al., 2020). For subunit vaccines to effectively elicit humoral as well as cellular immune responses, the selection of suitable B-cell epitopes (BCEs) and T-cell epitopes (TCEs) is essential (Parvizpour et al., 2020). For an effective vaccine, these epitopes must be correctly linked with adjuvants (Parvizpour et al., 2020). The right combination of these elements is essential for the best immunological stimulation. Due to the absence of potentially harmful components like allergens and toxins, epitope-based vaccinations are regarded as safe (Parvizpour et al., 2020). The steps in the in silico epitope-based vaccine construction process are visualized in Figure 1. The creation of epitope vaccines and important immunological variables are briefly discussed in this work.

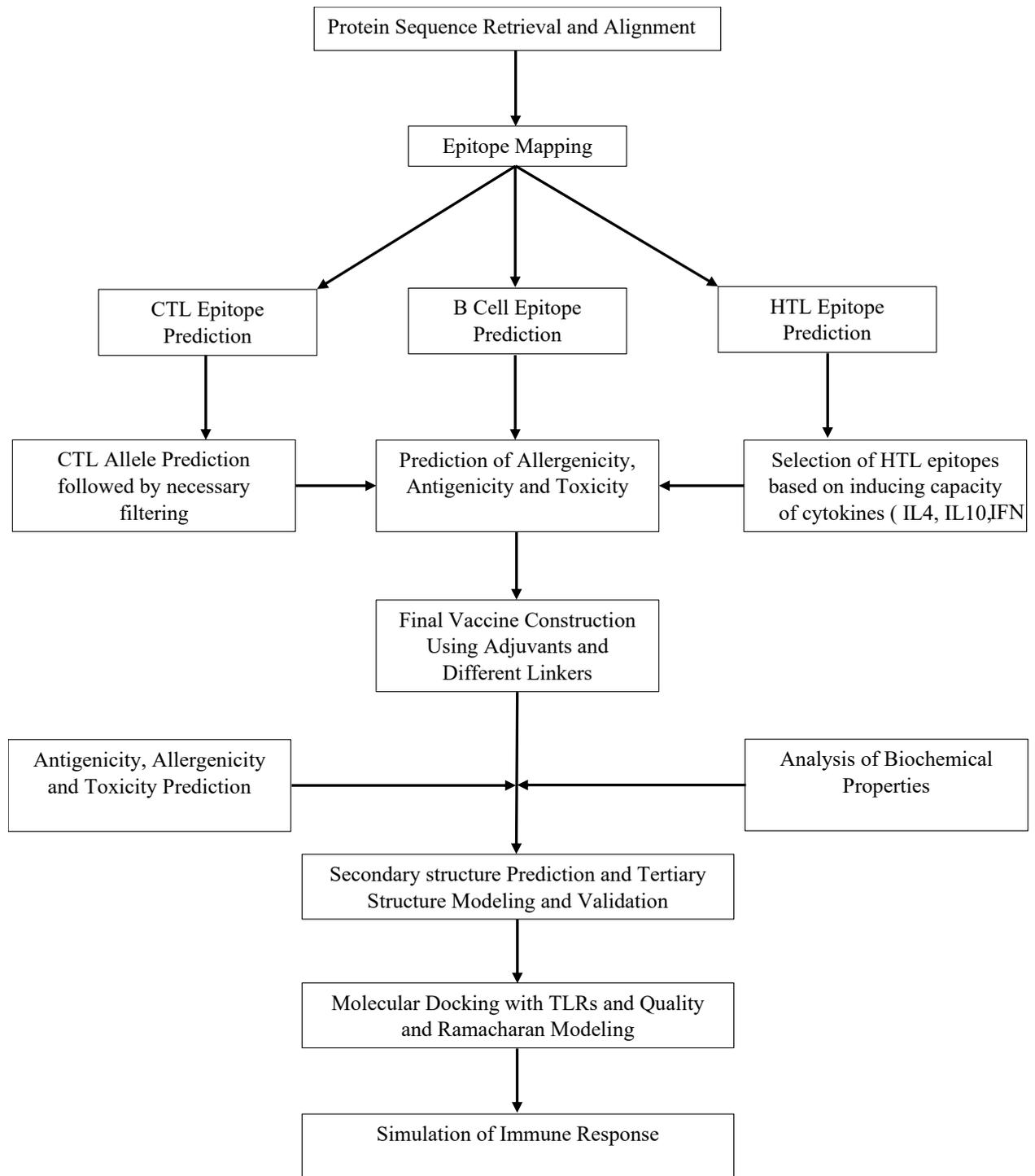


Figure 1: Stages of in-silico epitope-based vaccine design process

2.2 Protein Sequences Retrieval and Prioritization

NCBI was used to get the protein sequences of high-risk HPV 16 L1 proteins. With "virus" as the chosen organism, analysis was carried out in Vaxijen v2.0 and a constant 0.5 threshold is used. By using the autocross covariance (ACC) approach, the latter server achieves a prediction accuracy of 70–89% (Doytchinova & Flower, 2007). Since higher antigenicity scores are necessary to maintain a respectable antigenic score while conserved sequences are produced, few L1 protein sequences were chosen based on their antigenicity scores.

2.3 Identifying CTL Epitopes

According to Maleki et al.'s description from 2021, three chosen proteins underwent MHC class I epitope detection utilizing the NetCTL 1.2 server. The strategy includes evaluation of effectiveness of the transporter linked to antigen processing (TAP), prediction of peptide MHC class I binding, also proteasomal C-terminal cleavage. The website uses artificial neural networks to predict CTL epitopes for 12 MHC class I super types based on MHC binding and proteasomal cleavage analysis. A weight matrix is used to forecast the TAP transport efficiency. The C-terminal cleavage threshold of 0.15, the TAP transport efficiency of 0.05, and the epitope identification threshold of 0.75 are the parameters established by Maleki et al. (2002). The overall results are organized and evaluated to demonstrate the process of identifying MHC class I epitopes via using NetCTL 1.2 server.

2.4 MHC-I Binding Prediction

Using Egyptian HLA haplotypes that are often seen (HLA-A0101, HLA-A0210, HLA-B03401, and HLA-B4101), the NetMHCpan 4.1 server predicted the binding of the virus. With a threshold of two for weak binders and a 0.5% score for strong binders, the study focused on the HPV16 strain L1 protein (Awad et al., 2022). There were BA forecasts listed. In addition to eluted ligands from mass spectrometry, NetMHCpan4.1 also uses artificial neural networks for prediction that have been trained on multiple quantitative binding affinities (Awad et al., 2022). Only the strong binders with their accompanying HLA haplotypes made it into the final epitopes, which were then screened (Awad et al., 2022). Then, for each suggested epitope that had been predicted before, Vaxijen assessed the antigenicity response. A likely antigen threshold of 0.5 is used to implement Vaxijen (Awad et al., 2022). The optimum prediction threshold for the antigenicity response of the epitopes was chosen to be 0.5. Additionally, it was previously reported that this score served to confirm the antigenicity response of the predicted epitopes (Awad et al., 2022). ToxinPred (<http://crdd.osdd.net/raghava/toxinpred/>) were used to detect toxicity and AllerTop 2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) servers were used to estimate allergic response of the suggested epitopes, respectively.

2.5 Determination of HTL Epitopes

The IEDB MHC-II binding tool was used by Mohammadi et al. (2023) to determine the 15-mer HTL epitopes for eight MHC class II supertype alleles. Using artificial neural networks, the NetMHCpan4.0 service predicted sequence-based peptide binding to certain MHC molecules. More than 180,000 mass spectrometry binding observations and ligands, representing 172 MHC molecules across several species, were included in the training

dataset.. As they have a low adjusted rank, epitopes were chosen because they are effective binders (Mohammadi et al., 2023). With VaxiJen v2.0 and ToxinPred, the chosen epitopes were evaluated for toxicity and antigenicity. Then, in the research carried out by Mohammadi et al. (2023), the AllerTOP 2.0 program evaluated the allergenicity of the non-toxic and antigenic epitopes.

2.6 Identification of Cytokine Stimulation of Selected HTL epitopes

We evaluated three important cytokines that are involved with the epitope selection procedure methodically. The IFNepitope platform evaluated interferon (IFN) (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>) using a hybrid motif as well as support vector machine (SVM) approach, and the IL4pred tool (<https://webs.iitd.edu.in/raghava/il4pred/design.php>) evaluated interleukin (IL)-4 inducibility using motif analysis. Furthermore, the IL10pred server (<https://webs.iitd.edu.in/raghava/il10pred/design.php>) was used to investigate IL-10 inducibility. In order to construct the final vaccine, certain epitopes that satisfied strict requirements were included, such as significant antigenic potential, non-allergenicity, non-toxic properties, and inducibility of each of those cytokines (Mohammadi et al., 2023). This succinct synopsis describes the rigorous epitope selection procedure that guarantees the most effective candidates are included in the final vaccine.

2.7 Evaluation and Prediction of Linear B Lymphocyte (LBL) Epitopes

In this study, we used the Bepipred linear epitope prediction 2.0 tool (<http://tools.iedb.org/bcell/>) to predict Linear B Lymphocyte (LBL) epitopes, using the random forest technique with a default threshold of 0.5. Using a 0.5 threshold, the VaxiJen

v2.0 webservice was used to determine the antigenic value of the identified LBL epitopes. The antigenic epitopes were then evaluated using ToxinPred for toxicity and AllerTOP 2.0 for allergenicity (Mohammadi et al., 2023).

2.8 Vaccine Construct Design with Adjuvants and Specific Linkers

The vaccine design was put together using the recognized CTL and HTL epitopes, adding the right linkers to assure epitope separation and versatility of residues of amino acids for optimal bending (Sanami et al., 2022). Adjuvant sequences were associated with particular B and T cell epitopes, and the adjuvant and PADRE motifs were linked via the EAAAK linker (Kolla et al., 2021). CTL, HTL, and B cell epitopes were connected using linkers, specifically AAY, GPGPG, and KK (Kolla et al., 2021). In multi-epitope vaccine design, GPGPG, AAY, and KK are often utilized linkers for epitope mapping with the goal of enhancing pathogen-specific immunity and preventing junctional immunogenicity (Kolla et al., 2021).

2.9 Biochemical Analysis and Evaluation of Constructed Vaccine

The physical and chemical characteristics of the developed vaccine were examined using the ProtParam analysis tool (<https://web.expasy.org/protparam/>). Elshafei et al. (2022) computed a number of measures, including theoretical isoelectric point (pI), molecular weight, instability index, amino acid content, as well as Grand Average of Hydropathicity Index (GRAVY). Average the molecular masses of the amino acids in a protein to find its molecular weight. A typical amino acid has a molecular weight of about 110 daltons (Da). Important for methods such as isoelectric focused electrophoresis along with ion exchange chromatography, the isoelectric point (pI) refers to the pH at which a protein bears no net

charge. Under test-tube circumstances, proteins having an instability index of less than 40 are deemed stable. Thermal stability is measured by the aliphatic index, which reflects the fraction of aliphatic side chains. A protein sequence's GRAVY value is computed by dividing the total sum of the hydrophathy values of its amino acids by the amount of residues. The antigenicity of the vaccine was predicted using VaxiJen v2.0, and its allergenicity was evaluated by AllergenOnline. In the study conducted by Sanami et al., the T3DB server was utilized to forecast the vaccine's toxicity.

2.10 Establishing, Improving, and Evaluating 3D Structures Using Homology Modeling

PHYRE2 (Protein Homology/analogy Recognition Engine version 2) from Imperial College London, which may be found at (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>), was used to create the protein's tertiary structure (Al-Khayyat; et al., 2016). The final vaccine protein sequence was entered in plain format with email address for each constructed vaccine in the server. To choose the best model and evaluate its quality, the revised models were subjected to many validation tests (Al-Khayyat; et al., 2016). The image in the Phyre2 result is a representation of the model created for your sequence using that template. The coordinates of the model in PDB format may be downloaded by clicking the image and then entered into any other viewing or analysis software you may have. 'Confidence' is the column after that. This shows the likelihood (on a scale of 0 to 100) that your sequence and this template are indeed homologous. Despite the fact that the two are closely connected, it does not indicate the predicted accuracy of the model. The 'Normal modeling' mode was selected in Phyre 2. For each vaccination, this site sent a PDB file to the specified email address.

The PDB data were opened using the program Discovery Studio Visualizer, which also allowed for interactive 3D structure visualization.

2.11 Vaccine Generating Ramachandran Plots and Z-score

Ramachandran plots, local and Z-scores were used to validate HPV protein 3D models. The Ramachandran plot was generated using SWISS-MODEL (<https://swissmodel.expasy.org/assess/help>) to analyze the conformational, stereochemical, as well as quality of structure of the 3D modeled HPV structural as well as accessory proteins (Elalouf, 2023). The prominent sections of the Ramachandran plot, which include almost 90% of the residues in amino acids, demonstrate the high standard of the three-dimensional modeled structure (Elalouf, 2023).

Whether protein structure models were obtained experimentally or computationally (for example, via homology modeling or ab initio prediction), the Ramachandran plot has given a strong validation check. Its main advantage is that, if the structure has been developed without any constraints, it is a tough metric to artificially 'fix'. By deliberately following a protein chain backwards via the electron density and then improving the backwards model until acceptable quality indicators were found, this was demonstrated. The Ramachandran plot refused to give in, showing numerous residues in the forbidden zones and showing none of the typical clustering of residues in the most advantageous regions, but several quality metrics could be modified to reasonable-looking levels. In the ProSa-web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) the PDB file obtained from the Phyre2 server was uploaded. ProSA-web requires the atomic coordinates of the model to be evaluated. The model's atomic coordinates must be assessed for the ProSA-web. The z-score evaluates the departure of the structure's total energy from an energy distribution calculated from random confirmations and represents the general model quality. Z-scores

that fall outside of a typical range for natural proteins signify incorrect structures. Its specific value is shown in a plot that includes the z-scores of all experimentally determined protein chains in order to make it easier to comprehend the z-score of the given protein.

