In-Silico Vaccine Design Construction of Human Papillomavirus

By

Saurav Mondal Shuvo
ID: 19346044

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
May 2025

Declaration

It is hereby declared that

1. The thesis submitted is my 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 and Signature:

Saurav Mondal Shuvo
Student ID: 19346044

Approval

The thesis titled “In-Silico Vaccine Design Construction of Human Papillomavirus” submitted by Saurav Mondal Shuvo (ID: 19346044) of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised

Supervisor:

_____

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

Approved By

Dean:

_____

A F M Yusuf Haider, PhD
Acting Dean, School of Pharmacy
Professor, Department of Mathematics and Natural Sciences
BRAC University

Ethics Statement

The thesis was completed following all the required ethical standards. There was no involvement of any human or animal trial.

Abstract

Human Papillomavirus (HPV) is a leading cause of oropharyngeal cancer, causes a significant health risk worldwide. Traditional vaccine development processes are often time-consuming and costly, making alternative approaches necessary. This study focuses on the use of computational methods to design a vaccine against HPV-related oropharyngeal cancer. Using in-silico techniques to find CTL, HTL and B CELL of HPV proteins, known as epitopes, were identified for their ability to stimulate the immune system. These epitopes were carefully selected based on their potential to generate a strong immune response while ensuring safety for human use. The chosen epitopes were assembled into a multi-epitope vaccine construct, enhanced with adjuvants to improve immune activation. The vaccine's structure was modeled and optimized for stability and effectiveness. Molecular docking studies were conducted to evaluate the interaction between the vaccine and human immune receptors, demonstrating a strong potential for immune system activation. Immune simulation results suggested that the vaccine could trigger both cellular and humoral responses, offering long-term protection against HPV-induced oropharyngeal cancer. This study highlights the potential of in-silico vaccine design as a faster and more cost-effective method for developing vaccines, with promising implications for preventing HPV-related oropharyngeal cancer.

Keyword: Human Papillomavirus, CTL, HTL, B CELL, Molecular docking, Immune simulation

Acknowledgement

I would like to express my sincere gratitude to my supervisor Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, BRAC University for his continuous support, patience, motivation and valuable feedback. His guidance and support helped me all the time while performing this research.

Table of Contents

List of Figures

List of Tables

List of Acronyms

CTL	Cytotoxic T- Lymphocyte
GRAVY	The Grand Average of Hydropathicity
HPV	Human Papillomavirus
HTL	Helper T-Lymphocyte
MHC	Major Histocompatibility complex

Chapter 1

Introduction

Oropharyngeal cancer is a type of cancer that develops in the back of the throat, including the tonsils and base of the tongue. One of the main causes of this cancer is Human Papillomavirus (HPV), particularly high-risk types such as HPV-16 (Burd & Dean, 2016). Over the years, Number of HPV-related oropharyngeal cancer has increased, especially among younger individuals and those who do not smoke or drink alcohol. HPV is a common virus that spreads through close skin-to-skin contact, including oral transmission (Burd & Dean, 2016). When the virus infects the cells of the throat, it can cause changes that lead to abnormal cell growth and, over time, cancer. Unlike other throat cancers that are mainly linked to smoking and heavy alcohol use, HPV-related oropharyngeal cancer is primarily caused by the virus itself (Munger et al., 2004). People with HPV-related oropharyngeal cancer may experience symptoms such as a persistent score throat, difficulty swallowing, ear pain and a lump in the neck. Early diagnosis is important because HPV-positive oropharyngeal cancers usually respond well to treatment, which may include surgery, radiation, and chemotherapy (Munger et al., 2004). Preventing HPV infection is one of the best ways to minimize the risk of this cancer. The HPV vaccine has been proven effective in protecting against high-risk strains of the virus, lowering the chances of infection and related cancers (Moody & Laimins, 2010). Scientists are also exploring new ways to develop vaccines using advanced computer-based (in-silico) techniques. These modern approaches help design vaccines more quickly and accurately, offering new possibilities for preventing HPV-related oropharyngeal cancer in the future (Moody & Laimins, 2010).

1.1 Genomic Structure of Human Papillomavirus

The Human Papillomavirus (HPV) genome is made of circular, double-stranded DNA with about 8,000 DNA pairs, divided into three key regions:

1. Early Region: This section contains genes that help the virus grow and infect cells. The E6 and E7 genes play a role in disrupting the body's defense system, which can lead to cancer.
2. Late Region: This part includes genes that form the outer protective layer (capsid) of the virus. The L1 gene is important for vaccine development (Taberna et al., 2017).
3. Long Control Region: A non-coding section that regulates the virus's activity and helps in its replication.

Since HPV cannot reproduce on its own, it depends on human cells to make more copies. Some HPV types cause harmless warts, while others increase the risk of cancer (Taberna et al., 2017).

1.2 Lifecycle and Pathogenesis of Human Papillomavirus

Human Papillomavirus (HPV) go into the body through small cuts in the skin or mucosa. It infects basal cells, multiplies, and matures as these cells move upward (Stein et al., 2015). When the infected cells shed, the virus is released and can spread to new hosts. HPV has different types: Low-risk strains cause harmless warts on the skin and genitals. High-risk strains can disrupt normal cell function, increasing the risk of cervical and other cancers (Stein et al., 2015). Most infections disappear naturally, but some persist and lead to complications. Vaccination helps prevent infection from the most harmful HPV types.

Chapter 2

Methodology

The following outlines that were followed in developing and design the vaccine.