2.12 Molecular Docking of Vaccine with Relevant Human Receptors

TLR4 (4g8a) was docked with the multi-epitope vaccine construct containing the specified epitopes (Rahman et al., 2020). From Protein Databank (PDB) (<https://www.rcsb.org>), the TLR4 complex (TLR4 PDB ID:4g8a) was downloaded. For molecular docking and docking refinement, the online services ClusPro server were utilized, respectively (Rahman et al., 2020). In order to address critical issues in vaccine development, this paper includes the ClusPro operational framework, a well-known protein-ligand docking service (Kozakov et al., 2017). ClusPro makes use of a sophisticated algorithm built on the PIPER core to make it easier to explore ligand conformations and their interactions with receptor proteins (Kozakov et al., 2017). For rotational and translational exploration of ligands, the server uses a grid-based sampling approach, with rotational space being effectively sampled on a sphere-based grid (Kozakov et al., 2017). This leads to the evaluation of a wide variety of conformations for normal proteins, numbering in the region of 10^9 - 10^{10} (Kozakov et al., 2017).

2.13 Immune Response Simulations

Through the web server C-ImmSim (<http://kraken.iac.rm.cnr.it/C-IMMSIM/>), immune simulation of the vaccine design was carried out. It is an agent-based model implementation that gathers data on the humoral and cellular responses of the mammalian immune system that are triggered by antigen at the cellular level (Rahman et al., 2020). After signing in as

a temporary user, the vaccination sequence and three doses were entered into the C-ImmSim server. According to Kaba et al. (2018), 8 h corresponds to one cell division cycle in real life. The time steps of injection numbers 1, 2, and 3 were 1, 84, and 168, respectively. Each time step corresponds to a specified number of hours. Three injections spaced out over the course of 28 days were the recommended dosing schedule for the immunization. Three injections spaced out over the course of 28 days were the recommended dosing schedule for the immunization. The number of simulation steps was 300. The comparison between the construct and the positive control was made, and the findings were evaluated.

2.14 Significance of the Methodology

In genetics, biotechnology, and molecular biology, bioinformatics tools are used to organize and store biological data (Sunita et al., 2020). Prior to lab experimentation, the use of computational tools is more advantageous because they are cost effective and take less time to operate (Sunita et al., 2020). These instruments are entirely based on statistical and machine learning systems, and they have a solid reputation for analyzing and modeling molecular interactions that occur during antigen presentation and processing (Sunita et al., 2020). This strategy is helpful since it involves both identifying the proteins that might potentially serve as antigens and studying the pathogen's whole genome (Sunita et al., 2020). This enables flexible analysis that is not possible with conventional techniques (Sunita et al., 2020)

Chapter 3

Result

3.1 Antigenicity of L1 Major Capsid Protein

Using an amino acid size screening method, the first set of proteins for this study were obtained from NCBI Protein Database, which may be accessed at <https://www.ncbi.nlm.nih.gov/protein>. The VaxiJen v2.0 server found that two distinct variations of HPV L1 major capsid protein had the greatest levels of antigenicity among the primary proteins evaluated. AAM74159.1 and AEA76067.1 are the GenBank Sequence Accession Numbers used to identify these two protein variations. For work purpose, protein having AAM74159.1 as its sequence accession was labelled as “Protein no: 1” and protein having AEA76067.1 as its sequence accession was labelled as “Protein no: 2”. The two L1 major capsid protein's amino acid sequence are listed below,

Protein no: 1: (419 amino acids)

MQVTFIYILVITCYENDVNVYHIFQMSLWLPSEATVYLPVPVSKVVSTDEYVAR
TNIYYHAGTSRLLAVGHPYFPIKKNKILVPKVSGLQYRVFRIHLDPNKFQFSDT
SFYNPDTQRLVWACVGVEVGRGQPLGVGISGHPLLKDDTENASVYAANAGVD
NRECISMDYKQTQLCLIGCKPPIGEHWGKGSPTNVAVNPGDCPPLELINTVIQDG
DMVHTGFGAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDSLFFYLRRQ
MFVRHLFNRAVAVGENVPDDLYIKGSGSTANLASSNYFPTPSGSMVTSDAQIFNKP
YWLQRAQGHNNGICWGNQLFVTVVDTRSTNMSLCAAISTSETTYKNTNFKEYL
RHGEEYDLQFIFQLCKITLTADVTTYIHSM

Protein no.2: (439 amino acids)

TDEYVARTNIYYHAGTSRLLAVGHPYFPIKKNKILVPKVSGLQYRVFRIHLDP
PNKFGFPDTSFYNPDTQRLVWACVGVEVGRGQPLGVGISGHPLLKDDTENASA
YAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKGSPTNVAVNPGDCPPLE
LINTVIQDGDMDVDTGFGAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDS

LFFYLRREQMFVRHLFNRAGAVGENVPDDL YIKGSGSTANLASSNYFPTPSGSMV
TSDAQIFNKPYWLQRAQGHNNGICWGNQLFVTVVDTTRSTNMSLCAAISTSETTY
KNTNFKEYLRHGEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPG
GTLEDTYRFVTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKEKFSADLDQF

The aforementioned L1 major capsid protein variations produced antigenicity scores of 0.5209 and 0.4912 using the VaxiJen v2.0 server, a computational method for antigenicity prediction. These results support the potential importance of these proteins in the context of HPV-related research and indicate a high chance of antigenicity, as shown graphically in Figure 2 of this article.



Figure 2: Antigenicity Scores of The Two L1 proteins obtained from VaxiJen v2.0 server

According to the methods described by Larsen et al. in 2007, it is concentrated on the identification of cytotoxic T lymphocyte (CTL) epitopes utilizing the MHC supertype A1 and a strict threshold of 0.75. A total of 17 epitopes of Protein no: 1 and 15 epitopes for Protein no: 2 with the potential to elicit CTL responses were found as a result of this screening procedure; these epitopes are described in detail in Figure 3 of this paper. The study has also tabulated the combined scores linked to these discovered CTL epitopes, and Table 1 shows the full tabulation of these values.

Table 1: CTL epitopes of Protein no: 1 & 2 with their respective combined scores on the NetCTL1.2 server

Protein no: 1		Protein no: 2	
CTL Epitopes	Combined Scores	CTL Epitopes	Combined Scores
YVARTNIYY	2.385	YVARTNIYY	2.385
AISTSETTY	1.4902	AISTSETTY	1.4902
TANLASSNY	1.4077	TANLASSNY	1.4077
WLPSEATVY	1.2886	TLTADVMTY	1.3605
TLTADVTTY	1.2556	TSICKYPDY	1.1836
STDEYVART	1.2305	YIKMVSEPY	0.9762
TSICKYPDY	1.1836	TSDAQIFNK	0.9699
YIKMVSEPY	0.9762	DICTSICKY	0.9529
TSDAQIFNK	0.9699	NRECISMDY	0.9044
DICTSICKY	0.9529	LTADVMTYI	0.899
LTADVTTYI	0.9383	MVDTGFGAM	0.8822
CYENDVNVY	0.9377	STILEDWNF	0.8812
NRECISMDY	0.9044	ETTYKNTNF	0.8053
ETTYKNTNF	0.8053	RLLAVGHPY	0.795
RLLAVGHPY	0.784	YKNTNFKEY	0.7741
YKNTNFKEY	0.7741	-	-
HTGFGAMDF	0.7522	-	-

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

Protein no: 1

53	ID	fasta	pep	YVARTNIYY	aff	0.4948	aff_rescale	2.1010	cle	0.9284	tap	2.8930	COMB	2.3850	<-E
372	ID	fasta	pep	AISTSETTY	aff	0.2774	aff_rescale	1.1778	cle	0.9672	tap	3.3470	COMB	1.4902	<-E
308	ID	fasta	pep	TANLASSNY	aff	0.2641	aff_rescale	1.1212	cle	0.9384	tap	2.9160	COMB	1.4077	<-E
30	ID	fasta	pep	WLPSEATVY	aff	0.2374	aff_rescale	1.0078	cle	0.8886	tap	2.9500	COMB	1.2886	<-E
407	ID	fasta	pep	TLTADVTTY	aff	0.2272	aff_rescale	0.9647	cle	0.9745	tap	2.8960	COMB	1.2556	<-E
49	ID	fasta	pep	STDEYVART	aff	0.2915	aff_rescale	1.2379	cle	0.2183	tap	-0.8030	COMB	1.2305	<-E
251	ID	fasta	pep	TSICKYPDY	aff	0.2351	aff_rescale	0.9981	cle	0.2350	tap	3.0060	COMB	1.1836	<-E
259	ID	fasta	pep	YIKMVSEPY	aff	0.1665	aff_rescale	0.7068	cle	0.8320	tap	2.8920	COMB	0.9762	<-E
326	ID	fasta	pep	TSDAQIFNK	aff	0.2078	aff_rescale	0.8821	cle	0.4669	tap	0.3550	COMB	0.9699	<-E
248	ID	fasta	pep	DICTSICKY	aff	0.1750	aff_rescale	0.7432	cle	0.4665	tap	2.7940	COMB	0.9529	<-E
408	ID	fasta	pep	LTADVTTYI	aff	0.1820	aff_rescale	0.7729	cle	0.9214	tap	0.5440	COMB	0.9383	<-E
13	ID	fasta	pep	CYENDVNVY	aff	0.1535	aff_rescale	0.6518	cle	0.9014	tap	3.0140	COMB	0.9377	<-E
168	ID	fasta	pep	NRECISMDY	aff	0.1446	aff_rescale	0.6138	cle	0.9757	tap	2.8850	COMB	0.9044	<-E
377	ID	fasta	pep	ETTYKNTNF	aff	0.1311	aff_rescale	0.5565	cle	0.8833	tap	2.3270	COMB	0.8053	<-E
67	ID	fasta	pep	RLLAVGHPY	aff	0.1163	aff_rescale	0.4937	cle	0.8485	tap	3.2600	COMB	0.7840	<-E
380	ID	fasta	pep	YKNTNFKEY	aff	0.1145	aff_rescale	0.4862	cle	0.9628	tap	2.8700	COMB	0.7741	<-E
227	ID	fasta	pep	HTGFGAMDF	aff	0.1275	aff_rescale	0.5414	cle	0.6368	tap	2.3060	COMB	0.7522	<-E
328	ID	fasta	pep	DAQIFNKPY	aff	0.1114	aff_rescale	0.4732	cle	0.9011	tap	2.6830	COMB	0.7425	<-E

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000 Protein no: 2

4	ID	fasta	pep	YVARTNIYY	aff	0.4948	aff_rescale	2.1010	cle	0.9284	tap	2.8930	COMB	2.3850	<-E
324	ID	fasta	pep	AISTSETTY	aff	0.2774	aff_rescale	1.1778	cle	0.9672	tap	3.3470	COMB	1.4902	<-E
260	ID	fasta	pep	TANLASSNY	aff	0.2641	aff_rescale	1.1212	cle	0.9384	tap	2.9160	COMB	1.4077	<-E
359	ID	fasta	pep	TLTADVMTY	aff	0.2520	aff_rescale	1.0698	cle	0.9730	tap	2.8960	COMB	1.3605	<-E
203	ID	fasta	pep	TSICKYPDY	aff	0.2351	aff_rescale	0.9981	cle	0.2350	tap	3.0060	COMB	1.1836	<-E
211	ID	fasta	pep	YIKMVSEPY	aff	0.1665	aff_rescale	0.7068	cle	0.8320	tap	2.8920	COMB	0.9762	<-E
278	ID	fasta	pep	TSDAQIFNK	aff	0.2078	aff_rescale	0.8821	cle	0.4669	tap	0.3550	COMB	0.9699	<-E
200	ID	fasta	pep	DICTSICKY	aff	0.1750	aff_rescale	0.7432	cle	0.4665	tap	2.7940	COMB	0.9529	<-E
120	ID	fasta	pep	NRECISMDY	aff	0.1446	aff_rescale	0.6138	cle	0.9757	tap	2.8850	COMB	0.9044	<-E
360	ID	fasta	pep	LTADVMTYI	aff	0.1952	aff_rescale	0.8289	cle	0.2865	tap	0.5440	COMB	0.8990	<-E
177	ID	fasta	pep	MVDTGFGAM	aff	0.1720	aff_rescale	0.7304	cle	0.9539	tap	0.1750	COMB	0.8822	<-E
373	ID	fasta	pep	STILEDWNF	aff	0.1429	aff_rescale	0.6068	cle	0.9044	tap	2.7760	COMB	0.8812	<-E
329	ID	fasta	pep	ETTYKNTNF	aff	0.1311	aff_rescale	0.5565	cle	0.8833	tap	2.3270	COMB	0.8053	<-E
18	ID	fasta	pep	LLAVGHPY	aff	0.1163	aff_rescale	0.4937	cle	0.9221	tap	3.2600	COMB	0.7950	<-E
332	ID	fasta	pep	YKNTNFKEY	aff	0.1145	aff_rescale	0.4862	cle	0.9628	tap	2.8700	COMB	0.7741	<-E
421	ID	fasta	pep	YTFWEVNLK	aff	0.1369	aff_rescale	0.5813	cle	0.8920	tap	0.6360	COMB	0.7469	<-E

Figure 3: CTL epitopes obtained from NetCTL1.2 output

3.2 The Characterization of MHC Class I Alleles Distinctive to Cytotoxic T Lymphocyte (CTL) Epitopes

In this study, the use of multi-step procedure was utilized to locate CTL (cytotoxic T-lymphocyte) epitopes. Prior to collecting the appropriate alleles for these supertypes from the NetMHC Pan 4.1 server for further CTL epitope analysis, first receiving of 17 A1 supertypes for Protein No. 1 and 15 A1 supertypes for Protein No. 2 from the NetCTL-1.2 server was done. Lower scores showed greater binding affinities, which encouraged epitope selection. This was a key factor in our method of epitope selection. We set a standard of 2.0 as the minimal criterion to ensure a rigorous screening procedure. Epitopes that exceeded this threshold undergo additional analysis. According to their binding affinities, we systematically divided epitopes into two groups: strong binding peptides with a threshold of 0.500 or lower, indicating robust binding affinity to the chosen MHC I alleles, and weak binding peptides with a threshold at or above 2.000, indicating relatively weaker binding affinity against the selected MHC I alleles. We were able to distinguish between strong and weak binders among the discovered CTL epitopes using this methodical methodology, which gave us important information about their immunological applicability to our study goals.