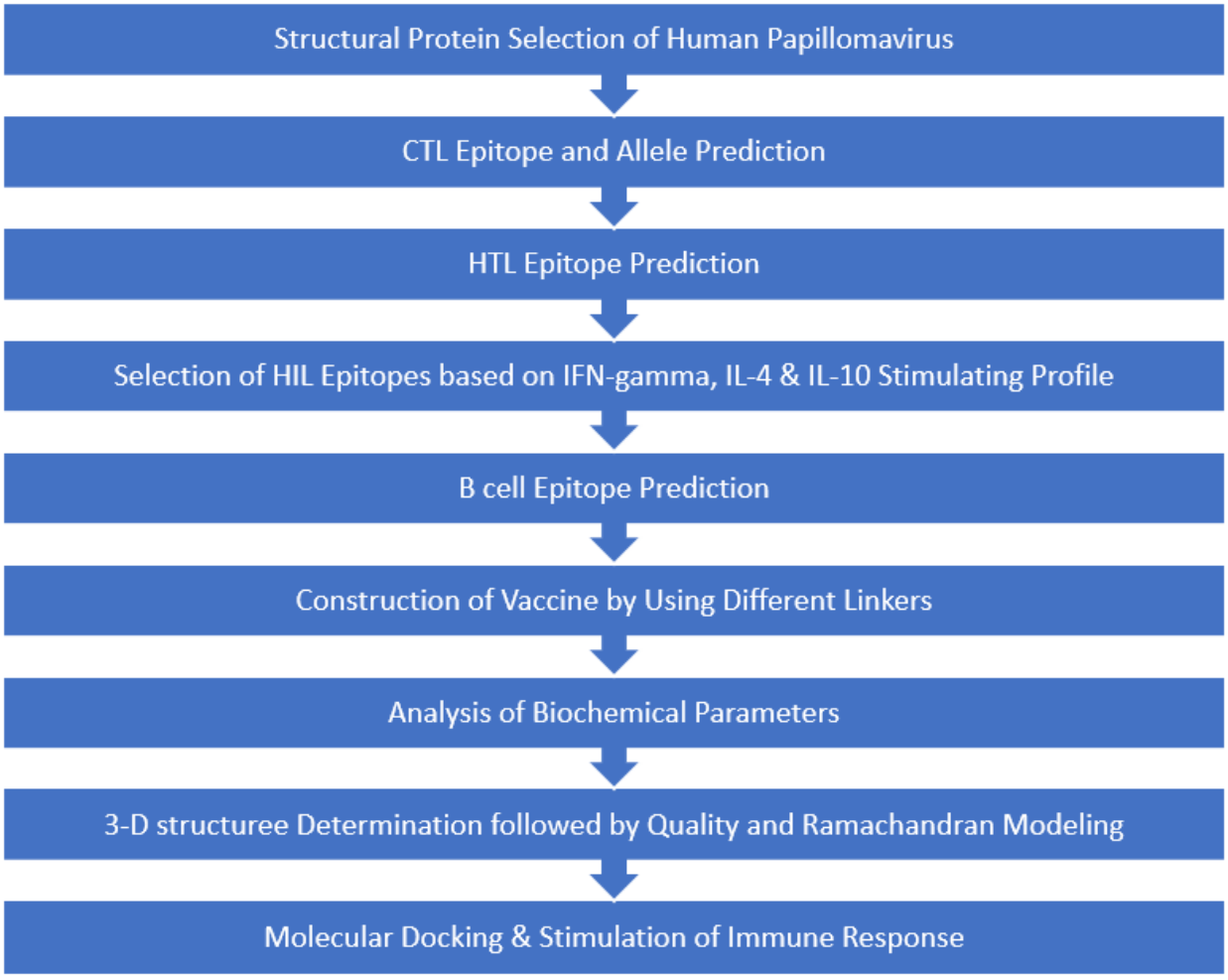


Figure 1 : Step-by-Step Schematic Overview of In-Silico vaccine Development

2.1 Selecting a Suitable Protein Sequence

To select the protein sequence, the Uniprot protein database was utilized. Various protein sequences were retrieved from this database, which contains a total of 3,450 Human Papillomavirus protein sequences. Among these, 45 sequences are reviewed in the Swiss-Prot section, while the remaining 3,405 are unreviewed in the TrEMBL section (Bairoch et al., 2005). To ensure the selection of a suitable protein sequence, antigenicity analysis was performed. The sequences were downloaded in FASTA format for this purpose. Antigenicity evaluation was carried out using the VaxiJen v.2.0 server, where "virus" was set as the target organism (Doytchinova & Flower, 2007). A threshold value of 0.5 was applied to minimize the risk of false-positive results. The sequence with the highest antigenicity score was chosen for further analysis (Bairoch et al., 2005).

2.2 Identification of CTL Epitopes

Cytotoxic T lymphocyte (CTL) epitopes play an important role in eliciting immune responses within host cells. To identify these epitopes, the NetCTL 1.2 tool was utilized (Williams & Bevan, 2007). This tool evaluates three key parameters: Transport efficiency, MHC-I binding affinity and C-terminal cleavage. The prediction of proteasomal C-terminal cleavage is carried out using artificial neural networks, while TAP transport efficiency is determined through a weight matrix approach (Williams & Bevan, 2007). For epitope prediction, in FASTA format the protein sequence was provide, and a selection threshold of 0.75 was applied. This ensured that only the most relevant epitopes were identified for further evaluation.

2.3 Assessment of CTL Epitopes

Evaluation of the identified CTL epitopes was conducted applying NetMHCpan 4.1, a tool designed to predict MHC-I binding alleles. This server employs an artificial neural network to analyze MHC molecules (Reynisson et al., 2020). In this study, 9-mer peptides were chosen, and all human leukocyte antigen (HLA) alleles were included in the assessment. A binding affinity threshold of 0.5 was set for strong binding peptides, whereas a threshold of 2 was used for weak binding peptides. The tool ranked the results based on binding affinity scores, and only peptides exhibiting strong binding interactions were selected for further analysis (Reynisson et al., 2020). To assure the safety and efficacy of the selected peptides, additional evaluations were performed. The antigenicity of the peptides was assessed using the VaxiJen v.2.0 server. The AllerTop v.20 server was employed to determine allergenicity, while toxicity screening was conducted using the ToxinPred server.

2.4 Identification of HTL Epitopes

Helper T lymphocyte (HTL) epitopes play an important role in the adaptive immune response. To identify these epitopes, the NetMHC-IIpan 4.0 server was utilized, which employs artificial neural networks for prediction (Reynisson et al., 2020). The sequence of protein was input in FASTA format, and peptides of 15 amino acids in length were selected. A maximum of 20 alleles were chosen per submission, and this process was repeated until all necessary alleles were included. Binding affinity thresholds were set at 1% for strong binders and 5% for weak binders. The results displayed the alleles that met these criteria, and only those with high binding affinity were selected for further analysis (Reynisson et al., 2020).

2.5 Assessment of HTL Epitopes

HTL epitopes contribute to immune responses by triggering cytokine release, particularly interferon-gamma (IFN-γ), interleukin-4 (IL-4), and interleukin-10 (IL-10) (Wong et al., 2001). These cytokines are essential in determining the immune system's reaction following vaccination. IFN-γ is responsible for activating major histocompatibility complex (MHC) molecules, and its induction by HTL epitopes was assessed using the IFNepitope server. IL-4 plays a key role in B cell maturation and proliferation, and its presence was evaluated through the IL-4 pred server. IL-10 is vital for immune homeostasis, and its induction was checked using the IL-10 pred server. Epitopes that meet all three criteria were then assessed for antigenicity, allergenicity and toxicity. Only those that were antigenic, non-allergenic, and non-toxic were considered for further study (Larsen et al., 2007).

2.6 Prediction and Evaluation of B Cell Epitopes

Prediction of B cell epitopes was conducted using the IEDB Resource Analysis server. The "Bepipred Linear Prediction 2.0" method was selected for this purpose (Ponomarenko et al., 2010). The protein sequence was submitted, and the output provided both a graphical representation and a tabular dataset based on antigenicity. A threshold of 0.5 was applied to select relevant epitopes. These epitopes were then further examined for antigenicity, allergenicity and toxicity, following the same criteria as the HTL epitopes. Only those that met the necessary requirements were considered for subsequent analysis (Ponomarenko et al., 2010).

2.7 Vaccine Construction using linkers and Antigenicity Assessment

A vaccine was designed in silico by linking various epitopes with specific linkers. Unlike conventional vaccine designs that incorporate four linkers, this study utilized three. The primary protein sequence, serving as an adjuvant, was linked to CTL epitopes using the EAAAK linker. The connection between CTL and HTL epitopes was established through the GPGPG linker, while the KK linker was employed to connect HTL epitopes with B cell epitopes (Garg et al., 2016). Following the construction of the vaccine, its antigenicity was evaluated using the VaxiJen v2.0 server to determine its potential as an effective immunogen (Doytchinova & Flower, 2007).

2.8 Biochemical Analysis of the Prepared Vaccine

The physicochemical properties of the constructed vaccine were assessed using the ProtParam server (Garg et al., 2016). This tool provided crucial insights, including the total number of amino acids, molecular of weight, index of instability, grand average of hydropathicity (GRAVY), and other physicochemical attributes. These parameters play a vital role in evaluating the vaccine's stability and potential efficacy (Garg et al., 2016).

2.9 Allergenicity and Toxicity Prediction of the Prepared Vaccine

To ensure the vaccine's safety, its allergenicity was examined using an Allergenonline server, where the sequence was submitted in plain text format (Goodman et al., 2016). Additionally, toxicity assessment was performed using the T3DB database, which helped determine whether the vaccine construct exhibited toxic properties (Wishart et al., 2015).

2.10 3D Structural Modeling of the Prepared Vaccine

The 3D structure of the vaccine was predicted using the Phyre2 server. The vaccine sequence was provided in plain format, and Phyre2 generated both primary and tertiary structures (Kelley et al., 2015). The model's confidence and coverage were key factors in evaluating its structural reliability. The resulting PDB file was received via email and visualized using PyMOL software for structural analysis (Kelley et al., 2015).

2.11 Validation of Ramachandran Plots

To validate the structural integrity of the modeled vaccine, a Ramachandran plot was generated using the Swiss Model Expasy structure assessment tool (Kiefer et al., 2009). The PDB file received from Phyre2, and the plot displayed the phi and psi angles of the polypeptide chain. This analysis ensured the proper folding and stability of the protein structure (Kiefer et al., 2009).

2.12 Z-Score Analysis for 3D- Model

The ProSA-web server was applied to determine the Z-score of the constructed vaccine model. This tool evaluated the overall quality of the 3D structure and analyzed the energy distribution throughout the model (Wiederstein & Sippl, 2007). Upon uploading the PDB file, the output included a Z-score, a Z-score versus residue count graph, and a knowledge-based energy versus sequence position graph. These analyses were crucial in determining the structural reliability of the vaccine (Wiederstein & Sippl, 2007).

2.13 Molecular Docking Analysis

Molecular docking was conducted to evaluate the binding interactions between the designed vaccine and a suitable human toll-like receptor (TLR). The ClusPro server (https://cluspro.bu.edu/) was utilized for this purpose (Kozakov et al., 2017). The PDB file of the TLR was obtained from the RCSB protein Data Bank, while the vaccine's PDB file was uploaded as the ligand molecule. ClusPro generated ten docked complexes, and the one with the highest binding energy score was selected for further analysis (Kozakov et al., 2017).

2.14 Simulation of Immune Response

The immunogenic potential of the designed vaccine was evaluated using the C-ImmSim 10.1 server, which simulates immune responses (Rapin et al., 2011). The vaccine sequence was submitted in FASTA format, and three injections were simulated at time steps of 1, 84 and 168, with each time steps representing eight hours. This simulation generated graphical representations of antigen levels, lymphocyte responses, and immunoglobulin production, providing insight into both humoral and cellular immune responses following vaccine administration (Rapin et al., 2011).

2.15 Methodological Considerations

This study focused on the in-silico development of a vaccine candidate through computational screening, prediction, and structural analyses (Chaturvedi et al., 2011). While this approach is efficient in identifying potential vaccine candidates, it does not provide direct insights into the efficiency and safety of the designed vaccine in biological systems. These computational tools rely entirely on statistical and machine learning approaches and are well-regarded for their ability to analyze and model molecular interactions involved in antigen processing and presentation (Kevin et al., 2013). This method is particularly useful as it not only aids in identifying proteins that have the potential to act as antigens but also enables a comprehensive examination of the pathogen's entire genome. Therefore, further experimental validation is required to confirm its effectiveness as a potential vaccine against Human Papillomavirus (Chaturvedi et al., 2011).

Chapter 3

Result

3.1 Selection of Protein Sequences Based on Antigenicity

The capsid protein of Human Papillomavirus (HPV) was chosen based on its strong antigenic properties, as it demonstrated the highest antigenicity during the initial screening process (Doytchinova & Flower, 2007). The amino acid sequence of the selected protein is provided below:

MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVETL
KEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIRT
GLIKGSGTAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGL
ISLQVKQKGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVD
MVFASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGI
EIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGAD
CIMLSGETAKGDYPLEAVRMQHLIAREAEAAIYHLQLFEELRRLAPITSDPTEATAVGAV
EASFKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPV
QEAWAEDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPVP

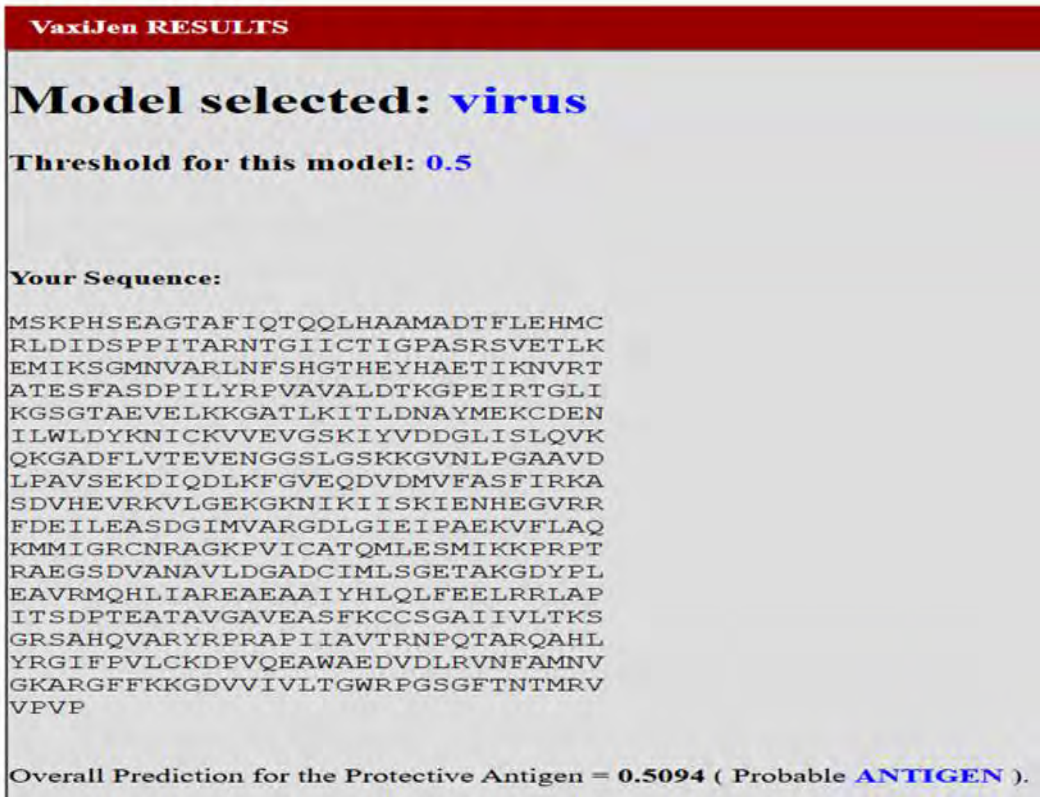


Figure 2 : Antigenicity score of the Protein Sequence

The Antigenicity of protein sequence was 0.5094 (Probable Antigen).