Table 2: Sequence Number, Length, Score_EL, Percentile Rank, and Binding Level of Major Histocompatibility Complex Class I (MHC I) Alleles Particular to Epitopes

Protein no: 1					
Allele Peptide	Sequence no.	Score EL	%Rank_EL	%Rank_BA	Aff(nM)
HLA-A*01:01	LTADVTTY	0.436111	0.356	0.681	1745.1
HLA-A*01:01		0.436111	0.356	0.681	1745.1
HLA-A*01:01	AISTSETTY	0.368145	0.435	1.571	5293.94
HLA-B*15:01		0.816503	0.101	1.147	270.64
HLA-A*01:01	TANLASSNY	0.427554	0.365	0.763	2025.92
HLA-A*26:01		0.24804	0.446	1.218	4261.75
HLA-A*01:01	YVARTNIYY	0.770331	0.117	0.09	106.69
HLA-A*26:01		0.812479	0.032	0.04	72.58
HLA-B*15:01	YVARTNIYY	0.703699	0.205	0.236	43.58
HLA-A*26:01	DICTSICKY	0.386237	0.252	0.342	987.08
HLA-A*26:01	ETTYKNTNF	0.550311	0.136	0.224	569.01
HLA-A*26:01	TLTADVTTY	0.246793	0.449	1.41	4957.37
HLA-B*15:01		0.715776	0.19	1.04	230.16
HLA-A*26:01	YIKMVSEPY	0.333729	0.308	0.195	462.17
HLA-B*15:01	RLLAVGHPY	0.772606	0.139	0.142	26.57
HLA-B*15:01	WLPSEATVY	0.515236	0.438	0.785	153.31
HLA-B*15:01	YIKMVSEPY	0.799254	0.115	0.019	9.95
Protein no: 2					
Allele Peptide	Sequence no.	Score EL	%Rank_EL	%Rank_BA	Aff(nM)
HLA-A*01:01	YVARTNIYY	0.770331	0.117	0.09	106.69
HLA-A*26:01		0.812479	0.032	0.04	72.58
HLA-B*15:01	YVARTNIYY	0.703699	0.205	0.236	43.58
HLA-A*01:01	AISTSETTY	0.368145	0.435	1.571	5293.94
HLA-B*15:01		0.816503	0.101	1.147	270.64
HLA-A*01:01	TLTADVMTY	0.364199	0.441	0.815	2215.42
HLA-A*26:01		0.251062	0.44	1.056	3652.06
HLA-B*15:01	TLTADVMTY	0.632621	0.283	0.946	200.01
HLA-A*01:01	TANLASSNY	0.427554	0.365	0.763	2025.92
HLA-A*26:01		0.24804	0.446	1.218	4261.75
HLA-A*01:01	LTADVMTY	0.41007	0.382	0.496	1104.85

HLA-A*26:01	YIKMVSEPY	0.333729	0.308	0.195	462.17
HLA-B*15:01	YIKMVSEPY	0.799254	0.115	0.019	9.95
HLA-A*26:01	DICTSICKY	0.386237	0.252	0.342	987.08
HLA-A*26:01	ETTYKNTNF	0.550311	0.136	0.224	569.01
HLA-B*15:01	RLLAVGHPY	0.772606	0.139	0.142	26.57
HLA-B*58:01	STILEDWNF	0.579572	0.404	0.503	107.06

3.3 Antigenicity, Allergenicity, and Toxicity Determination of Selected CTL Epitopes

Five epitopes with antigenic characteristics for Protein no: 1 and six epitopes with antigenic characteristics for Protein no: 2 were eventually found in this work using VaxiJen v2.0, which was used to predict the immunogenicity of all detected epitopes (Table 3). All the antigenic CTL epitopes were found to be non-toxic when ToxinPred, an SVM-based technique, was used to evaluate the toxicity of the CTL epitopes. Additionally, it is used to thoroughly assess several physicochemical characteristics for the discovered CTL epitopes, such as toxicity, hydrophobicity, hydrophaticity, hydrophilicity, molecular weight, and charge (Figure 4). Moreover, for the first protein, one of the antigenic epitopes were found to be non-allergic and two of the antigenic epitopes of protein no: 2 were found to be non-allergic shown in Table 3. Allergenicity was predicted by using AllerTOP v2.0 server.

Table 3: Selected CTL epitopes with their antigenicity, allergenicity an toxicity result

Protein No: 1			
CTL epitopes	Antigenicity	Toxicity	Allergenicity
AISTSETTY	0.7960 (Probable ANTIGEN)	Non-Toxin	Allergen
ETTYKNTNF	1.6024 (Probable ANTIGEN)	Non-Toxin	Allergen
RLLAVGHPY	0.7359 (Probable ANTIGEN)	Non-Toxin	Allergen
TANLASSNY	0.6186 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
YIKMVSEPY	0.5855 (Probable ANTIGEN)	Non-Toxin	Allergen
YVARTNIYY	Non-Antigen	Non-Toxin	-
TLTADVTTY	Non-Antigen	Non-Toxin	-
DICTSICKY	Non-Antigen	Non-Toxin	-
LTADVTTY	Non-Antigen	Non-Toxin	-
WLPSEATVY	Non-Antigen	Non-Toxin	-

With the help of ToxinPred server, toxicity of the selected antigenic CTL epitopes of the two best candidate L1 proteins were analyzed (Figure 4).

Query Peptides								Protein No: 1
Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydrophobicity	Hydrophilicity	Charge	Mol wt
seq1	AISTSETTY	-0.99	Non-Toxin	-0.08	-0.24	-0.24	-1.00	972.13
seq1	DICTSIQKY	-0.74	Non-Toxin	-0.08	0.42	-0.22	0.00	1045.35
seq1	ETTYKNTNF	-0.69	Non-Toxin	-0.32	-1.67	0.04	0.00	1117.30
seq1	RLAVGHPY	-0.99	Non-Toxin	-0.02	0.29	-0.60	1.50	1025.35
seq1	TANLASSNY	-0.67	Non-Toxin	-0.10	-0.36	-0.50	0.00	940.09
seq1	TLTADVTTY	-0.88	Non-Toxin	-0.01	0.24	-0.52	-1.00	984.19
seq1	WLPSEATVY	-1.17	Non-Toxin	0.06	0.11	-0.73	-1.00	1065.31
seq1	YIKMVSEPY	-0.43	Non-Toxin	-0.05	-0.20	-0.32	0.00	1129.46

Query Peptides								Protein No: 2
Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydrophobicity	Hydrophilicity	Charge	Mol wt
seq1	LTADVMTY	-0.84	Non-Toxin	0.07	0.69	-0.65	-1.00	913.16
seq1	AISTSETTY	-0.99	Non-Toxin	-0.08	-0.24	-0.24	-1.00	972.13
seq1	DICTSIQKY	-0.74	Non-Toxin	-0.08	0.42	-0.22	0.00	1045.35
seq1	ETTYKNTNF	-0.69	Non-Toxin	-0.32	-1.67	0.04	0.00	1117.30
seq1	RLAVGHPY	-0.99	Non-Toxin	-0.02	0.29	-0.60	1.50	1025.35
seq1	STILEDVWF	-1.15	Non-Toxin	-0.02	-0.20	-0.38	-2.00	1124.34
seq1	TANLASSNY	-0.67	Non-Toxin	-0.10	-0.36	-0.50	0.00	940.09
seq1	TLTADVMTY	-0.96	Non-Toxin	0.04	0.53	-0.62	-1.00	1014.28
seq1	YIKMVSEPY	-0.43	Non-Toxin	-0.05	-0.20	-0.32	0.00	1129.46

Figure 4: ToxinPred Server Output of Selected CTL epitopes of Two Candidate Proteins

3.4 Strong Binding HTL Epitopes Prediction via NetMHCpan 4.0

The current research included the prediction of Helper T-Lymphocyte (HTL) protein epitopes with a focus on MHC class II epitopes (Andongma et al, 2023). Epitopes made up of 15-mer sequences were found in various alleles, such as, HLA_DRB, HLA_DP, HLA_DQ, and H-2-I (Andongma et al, 2023). A threshold value of 1 denoted strong binding epitopes, whereas a threshold score of 5 denoted weak binders. Additionally, a crucial metric known as Binding Affinity (BA) prediction was included. The core sequences of the peptides, as well as each one's score_EL and percentile rank, were included in the results presentation (Andongma et al, 2023). Strong binding peptides were at the top of the list in the precisely arranged output from the NetMHCIIpan 4.0 server, which was based on prediction scores. A total of 49 strong binding (SB) HTL epitopes were found in the first

protein sequence (Protein no. 1), whereas 48 SB HTL epitopes were found in the second protein sequence (Protein no. 2).

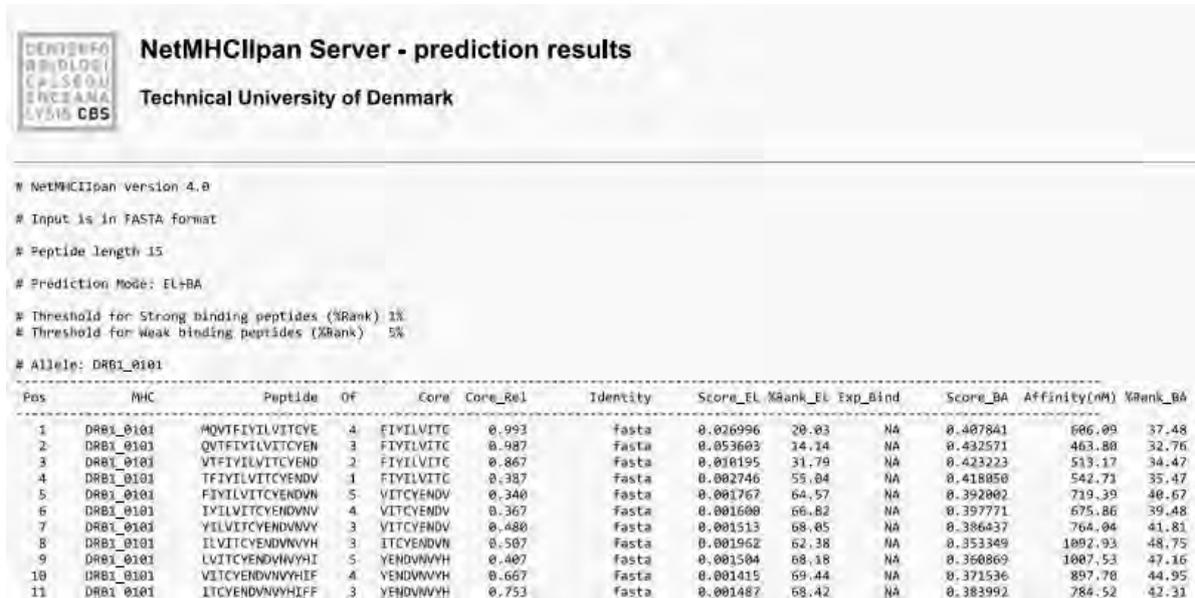


Figure 5: Preview of NetMHCpan 4.0 server output showing strong and weak binding peptides for different alleles

3.5 Cytokine Stimulating Ability of Obtained HTL Epitopes

In this study, the IL4pred, IL10pred, and IFNepitope servers were used to analyze the ability of strong binding Helper T-Lymphocyte (HTL) epitopes to meet three critical criteria: the production of cytokines IL-4, IL-10, and IFN-gamma. The goal of this rigorous study was to find epitopes that can trigger these major cytokines, which are important in immunological regulation and host defense mechanisms. Five HTL epitopes that fit all three requirements were found for the first protein candidate, indicating that they have the ability to trigger IL-4, IL-10, and IFN-gamma. Similar to the first protein candidate, 4 HTL epitopes were found for the second protein candidate that met these requirements, as shown in Table 4 and Table 5.

Table 4: HTL epitopes of protein candidate no: 1 with their IFN, IL4 and IL10 inducing capability

Protein no: 1				NASVYAANAGVDNRE	NEGATIVE	Non-Inducer	Non-Inducer
HTL peptide sequence	IFN	IL4 Prediction	IL10 Prediction	NIIYYHAGTSRLLAVG	POSITIVE	Non-Inducer	Inducer
AMDFTTLQANKSEVP	NEGATIVE	Inducer	Non-Inducer	NKPYWLQRAQGHNNG	NEGATIVE	Non-Inducer	Non-Inducer
ARTNIYYHAGTSRLL	NEGATIVE	Non-Inducer	Non-Inducer	NQLFVTVVDTRSTN	NEGATIVE	Inducer	Non-Inducer
ASSNYFPTPSGSMVT	NEGATIVE	Non-Inducer	Non-Inducer	NRECISMDYKQTQLC	NEGATIVE	Inducer	Inducer
ASVYAANAGVDNREC	POSITIVE	Non-Inducer	Non-Inducer	PDYIKMVSEPYGDSL	NEGATIVE	Inducer	Inducer
AVGHPYFPIKKNNNK	POSITIVE	Inducer	Inducer	PPELINTVIQDGD	NEGATIVE	Inducer	Non-Inducer
CKITLTADVTTYIHS	POSITIVE	Inducer	Non-Inducer	PYFPIKKNNNKILVP	NEGATIVE	Inducer	Non-Inducer
CKYPDYIKMVSEPYG	NEGATIVE	Inducer	Non-Inducer	QLFVTVVDTRSTNM	NEGATIVE	Inducer	Non-Inducer
CPPLELINTVIQDGD	POSITIVE	Non-Inducer	Non-Inducer	RECISMDYKQTQLCL	NEGATIVE	Inducer	Inducer
DCPPELINTVIQDG	NEGATIVE	Non-Inducer	Inducer	RTNIYYHAGTSRLLA	POSITIVE	Non-Inducer	Non-Inducer
DFTTLQANKSEVPLD	NEGATIVE	Inducer	Non-Inducer	SDTSFYNPDTQRLVW	NEGATIVE	Non-Inducer	Non-Inducer
DNRECISMDYKQTQL	NEGATIVE	Inducer	Inducer	SSNYFPTPSGSMVTS	NEGATIVE	Non-Inducer	Non-Inducer
DTSFYNPDTQRLVWA	NEGATIVE	Non-Inducer	Non-Inducer	STDEYVARTNIYYHA	POSITIVE	Non-Inducer	Non-Inducer
ENASVYAANAGVDNR	POSITIVE	Non-Inducer	Non-Inducer	SVYAANAGVDNRECI	NEGATIVE	Inducer	Non-Inducer
FGAMDFTTLQANKSE	NEGATIVE	Inducer	Inducer	TDEYVARTNIYYHAG	POSITIVE	Non-Inducer	Non-Inducer
FPIKKNNNKILVPKV	NEGATIVE	Inducer	Non-Inducer	TNIYYHAGTSRLLAV	POSITIVE	Non-Inducer	Non-Inducer
GAMDFTTLQANKSEV	NEGATIVE	Inducer	Non-Inducer	TSFYNPDTQRLVWAC	NEGATIVE	Inducer	Non-Inducer
GHPYFPIKKNNNKIL	POSITIVE	Inducer	Inducer	VARTNIYYHAGTSRL	NEGATIVE	Non-Inducer	Non-Inducer
GNQLFVTVVDTRST	NEGATIVE	Inducer	Non-Inducer	VGHPYFPIKKNNNKI	POSITIVE	Inducer	Inducer
HPYFPIKKNNNKILV	POSITIVE	Inducer	Inducer	WGNQLFVTVVDTRTS	NEGATIVE	Inducer	Non-Inducer
IYYHAGTSRLLAVGH	NEGATIVE	Non-Inducer	Inducer	YFPIKKNNNKILVPK	NEGATIVE	Inducer	Non-Inducer
KITLTADVTTYIHSM	NEGATIVE	Inducer	Non-Inducer	YPDYIKMVSEPYGDS	NEGATIVE	Inducer	Inducer
KYPDYIKMVSEPYGD	NEGATIVE	Inducer	Non-Inducer				
LASSNYFPTPSGSMV	NEGATIVE	Non-Inducer	Non-Inducer				
LAVGHPYFPIKKNNN	POSITIVE	Inducer	Inducer				
LCKITLTADVTTYIH	POSITIVE	Inducer	Non-Inducer				
MDFTTLQANKSEVPL	NEGATIVE	Inducer	Non-Inducer				