3.2 Identification and Assessment of CTL Epitopes

The NetCTL 1.2 tool was used to identify cytotoxic T lymphocyte (CTL) epitopes by analyzing three factors: TAP transport efficiency, MHC-I binding affinity, and C-terminal cleavage. Proteasomal cleavage was predicted using artificial neural networks, while TAP transport was assessed with a weight matrix method (Reynisson et al., 2020). The protein sequence, provided in FASTA format was screened with a selection threshold of 0.75 to ensure the identification of the most relevant epitopes for further analysis.

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

458	ID	FASTA	pep	QTARQAHLY	aff	0.7405	aff_rescale	3.1442	cle 0.4943	tap	2.7880 COMB	3.3578 <-E
436	ID	FASTA	pep	RSAHQVARY	aff	0.4900	aff_rescale	2.0804	cle 0.9220	tap	3.1960 COMB	2.3785 <-E
167	ID	FASTA	pep	VVEVGSKIY	aff	0.1992	aff_rescale	0.8459	cle 0.8563	tap	3.0820 COMB	1.1285 <-E
345	ID	FASTA	pep	GSDVANAVL	aff	0.1931	aff_rescale	0.8199	cle 0.9564	tap	0.4730 COMB	0.9870 <-E
22	ID	FASTA	pep	MADTFLEHM	aff	0.1862	aff_rescale	0.7906	cle 0.9585	tap	0.3820 COMB	0.9535 <-E
5	ID	FASTA	pep	HSEAGTAFI	aff	0.1887	aff_rescale	0.8011	cle 0.8763	tap	0.1950 COMB	0.9423 <-E
482	ID	FASTA	pep	WAEDVDLRV	aff	0.1754	aff_rescale	0.7448	cle 0.8237	tap	0.3500 COMB	0.8858 <-E
143	ID	FASTA	pep	TLDNAYMEK	aff	0.1631	aff_rescale	0.6926	cle 0.9530	tap	0.2960 COMB	0.8504 <-E
75	ID	FASTA	pep	NFSHGTHEY	aff	0.1099	aff_rescale	0.4666	cle 0.9742	tap	3.1730 COMB	0.7714 <-E
236	ID	FASTA	pep	DVDMVFASF	aff	0.1257	aff_rescale	0.5338	cle 0.8467	tap	2.1320 COMB	0.7674 <-E
382	ID	FASTA	pep	AREAEAAIY	aff	0.1094	aff_rescale	0.4645	cle 0.9497	tap	3.1400 COMB	0.7639 <-E
175	ID	FASTA	pep	YVDDGLISL	aff	0.1316	aff_rescale	0.5585	cle 0.9786	tap	0.9060 COMB	0.7506 <-E
97	ID	FASTA	pep	SFASDPILY	aff	0.1072	aff_rescale	0.4552	cle 0.6676	tap	2.9950 COMB	0.7051
189	ID	FASTA	pep	GADFLVTEV	aff	0.1257	aff_rescale	0.5335	cle 0.9710	tap	0.1130 COMB	0.6848
140	ID	FASTA	pep	LKITLDNAY	aff	0.0853	aff_rescale	0.3621	cle 0.8716	tap	3.0570 COMB	0.6457
21	ID	FASTA	pep	AMADTFLEH	aff	0.1202	aff_rescale	0.5104	cle 0.9265	tap	-0.0990 COMB	0.6444
248	ID	FASTA	pep	ASDVHEVRK	aff	0.1159	aff_rescale	0.4922	cle 0.7957	tap	0.5740 COMB	0.6402
158	ID	FASTA	pep	WLDYKNICK	aff	0.1339	aff_rescale	0.5683	cle 0.3520	tap	0.3290 COMB	0.6376
362	ID	FASTA	pep	SGETAKGDY	aff	0.1051	aff_rescale	0.4463	cle 0.2801	tap	2.6150 COMB	0.6190
111	ID	FASTA	pep	ALDTKGPEI	aff	0.1018	aff_rescale	0.4322	cle 0.9572	tap	0.5690 COMB	0.6042
286	ID	FASTA	pep	ASDGIMVAR	aff	0.1045	aff_rescale	0.4438	cle 0.5739	tap	1.3260 COMB	0.5962

Figure 3: NetCTL prediction for CTL epitopes

The identified CTL epitopes were analyzed using NetMHCpan 4.1 to predict MHC-I binding alleles. This tool, based on an artificial neural network, assessed 9-mer peptides across all HLA alleles. Strong binding peptides had a threshold of 0.5, while weak binders were set at 2. Peptides with high binding affinity were selected for further study. To ensure their safety and effectiveness, antigenicity was evaluated with VaxiJen v.2.0, allergenicity with AllerTop v.2.0, and toxicity with ToxinPred (Wong et al., 2001). Vaccine was prepared by using 2 standard CTL epitopes.

Table 1 : Determination of CTL Epitopes with Antigenicity, Allergenicity and Toxicity

CTL EPITOPE	ANTIGENICITY	ALLERGENICITY	TOXICITY
QTARQAHLY	ANTIGEN	NON-ALLERGEN	NON-TOXIN
RSAHQVARY	NON-ANTIGEN	NON-ALLERGEN	NON-TOXIN
VVEVGSKIY	ANTIGEN	NON-ALLERGEN	NON-TOXIN
YVDDGLISL	ANTIGEN	ALLERGEN	NON-TOXIN
TLDNAYMEK	NON-ANTIGEN	NON-ALLERGEN	NON-TOXIN
DVDMVFASF	ANTIGEN	ALLERGEN	NON-TOXIN

3.3 Identification and Assessment of HTL Epitopes

HTL epitopes play an important role in immune responses by stimulating cytokine release, including IFN-gamma, IL-4 and IL-10, which influence the immune system after vaccination. IFN-γ activates MHC molecules, and its induction was analyzed using the IFNepitope server. IL-4 supports B cell development and was evaluated through the IL-4 pred server, while IL-10, essential for immune regulation, was assessed using the IL-10 pred server. Epitopes fulfilling all three criteria were further examined for antigenicity, allergenicity, and toxicity (Larsen et al., 2007). Only antigenic, non-allergenic and non-toxic epitopes were selected for further study. Vaccine was prepared by using 2 standard HTL epitopes.

Table 2: Determination of HTL Epitopes with Antigenicity, Allergenicity, Toxicity, IL-4, IL-10 and IFN- Gamma

HTL EPITOPE	ANTIGENICITY	ALLERGENICITY	TOXICITY	IL4	IL10	IFN GAMMA
EAVRMQHLIAREAEA	NON-ANTIGEN	ALLERGEN	TOXIN	NON-INDUCER	INDUCER	POSITIVE
AVRMQHLIAREAEAA	ANTIGEN	NON-ALLERGEN	NON-TOXIN	NON-INDUCER	INDUCER	POSITIVE
FEELRRLAPITSDPT	ANTIGEN	ALLERGEN	NON-TOXIN	INDUCER	NON-INDUCER	NEGATIVE
GTHEYHAETIKNVRT	ANTIGEN	NON-ALLERGEN	NON-TOXIN	INDUCER	INDUCER	NEGATIVE
THEYHAETIKNVRTA	NON-ANTIGEN	NON-ALLERGEN	NON-TOXIN	INDUCER	INDUCER	NEGATIVE
GTAEVELKKGATLKI	ANTIGEN	ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE
TAEVELKKGATLKIT	NON-ANTIGEN	NON-ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE
GGSLGSKKGVNLPGA	ANTIGEN	ALLERGEN	NON-TOXIN	NON-INDUCER	NON-INDUCER	NEGATIVE
EVRKVLGEKGKNIKI	NON-ANTIGEN	ALLERGEN	NON-TOXIN	INDUCER	INDUCER	POSITIVE
VRKVLGEKGKNIKII	ANTIGEN	NON-ALLERGEN	TOXIN	NON-INDUCER	INDUCER	POSITIVE
QVARYRPRAPIIAVT	ANTIGEN	ALLERGEN	NON-TOXIN	INDUCER	INDUCER	POSITIVE
VARYRPRAPIIAVTR	ANTIGEN	NON-ALLERGEN	NON-TOXIN	INDUCER	INDUCER	POSITIVE
TATESFASDPILYRP	ANTIGEN	NON-ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE
ATESFASDPILYRPV	NON-ANTIGEN	NON-ALLERGEN	NON-TOXIN	INDUCER	INDUCER	POSITIVE
TESFASDPILYRPVA	ANTIGEN	ALLERGEN	NON-TOXIN	INDUCER	NON-INDUCER	POSITIVE

ESFASDPILYRPVAV	NON-ANTIGEN	NON-ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE
YRPVAVALDTKGPEI	ANTIGEN	NON-ALLERGEN	NON-TOXIN	NON-INDUCER	INDUCER	NEGATIVE
RPVAVALDTKGPEIR	ANTIGEN	ALLERGEN	NON-TOXIN	NON-INDUCER	INDUCER	NEGATIVE

KNIKIISKIENHEGV	NON-ANTIGEN	NON-ALLERGEN	TOXIN	NON-INDUCER	NON-INDUCER	POSITIVE
AETIKNVRTATESFA	ANTIGEN	ALLERGEN	NON-TOXIN	INDUCER	NON-INDUCER	NEGATIVE
ILYRPVAVALDTKGP	ANTIGEN	NON-ALLERGEN	NON-TOXIN	NON-INDUCER	INDUCER	NEGATIVE
EKGKNIKIISKIENH	NON-ANTIGEN	ALLERGEN	TOXIN	INDUCER	INDUCER	POSITIVE
KGKNIKIISKIENHE	ANTIGEN	NON-ALLERGEN	TOXIN	NON-INDUCER	INDUCER	POSITIVE
GKNIKIISKIENHEG	ANTIGEN	NON-ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE
YRPRAPIIAVTRNPQ	ANTIGEN	NON-ALLERGEN	NON-TOXIN	INDUCER	INDUCER	POSITIVE
RPRAPIIAVTRNPQT	ANTIGEN	ALLERGEN	NON-TOXIN	NON-INDUCER	INDUCER	POSITIVE
LFEELRRLAPITSDP	ANTIGEN	NON-ALLERGEN	NON-TOXIN	NON-INDUCER	NON-INDUCER	NEGATIVE
AHLYRGIFPVLCKDP	NON-ANTIGEN	ALLERGEN	NON-TOXIN	NON-INDUCER	NON-INDUCER	NEGATIVE
VDMVFASFIRKASDV	ANTIGEN	NON-ALLERGEN	TOXIN	NON-INDUCER	NON-INDUCER	POSITIVE
DMVFASFIRKASDVH	ANTIGEN	NON-ALLERGEN	TOXIN	INDUCER	INDUCER	NEGATIVE
KASDVHEVRKVLGEK	NON-ANTIGEN	ALLERGEN	TOXIN	INDUCER	NON-INDUCER	POSITIVE

3.4 Evaluation and Prediction of B cell Epitopes

B cell epitope prediction was carried out using the IEDB Resource Analysis server, utilizing the "Bepipred Linear Prediction 2.0" method. The protein sequence was submitted for analysis, generating both a graphical representation and a tabular dataset that ranked antigenic regions (Ponomarenko et al., 2010). To identify relevant epitopes, a threshold value of 0.5 was applied, ensuring the selection of antigenic regions with significant scores.

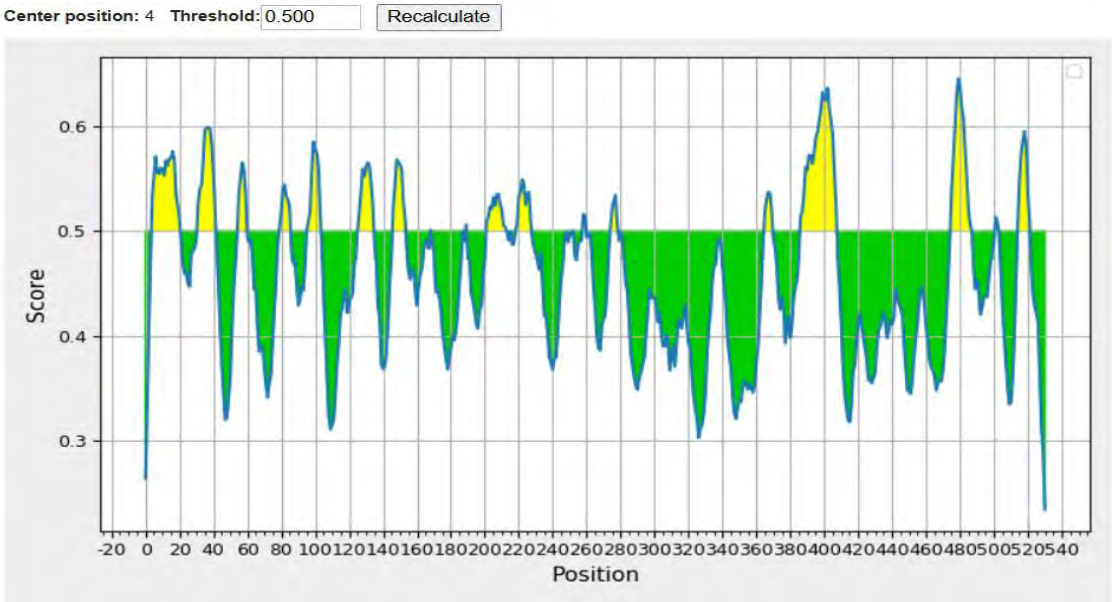


Figure 4 : Score vs Position graph of B Cell Epitopes

The selected epitopes underwent additional evaluation to assess their antigenicity, allergenicity and toxicity, using the same criteria applied to HTL epitopes. Only those that satisfied the required standards were chosen for further analysis. Vaccine was prepared by using 2 standard B CELL epitopes.