Table 5: HTL epitopes of protein candidate no: 1 with their IFN, IL4 and IL10 inducing capability

Protein no: 2				
HTL peptides sequences	IFN	IL10 Prediction	HTL peptide sequences	IL4 Prediction
AMDFTTLQANKSEVP	NEGATIVE	IL10 non- inducer	AMDFTTLQANKSEVP	IL4 inducer
ARTNIYYHAGTSRLL	NEGATIVE	IL10 non- inducer	AVGHPYFPIKKNNN	IL4 inducer
ASAYAANAGVDNREC	NEGATIVE	IL10 non- inducer	CKYPDYIKMVSEPYG	IL4 inducer
ASSNYFPTPSGSMVT	NEGATIVE	IL10 non- inducer	DFTTLQANKSEVPLD	IL4 inducer
AVGHPYFPIKKNNN	POSITIVE	IL10 inducer	DNRECISMDYKQTQL	IL4 inducer
CKYPDYIKMVSEPYG	NEGATIVE	IL10 non- inducer	DTYRFVTSQAIACQK	IL4 inducer
CPPLELINTVIQDGD	POSITIVE	IL10 non- inducer	EDTYRFVTSQAIACQ	IL4 inducer
DCPPELINTVIQDG	NEGATIVE	IL10 inducer	FGAMDFTTLQANKSE	IL4 inducer
DFTTLQANKSEVPLD	NEGATIVE	IL10 non- inducer	GAMDFTTLQANKSEV	IL4 inducer
DNRECISMDYKQTQL	NEGATIVE	IL10 inducer	GHPYFPIKKNNNKI	IL4 inducer
DTSFYNPDTQRLVWA	NEGATIVE	IL10 non- inducer	GNQLFVTVVDTRST	IL4 inducer
DTYRFVTSQAIACQK	NEGATIVE	IL10 inducer	GTLEDTYRFVTSQAI	IL4 inducer
EDTYRFVTSQAIACQ	NEGATIVE	IL10 inducer	HPYFPIKKNNNKIL	IL4 inducer
ENASAYAANAGVDNR	NEGATIVE	IL10 non- inducer	KYPDYIKMVSEPYGD	IL4 inducer

FGAMDFTTLQANKSE	NEGATIVE	IL10 inducer	LAVGHPYFPIKKPNN	IL4 inducer
GAMDFTTLQANKSEV	NEGATIVE	IL10 non- inducer	LEDTYRFVTSQAIAC	IL4 inducer
GHPYFPIKKPNNNKI	POSITIVE	IL10 non- inducer	MDFTTLQANKSEVPL	IL4 inducer
GNQLFVTVVDTRST	NEGATIVE	IL10 non- inducer	MTYIHSMNSTILEDW	IL4 inducer
GTLEDTYRFVTSQAI	NEGATIVE	IL10 inducer	NQLFVTVVDTRSTN	IL4 inducer
HPYFPIKKPNNNKIL	POSITIVE	IL10 inducer	NRECISMDYKQTQLC	IL4 inducer
IYYHAGTSRLLAVGH	NEGATIVE	IL10 inducer	PDYIKMVSEPYGDSL	IL4 inducer
KYPDYIKMVSEPYGD	NEGATIVE	IL10 non- inducer	PPELINTVIQDGM	IL4 inducer
LASSNYFPTPSGSMV	NEGATIVE	IL10 non- inducer	QLFVTVVDTRSTNM	IL4 inducer
LAVGHPYFPIKKPNN	POSITIVE	IL10 inducer	RECISMDYKQTQLCL	IL4 inducer
LEDTYRFVTSQAIAC	NEGATIVE	IL10 inducer	SAYAANAGVDNRECI	IL4 inducer
MOFTTLQANKSEVPL	NEGATIVE	IL10 non- inducer	TLEDTYRFVTSQAIA	IL4 inducer
MTYIHSMNSTILEDW	NEGATIVE	IL10 inducer	TYRFVTSQAIACQKH	IL4 inducer
NASAYAANAGVDNRE	NEGATIVE	IL10 non- inducer	VGHPYFPIKKPNNNK	IL4 inducer
NIYYHAGTSRLLAVG	POSITIVE	IL10 inducer	VMTYIHSMNSTILED	IL4 inducer
NKPYWLQRAQGHNG	NEGATIVE	IL10 non- inducer	WGNQLFVTVVDTRRS	IL4 inducer
NQLFVTVVDTRSTN	NEGATIVE	IL10 non- inducer	YPDYIKMVSEPYGDS	IL4 inducer
NRECISMDYKQTQLC	NEGATIVE	IL10 inducer	ARTNIYYHAGTSRLL	Non IL4 inducer
PDYIKMVSEPYGDSL	NEGATIVE	IL10 inducer	ASAYAANAGVDNREC	Non IL4 inducer
PPELINTVIQDGM	NEGATIVE	IL10 non- inducer	ASSNYFPTPSGSMVT	Non IL4 inducer
QLFVTVVDTRSTNM	NEGATIVE	IL10 non- inducer	CPPELINTVIQDGD	Non IL4 inducer
RECISMDYKQTQLCL	NEGATIVE	IL10 inducer	DCPPELINTVIQDG	Non IL4 inducer
RTNIYYHAGTSRLLA	POSITIVE	IL10 non- inducer	DTSFYNPDTQRLVWA	Non IL4 inducer
SAYAANAGVDNRECI	NEGATIVE	IL10 non- inducer	ENASAYAANAGVDNR	Non IL4 inducer
SSNYFPTPSGSMVTS	NEGATIVE	IL10 non- inducer	IYYHAGTSRLLAVGH	Non IL4 inducer
TDEYVARTNIYYHAG	POSITIVE	IL10 non- inducer	LASSNYFPTPSGSMV	Non IL4 inducer
TLEDTYRFVTSQAIA	NEGATIVE	IL10 inducer	NASAYAANAGVDNRE	Non IL4 inducer
TNIYYHAGTSRLLAV	POSITIVE	IL10 non- inducer	NIYYHAGTSRLLAVG	Non IL4 inducer
TYRFVTSQAIACQKH	NEGATIVE	IL10 inducer	NKPYWLQRAQGHNG	Non IL4 inducer
VARTNIYYHAGTSRL	NEGATIVE	IL10 non- inducer	RTNIYYHAGTSRLLA	Non IL4 inducer
VGHPYFPIKKPNNNK	POSITIVE	IL10 inducer	SSNYFPTPSGSMVTS	Non IL4 inducer
VMTYIHSMNSTILED	NEGATIVE	IL10 inducer	TDEYVARTNIYYHAG	Non IL4 inducer
WGNQLFVTWDTTRS	NEGATIVE	IL10 non- inducer	TNIYYHAGTSRLLAV	Non IL4 inducer
YPDYIKMVSEPYGDS	NEGATIVE	IL10 inducer	VARTNIYYHAGTSRL	Non IL4 inducer

After thoroughly screening HTL epitopes, those which showed inducing ability for all three cytokines (IFN, IL4 and IL10) were chosen for further analysis. In the VaxiJen v2.0 server, the antigenicity of the epitopes were predicted. The antigenic HTL epitopes then were entered into the AllerTOP v2.0 and ToxinPred server for allergenicity and toxicity prediction. The results are shown below in the Table 6, marked as green color.

Table 6: Selected HTL epitopes with their antigenicity, allergenicity and toxicity for Protein no: 1 & 2.

Protein No: 1			
HTL epitopes	Antigenicity	Toxicity	Allergenicity
AVGHPYFPIKKNNNK	0.6027 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
GHPYFPIKKNNKIL	Non-Antigen	Non-Toxin	-
HPYFPIKKNNNKILV	0.5182 (Probable ANTIGEN)	Non-Toxin	Allergen
LAVGHPYFPIKKNNN	0.9991 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
VGHPYFPIKKNNNKI	0.5410 (Probable ANTIGEN)	Non-Toxin	Allergen
Protein No: 2			
PPLELINTVIQDGD	Non-Antigen	Non-Toxin	Non-Allergen
HPYFPIKKPNNNKIL	Non-Antigen	Non-Toxin	Allergen
LAVGHPYFPIKKPNN	0.5867 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
VGHPYFPIKKPNNNK	Non-Antigen	Non-Toxin	Allergen

3.6 Determination of B-cell Epitopes

The main protein sequence in the IEDB Analysis Resource was examined using the Bepipred Linear Epitope Prediction 2.0 technique, which produced the results. A thorough analysis revealed a total of 11 unique peptides for Protein 1 and a total of 13 distinct peptides for Protein 2 that were classified as B cell epitopes. In the initial screening, epitopes with extremely short lengths were eliminated.

No.	Start	End	Peptide	Length	No.	Start	End	Peptide	Length
1	37	54	VYLPPVPVSKVVSTDEYV	18	1	34	34	N	1
2	107	120	KFGFSDTSFYNPDT	14	2	61	72	GFPDTSFYNPDT	12
3	137	138	PL	2	3	101	120	NKLDDTENASAYAANAGVDN	20
4	148	168	LNKLDDTENASVYAANAGVDN	21	4	142	154	IGEHMGKGSPECTN	13
5	190	202	IGEHMGKGSPECTN	13	5	172	173	IQ	2
6	220	222	IQD	3	6	190	197	LQANKSEV	8
7	237	245	TLQANKSEV	9	7	211	219	YIKMVSEPY	9
8	259	268	YIKMVSEPYG	10	8	241	263	AGAVGENVPDDLYIKGSGSTANL	23
9	289	311	AGAVGENVPDDLYIKGSGSTANL	23	9	284	284	F	1
10	330	333	QIFN	4	10	327	346	TSETTYKNTNFKEYLRHGEE	20
11	375	394	TSETTYKNTNFKEYLRHGEE	20	11	373	400	STILEDWNFGLQPPPGGTLEDTYRFVTS	28
					12	403	421	IACQKHTPPAPKEDPLKKY	19
					13	430	435	EKFSAD	6

Figure 6: B-Cell epitope prediction

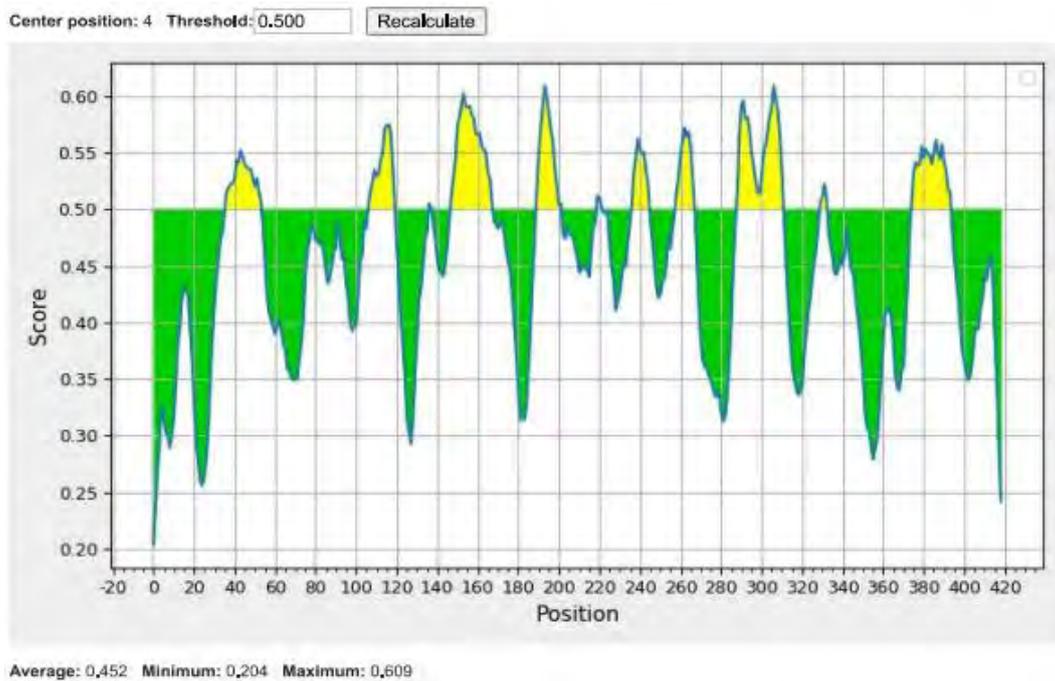


Figure 7: Diagram showing the connection between a score and position of B-cell epitope for protein no: 1(Jespersen et al., 2017)