Table 3: Determination of B Cell Epitopes with Antigenicity, Allergenicity and Toxicity

B CELL EPITOPE	ANTIGENICITY	ALLERGENICITY	TOXICITY
HSEAGTAFIQTQQLHAA	ANTIGEN	NON-ALLERGEN	TOXIN
RLDIDSPPITA	NON-ANTIGEN	ALLERGEN	NON-TOXIN
THEYHAE	NON-ANTIGEN	ALLERGEN	NON-TOXIN
ESFASDPIL	NON-ANTIGEN	ALLERGEN	TOXIN
GSGTAEVELK	ANTIGEN	NON-ALLERGEN	NON-TOXIN
AYMEKCDE	NON-ANTIGEN	ALLERGEN	TOXIN
SLGSKKGVNLPG	ANTIGEN	NON-ALLERGEN	NON-TOXIN
AVSEKDIQD	ANTIGEN	ALLERGEN	TOXIN
AAIYHLQLFEELRRLAPITSDP	NON-ANTIGEN	NON-ALLERGEN	NON-TOXIN
DPVQEAWAEDVD	NON-ANTIGEN	ALLERGEN	TOXIN
WRPGSGFT	ANTIGEN	ALLERGEN	NON-TOXIN

3.5 Construction of Vaccine using linkers and antigenicity check

In this study, a multi-epitope vaccine was designed using an in-silico approach, where different epitopes were connected through specific linkers. Unlike traditional vaccine designs that typically incorporate four linkers, this research utilized only three. The core protein sequence, functioning as an adjuvant, was fused to CTL epitopes via the EAAAK linker. To establish a connection between CTL and HTL epitopes, the GPGPG linker was used, whereas the KK linker facilitated the attachment of HTL epitopes to B cell epitopes (Garg et al., 2016). Vaccine was prepared by using 2 CTL epitopes, 2 HTL epitopes, 2 B CELL epitopes. The constructed vaccine sequence:

MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVETLKEMIKS
GMNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIRTGLIKGSGTAEV
ELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKQKGADFLVTE
VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDMVFASFIRKASDVHEVRKVLG

EKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMIGRCNRAGKPVIC

ATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEAVRMQHLIAREAEAAIY

HLQLFEELRRLAPITSDPTEATAVGAVEASFKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNP

QTARQAHLYRGIFPVLCKDPVQEAWAEDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGS

GFTNTMRVVPVPEAAAKQTARQAHLYEAAAKVVEVGSKIYGPGPGVARYRPRAPIIAVTRGPGP

GYRPRAPIIAVTRNPQKKGSGTAEVELKKKSLGSKKGVNLPG

VaxiJen RESULTS

Model selected: virus

Threshold for this model: 0.5

Your Sequence:

MSKPHSEAGTAFIQTQQLHAAMADTFLEHMC
RLDIDSPPITARNTGIICTIGPASRSVETLK
EMIKSGMNVARLNFSHGTHEYHAETIKNVRT
ATESFASDPILYRPVAVALDTKGPEIRTGLI
KGSGTAEVELKKGATLKITLDNAYMEKCDEN
ILWLDYKNICKVVEVGSKIYVDDGLISLQVK
QKGADFLVTEVENGGSLGSKKGVNLPGAAVD
LPAVSEKDIQDLKFGVEQDVDMVFASFIRKA
SDVHEVRKVLGEKGKNIKIISKIENHEGVRR
FDEILEASDGIMVARGDLGIEIPAEKVFLAQ
KMMIGRCNRAGKPVICATQMLESMIKKPRPT
RAEGSDVANAVLDGADCIMLSGETAKGDYPL
EAVRMQHLIAREAEAAIYHLQLFEELRRLAP
ITSDPTEATAVGAVEASFKCCSGAIIVLTKS
GRSAHQVARYRPRAPIIAVTRNPQTARQAHL
YRGIFPVLCKDPVQEAWAEDVDLRVNFAMNV
GKARGFFKKGDVVIVLTGWRPGSGFTNTMRV
VPVPEAAAKQTARQAHLYEAAAKVVEVGSKI
YGPGPGVARYRPRAPIIAVTRGPGPGYRPRA
PIIAVTRNPQKKGSGTAEVELKKKSLGSKKG
VNLPG

Overall Prediction for the Protective Antigen = 0.5253 (Probable ANTIGEN).

Figure 5 : The Antigenicity score of Protein Sequence

The Antigenicity of protein sequence was 0.5253 (Probable Antigen).

3.6 Biochemical Analysis of the Constructed Vaccine

The physicochemical characteristics of the designed vaccine were analyzed using the ProtParam server, which offered valuable data on various properties. This analysis included determining the total amino acid count, molecular weight, instability index and the grand average of hydropathicity (GRAVY), among other key attributes. These factors are essential in assessing the vaccine's stability and potential effectiveness. The values of the physicochemical parameters were also between acceptable range: the therapeutic pI was 9.18 which was within the range of (7-10), the instability index (24.65) was lower than 40, and the GRAVY value was negative (-0.187). This suggests that the vaccine is stable and has a hydrophilic feature (Garg et al., 2016).