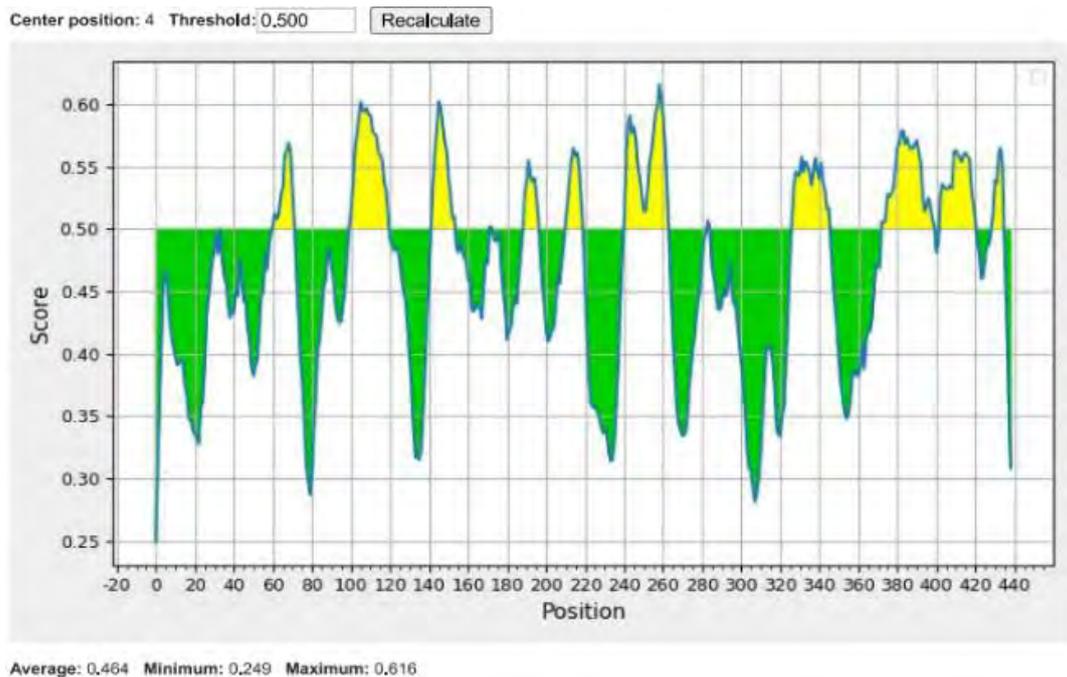


Figure 8: Diagram showing the connection between a score and position for protein no: 2 (Jespersen et al., 2017)

The average score of B cell epitopes was 0.452 and 0.464, respectively, with the minimum value being 0.204 and 0.249 and the maximum score being 0.606 and 0.616 (Jespersen et al., 2017).

Table 7: Antigenic B-cell epitopes with their ToxinPred and AllerTOP v2.0 server prediction for both Protein 1 & 2

	B-cell epitopes	Antigenicity	Toxicity	Allergenicity
Protein No: 1	VYLPPVPVSKVVSTDEYV	0.5004 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
	KFGFSDTSFYNPDT	1.0027 (Probable ANTIGEN).	Non-Toxin	Non-Allergen
	TSETTYKNTNFKEYLRHGEE	0.5883 (Probable ANTIGEN)	Non-Toxin	Non-Allergen
Protein No: 2	YIKMVSEPY	0.5855 (Probable ANTIGEN).	Non-Toxin	Allergen
	TSETTYKNTNFKEYLRHGEE	0.5883 (Probable ANTIGEN).	Non-Toxin	Non-Allergen
	STILEDWNFGLQPPPGGTLEDTYRFVTS	0.6008 (Probable ANTIGEN).	Non-Toxin	Non-Allergen

3.7 Assembly of Vaccine Candidates with Their Antigenicity, Allergenicity and Toxicity Profile

In this study, the primary adjuvant was synergistically coupled with a wide array of epitopes B-cell, Helper T cell and Cytotoxic T lymphocyte domains. This combination was made possible by the addition of linkers specific for each type of epitopes. A specific linker sequence known as "EAAAK" was used to create adjuvant-linked CTL epitopes, while another linker known as "AAY" was used to connect several CTL epitopes. Additionally, a novel fusion approach included joining HTL and CTL epitopes using the intermediary linker "GPGPG." B-cell epitopes were also linked to HTL epitopes using the linker sequence "KK." A crucial aspect of our research investigation is the versatile inclusion of these meticulously generated epitope-linker complexes.

Here are the two final constructed vaccines,

Protein No: 1: AAM74159.1

Constructed Vaccine (i):

MQVTFIYILVITCYENDVNVYHIFQMSLWLPSEATVYLPVPVSKVVSTDEYVAR
TNIYYHAGTSRLLAVGHPYFPIKKNKILVPKVSGLQYRVFRIHLDPNKFQFSDT
SFYNPDTQRLVWACVGVVGRGQPLGVLGSGHPLLKDDTENASVYAANAGVD
NRECISMDYKQTQLCLIGCKPPIGEHWGKSPCTNVAVNPGDCPPLELINTVIQDG
DMVHTGFGAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDSLFFYLRRREQ
MFVRHLFNRAVAVGENVPDDLVIKSGSTANLASSNYFPTPSGSMVTSDAQIFNKP
YWLQRAQGHNNGICWGNQLFVTVDTRSTNMSLCAAISTSETTYKNTNFKEYL
RHGEEYDLQFIFQLCKITLTADVTTIYHSM**EAAAK**TANLASSNY**GPGPGA**VGHPYF
PIKKNK**GPGPLA**VGHPYFPIKKNK**KK**VYLPVPVSKVVSTDEYV**KK**KFGFS
DTSFYNPDT**KK**TSETTYKNTNFKEYLRHGEE

Protein No: 2: AEA76067.1

Constructed Vaccine (ii):

TDEYVARTNIYYHAGTSRLLAVGHPYFPIKKNKILVPKVSGLQYRVFRIHLDP
PNKFGFPDTSFYNPDTQRLVWACVGVVGRGQPLGVLGSGHPLLKDDTENASA
YAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKSPCTNVAVNPGDCPPLE
LINTVIQDGDMDVTGFGAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDS
LFFYLRRREQMFVRHLFNRAVAVGENVPDDLVIKSGSTANLASSNYFPTPSGSMV
TSDAQIFNKPYWLQRAQGHNNGICWGNQLFVTVDTRSTNMSLCAAISTSETTY
KNTNFKEYLRHGEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPG
GTLEDTYRFVTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKEKFSADLDQF**EAAA**
KSTILEDWNF**AA**TANLASSNY**GPGPLA**VGHPYFPIKKN**KK**TSETTYKNTNFK
EYLRHGEE**KK**STILEDWNFGLQPPPGGTLEDTYRFVTS

Their antigenicity, allergenicity was evaluated through VaxiJen v2.0 server and AllergenOnline server. It was also evaluated that there is no presence of toxins in the vaccine via T3DB server. The output of the servers is given below in Table 8. Antigenicity score of both of the vaccines have increased.

Table 8: Antigenicity, Allergenicity and Toxin Presence Profile of Final Vaccines

	Antigenicity	Allergenicity	Toxin Presence
Vaccine (i)	0.5516 (Probable ANTIGEN)	Number of Sequences with hits: 0	No Results
Vaccine (ii)	0.5076 (Probable ANTIGEN)	Number of Sequences with hits: 0	No Results

3.8 In-silico Biochemical Analysis of Candidate Vaccines

We performed a study utilizing ExPASy's ProtParam Tool service to thoroughly assess possible changes in the physicochemical characteristics of both the final vaccines (Rodríguez-Ruiz et al., 2019). The number of amino acids was 531 and 522 for vaccine (i) and (ii) respectively. Molecular weight of these vaccines was 59186.19 and 58307.61 kDa respectively. With the use of vaccine protein sequences, this server provides a firm basis for the theoretical prediction of several important physical and chemical properties. In addition to the molecular weight (MW), theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient (EC), estimated half-life, instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY), these characteristics also include the estimated half-life (Rodríguez-Ruiz et al., 2019). Theoretical isoelectric points (pI) were computed for the final vaccine designs and found to be 7.87 and 5.63, respectively. The EC measurement is useful for estimating protein concentrations in solution, an important factor in purifying procedures, in addition to providing information on the protein's light-absorbing characteristics (Rodríguez-Ruiz et al., 2019). In addition, an analysis of the instability index (II), a measure of protein stability in a laboratory context,

was conducted (Rodríguez-Ruiz et al., 2019). Notably, the instability index (II) values for both vaccine formulations were below the threshold point of 40, indicating that both designs would likely have overall protein stability (Rodríguez-Ruiz et al., 2019). The aliphatic index (AI) was calculated, which measures the relative volume filled by aliphatic side chains such as alanine, valine, leucine, and isoleucine (Rodríguez-Ruiz et al., 2019). This index is thought to be a benefit that increases the thermal stability of globular proteins (Rodríguez-Ruiz et al., 2019). The GRAVY (grand average of hydropathicity) value for the proteins under study was then established. This metric accounts for both the length of the amino acid sequence and the overall hydrophobic or hydrophilic properties of a protein (positive or negative values) (Rodríguez-Ruiz et al., 2019). It was determined that vaccines (i) and (ii) had GRAVY values of -0.314 and -0.435, respectively. These precise and thorough investigations have offered a thorough grasp of the physicochemical properties of the examined proteins, substantially improving our study work(Parvizpour et al., 2020).

ProtParam

User-provided sequence:

```

1Q      7Q      13Q      19Q      25Q      31Q
MSDFEFLVLY ITCVSMRWNY VHEFEPQLM LRSKATVLEF WYVSVVVDI QRYMAYPLSY
7Q      13Q      19Q      25Q      31Q      37Q
WAGYSKLLA MSHYYPPIKE AMMKLVKPV SGLQWVWREI HLRDMMAPDF SDTSFYVSDT
13Q     19Q     25Q     31Q     37Q     43Q
QGLVAVLVGQ EVNGRQDQYK QTSRPLNE LDOTENAFY WMAAYDREI ETRSDYKDTQ
19Q     25Q     31Q     37Q     43Q     49Q
LGLDGLPPI GHWKQKQSC TWYWRKSK RFLLEMLTYI QDQWVHTDF GSRDFTLQK
25Q     31Q     37Q     43Q     49Q     55Q
MKEPFLQIC YSTQKQVYI KMSEPRQDS LKFKYKSKQI FVDEEFAES KWRWVWQSL
31Q     37Q     43Q     49Q     55Q     61Q
YVLDGSLAR LASSMPTTF SGNSTQDAD IYKQVWQDI QDQWVHTDF SGLVYVYVD
47Q     53Q     59Q     65Q     71Q     77Q
TTSTWGLC AAKTSETTY ANIKPKYLR KKEVLDQET EQLCITLTA QVTTKDSDF
47Q     53Q     59Q     65Q     71Q     77Q
AAKATLAKG SYPQDQYK GHYFPIKAI MNGDQSLA MSHYYPPIKE WMAVYVLEP
53Q     59Q     65Q     71Q     77Q     83Q
WYVSVVVDI EVWKKQDQ'S QTSRPMDFE KTSSETTYKMT NFKYLRHGE F
    
```

Total number of negatively charged residues (Asp + Glu): 47
 Total number of positively charged residues (Arg + Lys): 49

Atomic composition:

Carbon	C	2669
Hydrogen	H	4889
Nitrogen	N	697
Oxygen	O	785
Sulfur	S	22

Formula: C₂₆₆₉H₄₈₈₉N₆₉₇O₇₈₅S₂₂
 Total number of atoms: 8242

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in w
 Ext. coefficient 78480
 Abs 0.1% (=1 g/l) 1.258, assuming all pairs of Cys residues form cystine

Ext. coefficient 73690
 Abs 0.1% (=1 g/l) 1.245, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).
 The estimated half-life is: 38 hours (mammalian reticulocytes, in vitro).
 >20 hours (yeast, in vivo).
 >10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 32.13
 This classifies the protein as stable.

Aliphatic index: 74.29

Grand average of hydropathicity (GRAVY): -0.314

[References](#) and [documentation](#) are available.

Number of amino acids: 83

Molecular weight: 93180.39

Theoretical pI: 7.87

Amino acid composition:

[CSV format](#)

Ala (A)	28	5.30
Arg (R)	15	2.89
Asn (N)	18	3.25
Asp (D)	25	4.73
Cys (C)	22	4.19
Gln (Q)	20	3.89
Glu (E)	26	4.93
Gly (G)	39	7.29
His (H)	14	2.62
Ile (I)	26	4.90
Leu (L)	30	5.65
Lys (K)	34	6.43
Met (M)	10	1.90
Phe (F)	26	4.93
Pro (P)	37	7.05
Ser (S)	35	6.63
Thr (T)	30	5.65
Trp (W)	8	1.52
Tyr (Y)	31	5.83
Val (V)	45	8.48
Py1 (O)	0	0.00
Sec (U)	0	0.00

(R) 0 0.00
 (E) 0 0.00
 (S) 0 0.00

Figure 9: ProtParam Tool Result for Vaccine (i)

```

77Q      83Q      89Q      95Q      101Q     107Q
TADWYTDYHS RMTLLEIDWQ FGLDPRRGGT LEDYVRFVTS DALACQHTF PWRKEDLAK
45Q      51Q      57Q      63Q      69Q      75Q
VTFKAVLEK FESMLDQTF AAKTSTLLEI WFAAYTAKI SCRWDGKPC VETSETTYEN
49Q      55Q      61Q      67Q
TWREYRRIK ETRKSTLLEI WREGQPRK STLEITYRFPV F
    
```

Total number of negatively charged residues (Asp + Glu): 57
 Total number of positively charged residues (Arg + Lys): 47

Atomic composition:

Carbon	C	2612
Hydrogen	H	3966
Nitrogen	N	686
Oxygen	O	791
Sulfur	S	21

Formula: C₂₆₁₂H₃₉₆₆N₆₈₆O₇₉₁S₂₁
 Total number of atoms: 8076

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.
 Ext. coefficient 83490
 Abs 0.1% (=1 g/l) 1.432, assuming all pairs of Cys residues form cystine

Ext. coefficient 82740
 Abs 0.1% (=1 g/l) 1.419, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is T (Thr).
 The estimated half-life is: 7.2 hours (mammalian reticulocytes, in vitro).
 >20 hours (yeast, in vivo).
 >10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 31.94
 This classifies the protein as stable.