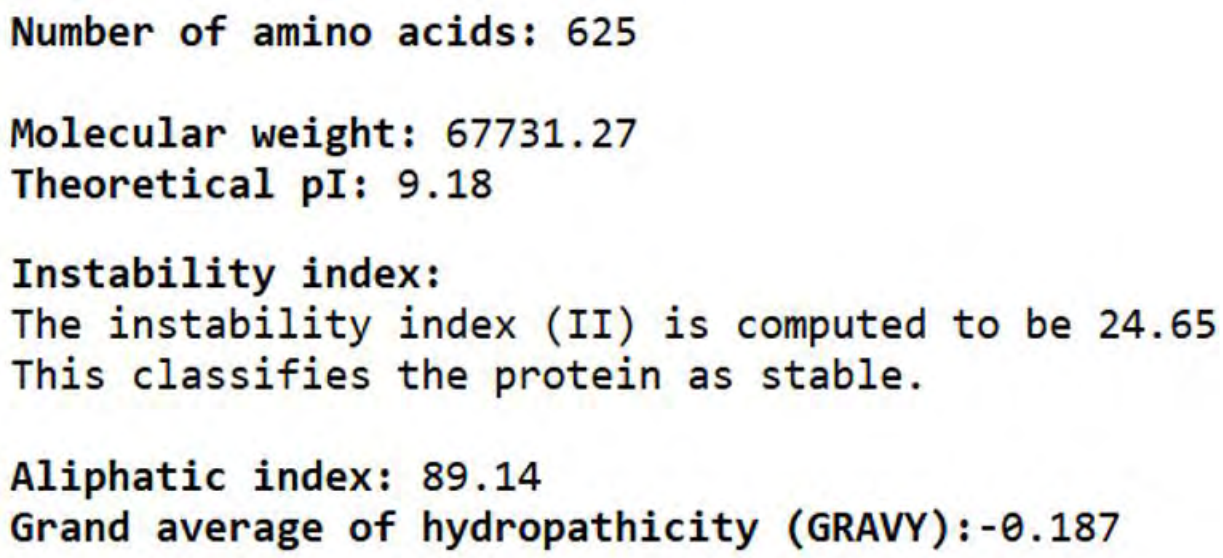
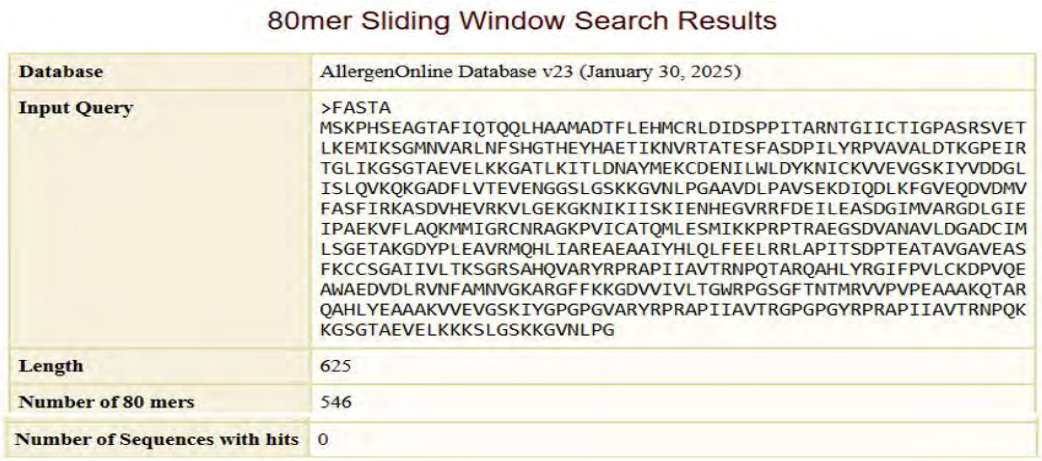


Figure 6 : Number of amino acids, Molecular weight, Theoretical Pi, Instability index and GRAVY of the prepared vaccine

3.7 Allergenicity and Toxicity Prediction of the Constructed Vaccine

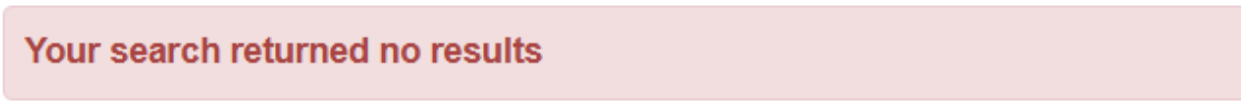
To evaluate the safety of the vaccine, its potential allergenicity was analyzed using the AllergenOnline server by submitting the sequence in plain text format. Furthermore, toxicity assessment was conducted through the T3DB database, allowing for the identification of any toxic properties associated with the vaccine construct (Goodman et al., 2016).



80mer Sliding Window Search Results

Database	AllergenOnline Database v23 (January 30, 2025)
Input Query	>FASTA MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVET LKEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIR TGLIKGSGTAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGL ISLQVKQKGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDMV FASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIE IPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIM LSGETAKGDYPLEAVRMQHLIAREAEAAIYHLQLFEELRRLAPITSDPTEATAVGAVEAS FKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPVQE AWAEDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPVPEAAAKQTAR QAHLYEAAAKVVEVGSKIYGPGPGVARYRPRAPIIAVTRGPGPGYRPRAPIIAVTRNPQK KGSGTAEVELKKKSLGSKKGVNLPG
Length	625
Number of 80 mers	546
Number of Sequences with hits	0

Figure 7: Allergenonline Database for Allergenicity prediction



Your search returned no results

Figure 8 : T3DB Database for Toxicity prediction

3.8 3D Structural Modeling of the Constructed Vaccine

The Phyre2 server was utilized to predict the three-dimensional structure of the vaccine. The sequence was submitted in plain text format, enabling Phyre2 to construct both primary and tertiary structures. To determine the reliability of the generated model, confidence levels and sequence coverage were assessed. The resulting PDB file was sent via email and later analyzed using PyMOL software for structural visualization (Kelley et al., 2015). The confidence was 100% and coverage is 84%. High confidence and good coverage improve vaccine accuracy and effectiveness.

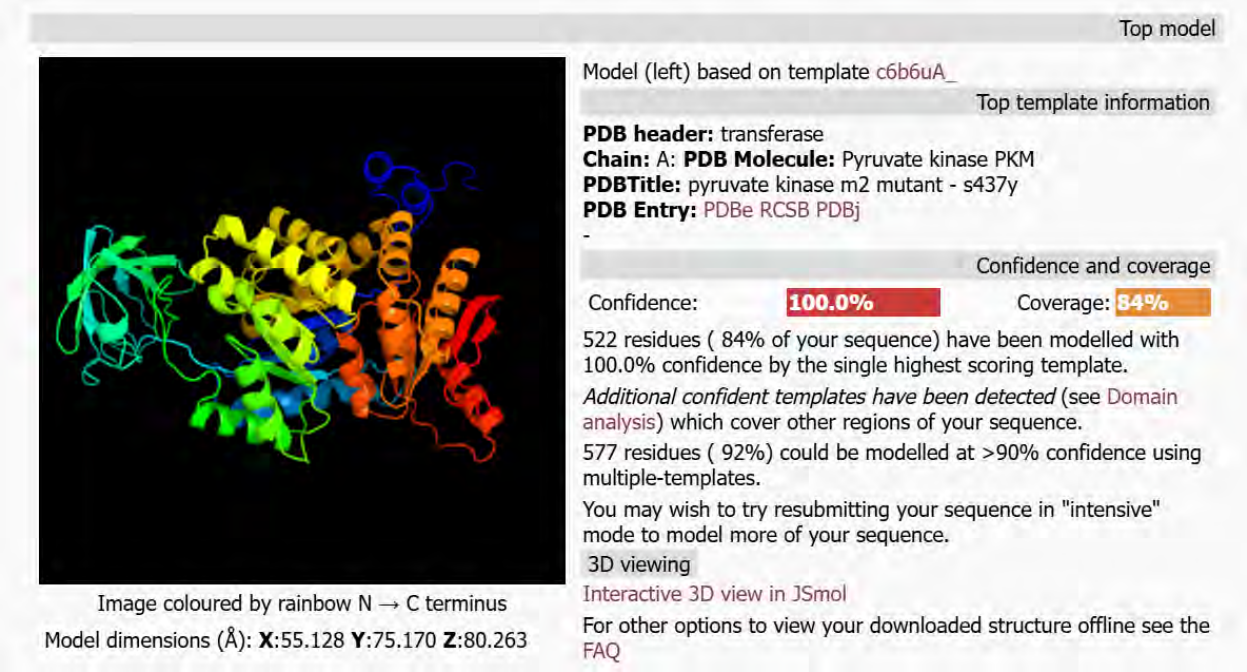


Figure 9 : Score of the homology modeling found from Phyre2

3.9 Analyzing the Ramachandran Plots

The structural integrity of the modeled vaccine was evaluated using a Ramachandran plot generated through the Swiss Model Expasy structure assessment tool. The PDB file obtained from Phyre2 was uploaded, allowing for the visualization of phi and psi angles within the polypeptide chain (Kiefer et al., 2009). This assessment was essential in verifying the correct folding and stability of the protein structure. The Ramachandran Favored region is greater than 92% (98.66%) and Ramachandran outlier is less than 2% (0.38%).