Aliphatic index: 69.58

Grand average of hydropathicity (GRAVY): -0.435

[References](#) and [documentation](#) are available.

Number of amino acids: 107

Molecular weight: 93307.61

Theoretical pI: 5.83

Amino acid composition:

[CSV format](#)

Ala (A)	32	6.13
Arg (R)	17	3.19
Asn (N)	34	6.38
Asp (D)	30	5.73
Cys (C)	32	6.05
Gln (Q)	19	3.59
Glu (E)	27	5.09
Gly (G)	39	7.29
His (H)	11	2.10
Ile (I)	24	4.52
Leu (L)	40	7.57
Lys (K)	30	5.65
Met (M)	0	0.00
Phe (F)	20	3.77
Pro (P)	35	6.57
Ser (S)	30	5.65
Thr (T)	45	8.28
Trp (W)	8	1.52
Tyr (Y)	20	3.77
Val (V)	28	5.24
Py1 (O)	0	0.00
Sec (U)	0	0.00

(R) 0 0.00
 (E) 0 0.00
 (S) 0 0.00

Figure 10: ProtParam Tool Result for Vaccine (ii)

3.9 Homology modeling

The homologous protein and the hypothetical protein were both created as precise 3D representations with their appropriate amino acid alignments using the software Phyre2. The result showed that the confidence was 100% with 79% coverage for vaccine (i) and 100% confidence with 85% coverage for vaccine (ii).

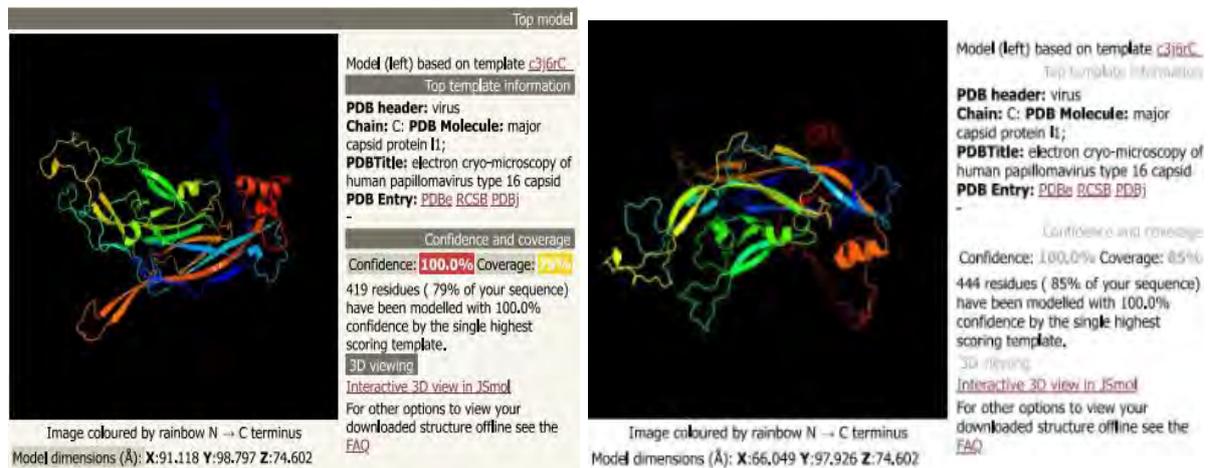


Figure 11: Phyre2 result showing homology modeling of vaccine (Kelley et al., 2015)

3.10 Z-Score and Ramachandran Plots Prediction

The final vaccine build was created as a Protein Data Bank (PDB) file using the Phyre 2 server, which was then accessible via the Discovery Studio program. Our biological research was advanced thanks in large part to this PDB dataset. An illustration of the Z-score in relation to the quantity of residues was produced as a result of the vaccine's PDB file being submitted to ProSAweb. In our example, the Z-score was found to be -3.6, as shown graphically in Figure 12, and is an important indicator for evaluating the overall quality of the model.

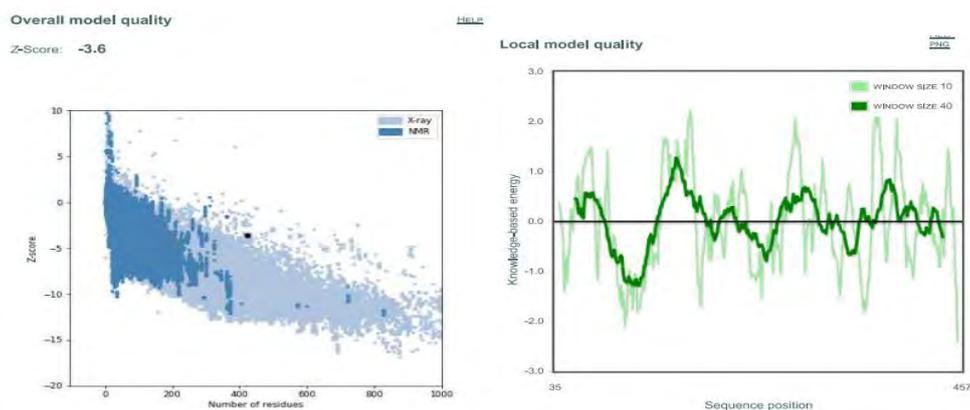


Figure 12: Vaccine (i): (A) The graph of the Z-score against the number of residues and the Z-score on the ProSAweb server.

(b) Local model of high quality linking knowledge-based energy to position in the sequence (Wiederstein & Sippl, 2007).

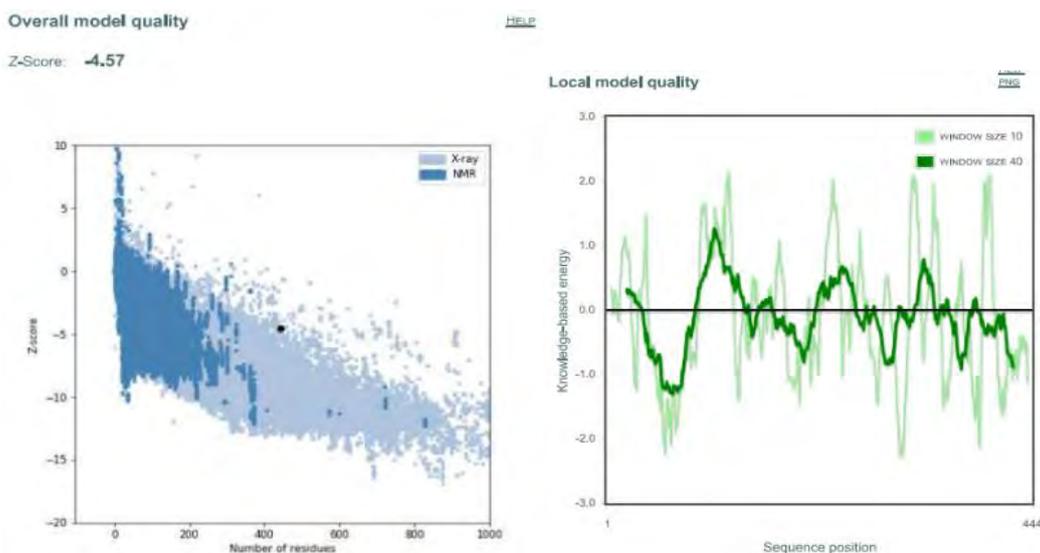


Figure 13: Vaccine (ii): (A) The graph of the Z-score against the number of residues and the Z-score on the ProSAweb server.

The PDB file then was uploaded in the <https://swissmodel.expasy.org/> server to generate the Ramachandran plot. The scores of the various parameter are shown below in Table 9.

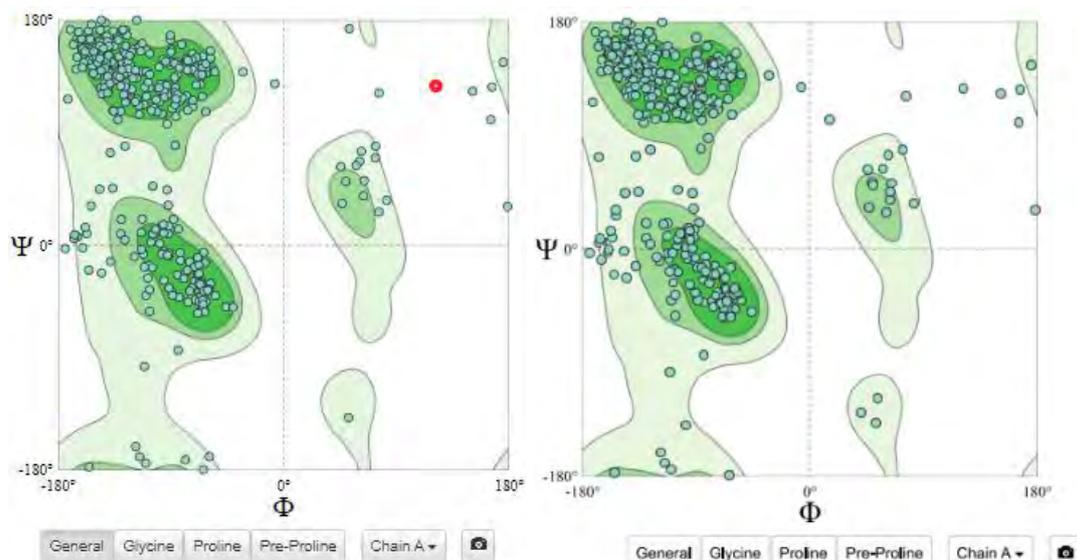


Figure 14: Ramachandran Plot of Vaccine (i) & (ii)

As determined by the Phyre2 server, the Swiss-Model Interactive Workplace was helpful in determining a number of important parameters relating to the anticipated three-dimensional structure. These factors included the Ramachandran preferred area, the Ramachandran Outliers, the MolProbity Score, and a wide range of additional assessment criteria, all of which are painstakingly described in detail in Table 9. This thorough evaluation, made possible by the Swiss-Model Interactive Workplace and the Phyre2 server, is essential for determining the structural quality and integrity of the protein model under investigation and improves the extent and rigor of our research findings.

Table 9: Summary of Biochemical analysis and 3D Structure Analysis

Server	Parameters	Vaccine (i)	Vaccine (ii)
ProtParam-Vaccine Stability (Physio-Chemical Parameters Values)	Number of amino acids	531	522
	Molecular weight	59186.19	58307.61
	Theoretical isoelectric point (pI)	7.87	5.63
	Aliphatic index	74.29	69.5
	Instability index	32.13	31.94
	Extinction coefficients (all pairs of Cys residues form cystines)	74440	83490
	Extinction coefficients (all Cys residues are reduced)	73690	82740
	Total number of negatively charged residues (Asp + Glu)	47	57
	Total number of positively charged residues (Arg + Lys)	49	47
	Grand average of hydropathicity (GRAVY)	-0.314	-0.435
Phyre2	Confidence	100%	100%
	Coverage	79%	85%
ProSaWeb	Z-Score	-3.6	-4.57
Ramachandran Plot (SWISS-MODEL)	MolProbity	2.82	2.78
	Ramachandran favored	85.51%	85.29%
	Ramachandran outliers	4.75%	4.07%
	Rotamer outlier	0.00%	0.00%
	C-Beta Deviations	0	0

3.11 Analysis of the Constructed Vaccine Candidate by Molecular Docking with the Relevant Human Receptor

The workflow for ClusPro consists of two main stages. Extensive sampling is used to produce docking conformations in the first phase (Kozakov et al., 2017). While evaluating various ligand orientations, the center of mass of the receptor remains stationary at the

coordinate system's origin (Kozakov et al., 2017). The translational locations are collected at a step size of 1, and the rotational space is discretized with roughly 70,000 rotations, each covering 5 degrees in Euler angles (Kozakov et al., 2017). The end result is an enormous variety of possible protein-ligand conformations (Kozakov et al., 2017). The next step is to group the 1,000 docked structures with the lowest energy, using pairwise Inverse Rotamer-Merger Square Deviation (IRMSD) as a distance measure (Kozakov et al., 2017). All structure pairs in the set have their IRMSD values calculated, and the center of the first cluster is determined to be the structure with the most neighbors within a 9- IRMSD radius(Kozakov et al., 2017). The initial cluster is then made up of structures that are located within this radius. Up to 30 clusters are produced by this method . After clustering, retained structures are subjected to energy minimization to reduce steric overlaps, resulting in minute conformational changes (Kozakov et al., 2017). The structures in the ten most populous clusters' centers are output by ClusPro in its default mode of operation. By streamlining the investigation of various protein-ligand binding modalities, this method helps scientists comprehend molecular interactions.

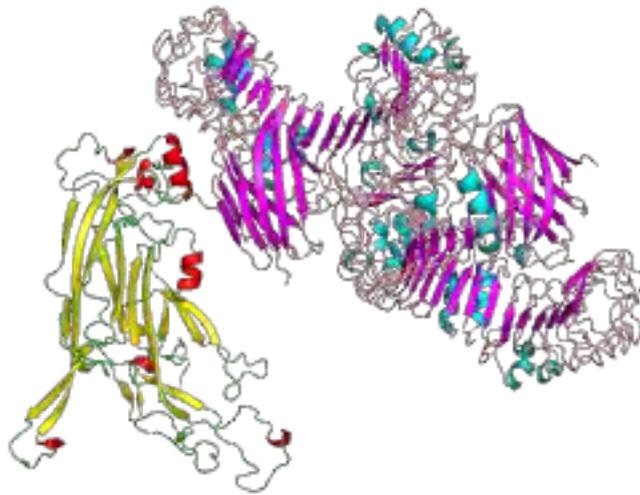


Figure 15: *Creating a docked complex in three dimensions between the final vaccine (i) and TLR4 using Discovery Studio*

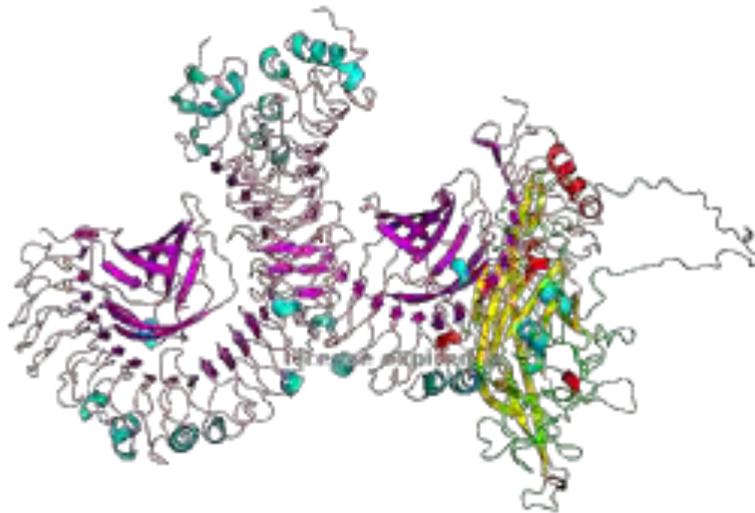


Figure 16: *Creating a docked complex in three dimensions between the final vaccine (ii) and TLR4 using Discovery Studio.*

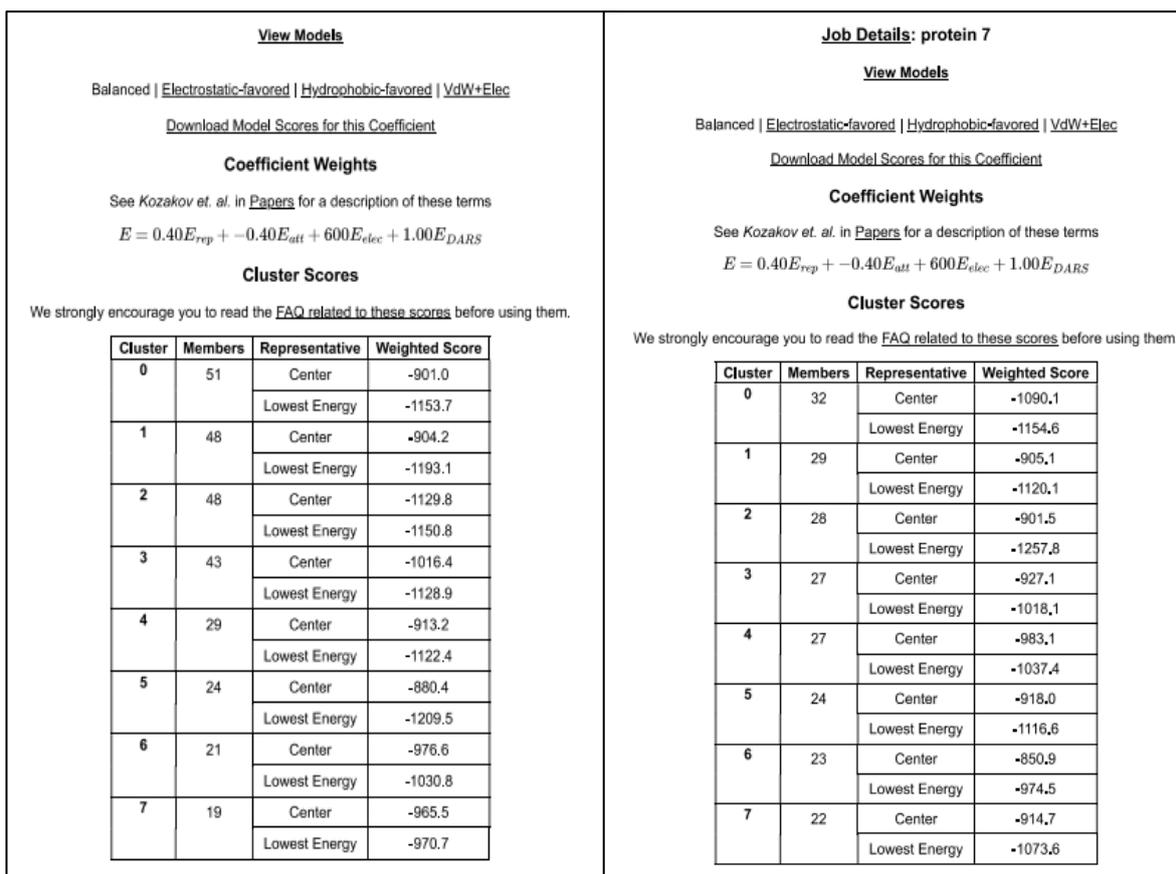


Figure 17: Protein-Protein docking ClusProv2.0 Model Score for Vaccine (i) & (ii)

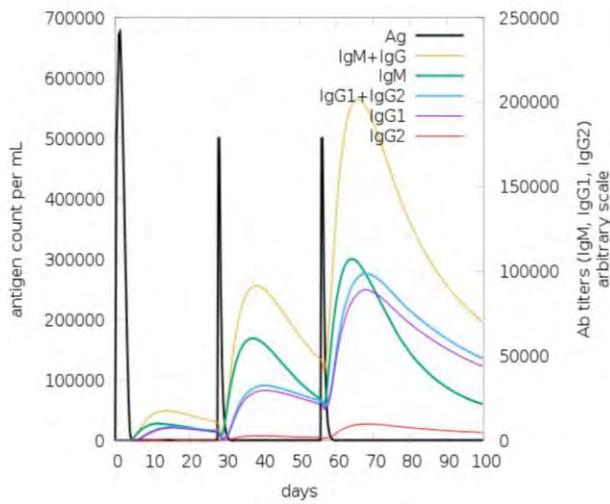
In this research, we used the ClusPro V2.0 docking server, which Nezafat et al. introduced in 2016, to evaluate the affinity of antigenic ligands with the human TLR4 receptor that have been suggested as possible vaccinations. The rotational investigation of the TLR4 receptor in combination with the antigenic ligands was made easier by the ClusPro V2.0 server, which produced 1000 different docking conformations for thorough study. Following the methods described by Comeau et al. in 2004, the best binding clusters were selected based on the criteria of both high cluster population and reduced energy. A total of 29 clusters were found among the docking results generated and rigorously assessed. In the end, the cluster designated as "Cluster 0" stood out as the top choice for additional research. This choice was made in light of Cluster 0's outstanding characteristics, including the greatest cluster membership (51 for vaccine candidate (i) and 32 for vaccine candidate (ii), respectively).

Additionally, Cluster 0 had the lowest energy ratings for vaccine candidates (i) and (ii), measuring -1153.7 and -1154.6, respectively. The fact that Cluster 0 was chosen as the best docked cluster highlights the possibility of this binding configuration between the molecules constituting the vaccination ligands and the TLR4 receptor. The specific docked cluster in question was then retrieved for in-depth investigation and additional use in our research projects.

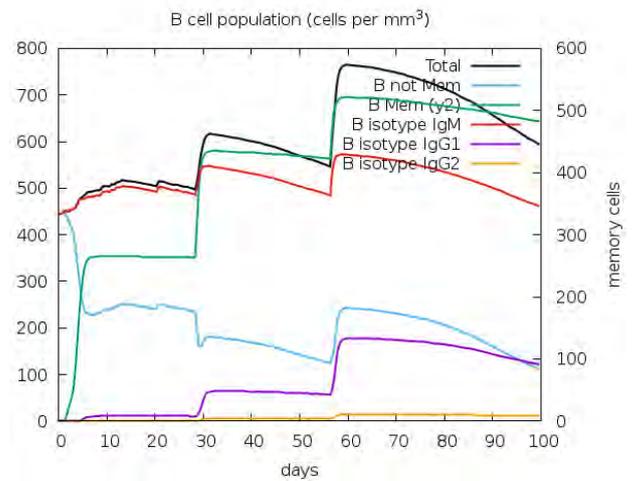
3.12 Immune Simulations

A crucial component of acquired immunity is the role that antibodies play in the body's defense against infection. Figure (a) provides an instance of the immune response kinetics following vaccination. It is noteworthy that there is a discernible increase in antigen count per mL, which is greatest after the first vaccination. Following immunization, IgM and IgG antibodies both exhibit a modest initial rise within the first 28 days, but their levels increase dramatically and steadily after about 60 days (Al-Khayyat; et al., 2016). This significant rise is attributable to the repeated injection of vaccine doses, demonstrating the successful production of the anticipated immune response. This immune response depends on B lymphocytes, which are immunological cells responsible for producing memory cells and antibodies (Parvizpour et al., 2020). Memory cells made from B cells, which speed up the immune response, give the body the ability to quickly recognize and combat the same illness during subsequent interactions (Ghorban Hosseini et al., 2017). In order to offer an immunization that is both effective and long-lasting, the production of developed memory cells becomes of the utmost importance. Figure (b) provides a detailed breakdown of the total number of memory cells and B lymphocytes after immunization using different isotypes. A thorough breakdown of B cell states is shown in Figure (c), highlighting the differences between active B cells, those that display antigens on Class II molecules, B

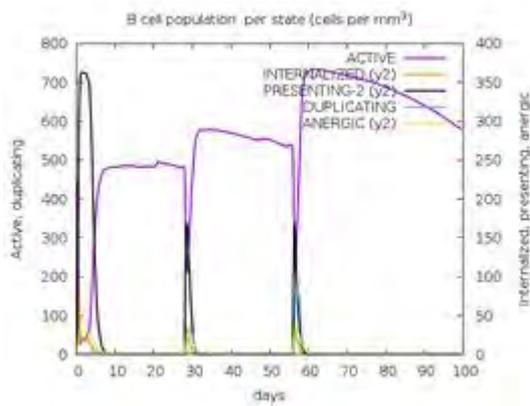
cells that internalize the antigen, and B cells that are duplicating and becoming anergic. The number of plasma B lymphocytes (PLB), further divided into isotypes such IgG1, IgG2, IgM, and IgM + IgG, is shown in Figure (d) (Kumar et al., 2022). Following each vaccine dosage, the count of CD4 T helper cells is predicted. Additionally, the amount of Helper T cells simulation while categorizing them into active, duplicative, resting, or anergic states using C-IMMSIM, the total and memory CD8 T cytotoxic (TC) cell count, which is divided into active, duplicative, resting, and anergic stages are shown in Figure 18. Dendritic cells are shown in multiple stages, such as active, resting, antigen-presenting, and internalized. It's important to notice that the last graph in Figure 18 shows the concentrations of interleukin-2 and danger signals, two essential components in controlling the immune response (Tan et al., 2023). The breadth and rigor of our study findings are increased by these in-depth analyses and graphical depictions, which combined help us grasp the complex dynamics underlying the post-vaccination immune response.



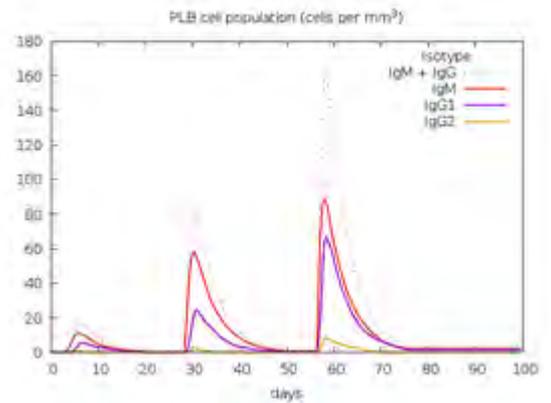
(a) Antigen count per mL and antibody titers (Rapin et al., 2010)



(b) Overall count of B lymphocytes and memory cells (Rapin et al., 2010)

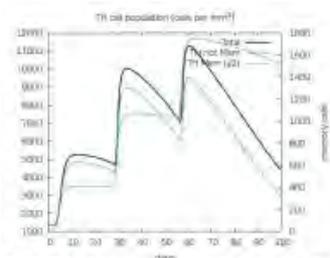


(c) Entity-state of B cell population (Rapin et al., 2010)

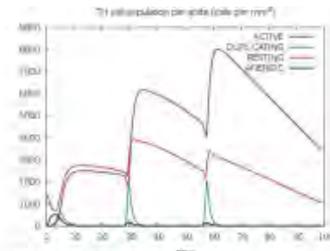


(d) Number of plasma B cells according to their isotypes

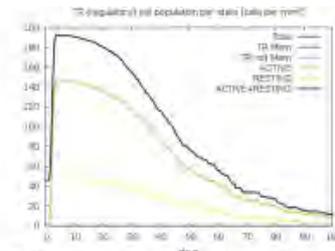
Figure 18: CimmSim Output showing Immune Simulation of Vaccine (i).



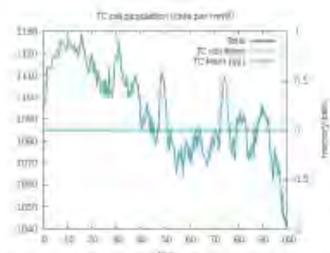
CD4 T-helper lymphocytes count. The plot shows total and memory counts.



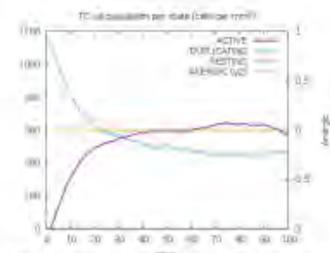
CD4 T-helper lymphocytes count subdivided per entity-state (i.e., active, resting, anergic and duplicating).



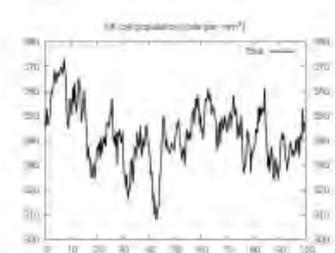
CD4 T-regulatory lymphocytes count. total, memory and per entity-state count plotted here.



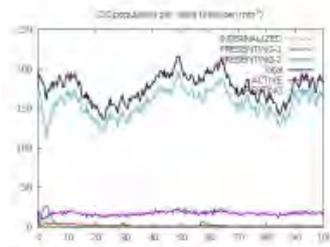
CD8 T-cytotoxic lymphocytes count. Total and memory shown.



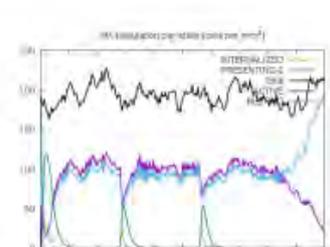
CD8 T-cytotoxic lymphocytes count per entity-state.



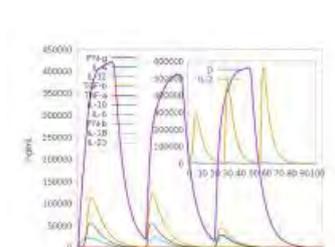
Natural Killer cells (total count).



Dendritic cells. DC can present antigenic peptides on both MHC class-I and class-II molecules. The curves show the total number broken down to active, resting, internalized and presenting the ag.



Macrophages. Total count, internalized, presenting on MHC class-II, active and resting macrophages.



Cytokines. Concentration of cytokines and interleukins. D in the inset plot is danger signal.

Figure 19: CimmSim server output for Vaccine (i)

Chapter 4

Discussion

In recent years, oncogenic infectious pathogens, notably Human Papillomavirus (HPV), have emerged as significant threats to human health (Mahmoudvand et al., 2022). Among these pathogens, HPV is particularly noteworthy due to its substantial role in the etiology of cervical cancer (Mahmoudvand et al., 2022). Nearly all cases of cervical cancer worldwide can be attributed to HPV infection, with HPV-16 being one of the most prominent and clinically relevant strains linked to this malignancy. Extensive epidemiological studies conducted in Iran have consistently identified HPV-16 as the predominant high-risk type, further accentuating its clinical relevance in the region. Given the pivotal clinical significance of HPV-16 in cervical cancer, the development of vaccination strategies represents a paramount preventive measure against HPV infection (Lei, Y. et al., 2019). In this regard, the L1 protein of HPV-16 emerges as an auspicious candidate for vaccine development (Mahmoudvand et al., 2022). The L1 protein boasts multiple well-characterized neutralizing epitopes, making it an ideal target for eliciting immune responses capable of blocking viral infection (Lei, Y. et al., 2019). This paper describes a deliberate strategy that resulted in the discovery of highly antigenic vaccines, which are distinguished by a decreased incidence of adverse responses. Within this framework, a particular protein sequence was painstakingly chosen due to its strong binding affinity with important epitopes found on B cells, helper T lymphocytes, and cytotoxic T lymphocytes. Through the addition of several antigenic epitopes, the antigenicity of the chosen protein sequence was increased (Lei, Y. et al., 2019). Notably, the completed protein sequence demonstrated a flawless safety profile, and further administrations led to a notable rise in the population's total antibody levels (Lei, Y. et al., 2019). Hydrophobicity, antigenicity, toxicity, allergenicity, surface accessibility, and flexibility are just a few of the many factors that must be carefully considered when choosing effective epitopes for vaccine design (Yokomine, M. et

al., 2017). The structural stability of a vaccine construct also turns out to be a key factor in determining how effective it will be. Structural stability has a significant effect on the presentation of T- and B-cell epitopes as well as the intracellular destiny of antigens during processing and presentation within antigen-presenting cells. In order to effectively present antigenic peptides on the Major Histocompatibility Complex (MHC), which is essential to strongly stimulate the immune system, a vaccine design must be as stable as possible (Yokomine, M. et al., 2017). A protein's destabilization may cause tertiary structures to unravel, resulting in the loss of conformational epitopes. The development of multiepitope vaccines, which are made up of areas containing highly immunogenic B- and T-cell epitopes, is a method capable of generating substantial humoral and cellular immune responses. The chosen protein sequence was confirmed to be a reliable model for vaccine development by achieving a confidence level of 100% and a coverage of 79% and 85% for vaccine (i) and (ii) respectively. The study also used computational methods to evaluate the structural properties of the chosen protein sequence. This study shows a Z score of -3.6 and -4.57 for the two vaccines with a recognizable black area in the X-ray region, denoting a score that is acceptable by the algorithm. The molecule's low Z score emphasizes the fact that only the N-terminal portion has positive properties. The incorporation of built-in adjuvants has emerged as a promising way to improve the presentation of pathogen epitopes to the immune system, in addition to structural stability and multiepitope design. Adjuvants are incorporated into epitope-based vaccination designs to aid innate immune responses, which are necessary for the generation of a powerful adaptive immunological response. Adjuvants can be shown on particular biomaterials like lipopeptides or can fuse with proteins that are known to activate Toll-like receptors (TLRs). TLR ligands are attractive candidates for vaccine development because TLRs have the capacity to trigger both innate and adaptive immune responses (Yokomine, M. et al., 2017).. Further highlighting its potential is the discovery that TLR4 functions as a key molecule in cellular and

humoral defense against the Human Papillomavirus (HPV) (Yokomine, M. et al., 2017). A promising method for focusing on particular antigenic areas of HPV 16 and reducing the possibility of side effects is epitope-based vaccine design (Pumchan, A. et al., 2020). It is feasible to direct the immune response exactly to the most important targets by choosing immunogenic epitopes from the virus (Pumchan, A. et al., 2020). These epitopes can cause potent cellular and humoral immune responses when they are included in a vaccine design.

Chapter 5

Conclusion

Immunoinformatic techniques have become essential tools for accelerating the creation of vaccines that are more effective and efficient while taking less time to produce. The goal of this Seadawy study from 2022 is to develop a multi-epitope HPV16 vaccination. This vaccine contains L1, a significant capsid protein known to increase viral pathogenicity and disease severity. The capability of the multi-epitope protein to stimulate cytotoxic T lymphocytes (CTLs) and helper T lymphocytes (HTLs) was investigated *in silico* (Seadawy, M. G. 2022). A possible candidate for immunizing against strains that contain the L1 protein is the suggested multi-epitope vaccination. While *in silico* evaluations point to the created candidate vaccine having a high level of efficacy, additional experimental study is necessary to confirm the expected results. This study emphasizes how crucial immunoinformatics is to modernizing vaccine development methods for better efficacy and affordability.

References

- Awad, N., Mohamed, R. H., Ghoneim, N. I., Elmehrath, A. O., & El-Badri, N. (2022). Immunoinformatics approach of epitope prediction for SARS-CoV-2. *Journal of Genetic Engineering and Biotechnology*, 20. <https://doi.org/10.1186/s43141-022-00344-1>
- Elalouf, A. (2023). In-silico Structural Modeling of Human Immunodeficiency Virus Proteins . *Biomedical Engineering and Computational Biology*, 14. <https://doi.org/10.1177/11795972231154402>
- Ghorban Hosseini, N., Tebianian, M., Farhadi, A., Hossein Khani, A., Rahimi, A., Mortazavi, M., Hosseini, S. Y., Taghizadeh, M., Rezaei, M., & Mahdavi, M. (2017). In Silico Analysis of L1/L2 Sequences of Human Papillomaviruses: Implication for Universal Vaccine Design. *Viral Immunology*, 30. <https://doi.org/10.1134%2FS0026893321050150>
- Kolla, H. B., Tirumalasetty, C., Sreerama, K., & Ayyagari, V. S. (2021). An immunoinformatics approach for the design of a multi-epitope vaccine targeting super antigen TSST-1 of *Staphylococcus aureus*. *Journal of Genetic Engineering and Biotechnology*, 19. <https://doi.org/10.1186/s43141-021-00160-z>
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, 12. <https://doi.org/10.1038/nprot.2016.169>
- Lei, Y., Zhao, F., Shao, J., Li, Y., Li, S., Chang, H., & Zhang, Y. (2019). Application of built-in adjuvants for epitope-based vaccines. *PeerJ*, 6. <https://doi.org/10.7717%2Fpeerj.6185>
- Kumar, A., Sahu, U., Kumari, P., Dixit, A., & Khare, P. (2022). Designing of multi-epitope chimeric vaccine using immunoinformatic platform by targeting oncogenic strain HPV 16 and 18 against cervical cancer. *Scientific Reports*, 12. <https://doi.org/10.1038/s41598-022-13442-4>

- M. Elshafei, A., A. Mahmoud, N., & A. Almofti, Y. (2022). Development of Multi-Epitopes Vaccine against Human Papilloma Virus16 Using the L1 and L2 Proteins as Immunogens. *Biosciences Biotechnology Research Asia*, 19. <https://doi.org/10.13005/bbra/3032>
- Maisner, A., Ringel, M., Behner, L., Heiner, A., Sauerhering, L., & Maisner, A. (2018). The Journal of Infectious Diseases Replication of a Nipah Virus Encoding a NuclearRetained Matrix Protein. *The Journal of Infectious Diseases* ®, 221. <https://doi.org/10.1093/infdis/jiz440>
- Mahmoudvand, S., Shokri, S., Makvandi, M., Taherkhani, R., Rashno, M., Jalilian, F. A., & Angali, K. A. (2022). In silico prediction of T-cell and B-cell epitopes of human papillomavirus type 16 L1 protein. *Biotechnology and Applied Biochemistry*, 69. <https://doi.org/10.1002/bab.2128>
- Maleki, A., Russo, G., Parasiliti Palumbo, G. A., & Pappalardo, F. (2021). In silico design of recombinant multi-epitope vaccine against influenza A virus. *BMC Bioinformatics*, 22. <https://doi.org/10.1186/s12859-022-04581-6>
- Mohammadi, Y., Nezafat, N., Negahdaripour, M., Eskandari, S., & Zamani, M. (2023). In silico design and evaluation of a novel mRNA vaccine against BK virus: a reverse vaccinology approach. *Immunologic Research*, 71. <https://doi.org/10.1007/s12026-022-09351-3>
- Parvizpour, S., Pourseif, M. M., Razmara, J., Rafi, M. A., & Omid, Y. (2020). Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discovery Today*, 25. <https://doi.org/10.1016/J.DRUDIS.2020.03.006>
- Pumchan, A., Krobthong, S., Roytrakul, S., Sawatdichaikul, O., Kondo, H., Hirono, I., Areechon, N., & Unajak, S. (2020). Novel Chimeric Multiepitope Vaccine for Streptococcosis Disease in Nile Tilapia (*Oreochromis niloticus* Linn.). *Scientific reports*, 10. <https://doi.org/10.1038/s41598-019-57283-0>

- Rahman, N., Ali, F., Basharat, Z., Shehroz, M., Khan, M. K., Jeandet, P., Nepovimova, E., Kuca, K., & Khan, H. (2020). Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: An immunoinformatics approach. *Vaccines*, 8. <https://doi.org/10.3390/vaccines8030423>
- Rodríguez-Ruiz, H. A., Garibay-Cerdenares, O. L., Illades-Aguiar, B., Montaña, S., Jiang, X., & Leyva-Vázquez, M. A. (2019). In silico prediction of structural changes in human papillomavirus type 16 (HPV16) E6 oncoprotein and its variants. *BMC Molecular and Cell Biology*, 20. <https://doi.org/10.1186/s12860-019-0217-0>
- Sanami, S., Rafieian-Kopaei, M., Dehkordi, K. A., Pazoki-Toroudi, H., Azadegan-Dehkordi, F., Mobini, G. R., Alizadeh, M., Nezhad, M. S., Ghasemi-Dehnoo, M., & Bagheri, N. (2022). In silico design of a multi-epitope vaccine against HPV16/18. *BMC Bioinformatics*, 23. <https://doi.org/10.1186/s12859-022-04784-x>
- Shahab, M., Guo, D., Zheng, G., & Zou, Y. (2023). *Design of a Novel and Potent Multi-Epitope Chimeric Vaccine against Human Papillomavirus (HPV): An Immunoinformatics Approach*. <https://doi.org/10.3390/biomedicines>
- Skowron, K., Bauza-Kaszewska, J., Grudlewska-Buda, K., Wiktorczyk-Kapischke, N., Zacharski, M., Bernaciak, Z., & Gospodarek-Komkowska, E. (2022). Nipah Virus— Another Threat From the World of Zoonotic Viruses. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/FMICB.2021.811157/BIBTEX>
- Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., & Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. *Journal of Biomedical Informatics*, 53. <https://doi.org/10.1016/J.JBI.2014.11.003> 31.

- Stanzione, F., Giangreco, I., & Cole, J. C. (2021). Use of molecular docking computational tools in drug discovery. *Progress in Medicinal Chemistry*, 60. <https://doi.org/10.1016/BS.PMCH.2021.01.004>
- Sun, B., Jia, L., Liang, B., Chen, Q., & Liu, D. (2018). Phylogeography, Transmission, and Viral Proteins of Nipah Virus. *Virologica Sinica*, 33. <https://doi.org/10.1007/S12250-018-0050-1> 33.
- Sunita, Sajid, A., Singh, Y., & Shukla, P. (2020). Computational tools for modern vaccine development. In *Human Vaccines and Immunotherapeutics*, 16. <https://doi.org/10.1080/21645515.2019.1670035>
- Vita, R., Mahajan, S., Overton, J. A., Kumar Dhanda, S., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2018). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, 47. <https://doi.org/10.1093/nar/gky1006>
- Tan, C., Zhu, F., Pan, P., Wu, A., & Li, C. (2023). Development of multi-epitope vaccines against the monkeypox virus based on envelope proteins using immunoinformatics approaches. *Frontiers in Immunology*, 14. <https://doi.org/10.3389/fimmu.2023.1112816>
- Ye, W., Schmitt, N. C., Ferris, R. L., & Allen, C. T. (2020). Improving responses to immunotherapy in head and neck squamous cell carcinoma. *Improving the Therapeutic Ratio in Head and Neck Cancer*, 6. <https://doi.org/10.1016/B978-0-12-817868-3.00006-8> 35.
- Yokomine, M., Matsueda, S., Kawano, K., Sasada, T., Fukui, A., Yamashita, T., Komatsu, N., Shichijo, S., Tasaki, K., Matsukuma, K., Itoh, K., Kamura, T., & Ushijima, K. (2017). Enhancement of humoral and cell mediated immune response to HPV16 L1-derived peptides subsequent to vaccination with prophylactic bivalent HPV L1 virus-like particle vaccine in healthy females. *Experimental and therapeutic medicine*, 13. <https://doi.org/10.3892/etm.2017.4150>