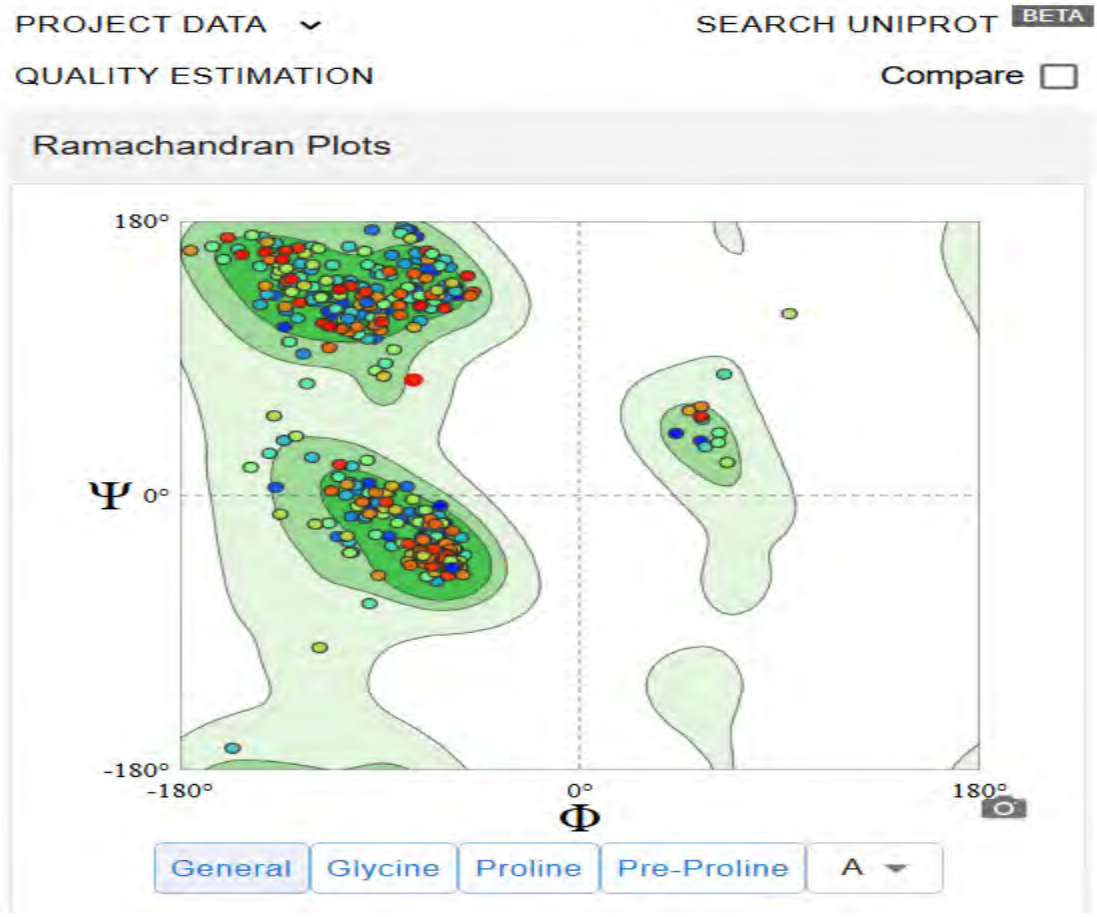


Figure 10: Ramachandran plots from Swiss-Model

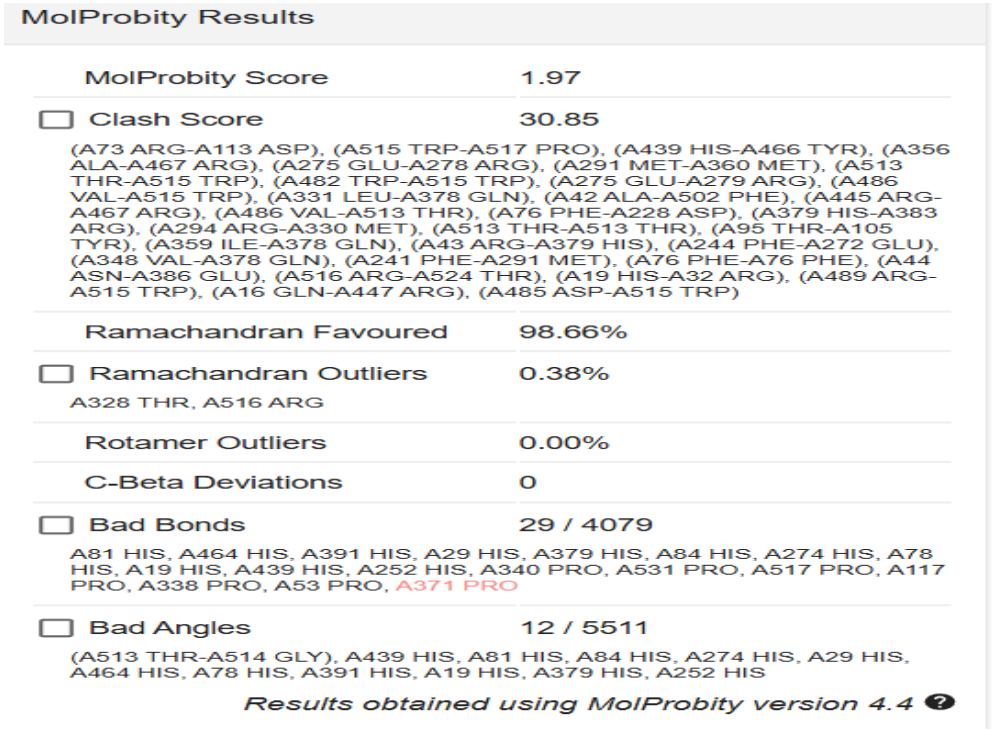


Figure 11 : Molprobity result from Swiss-Model

3.10 Z-Score analysis for 3D- Model

The ProSA-web server was utilized to assess the Z-score of the designed vaccine model. This platform analyzed the overall quality of the 3D structure by evaluating the energy distribution across the model. After uploading the PDB file, the server provided results including the Z-score, a graph comparing the Z-score to residue count, and another displaying energy distribution relative to sequence position (Wiederstein & Sippl, 2007). These evaluations played a key role in verifying the structural reliability of the vaccine. Z-score is -7.57 which was within the range of (-5 to -10).

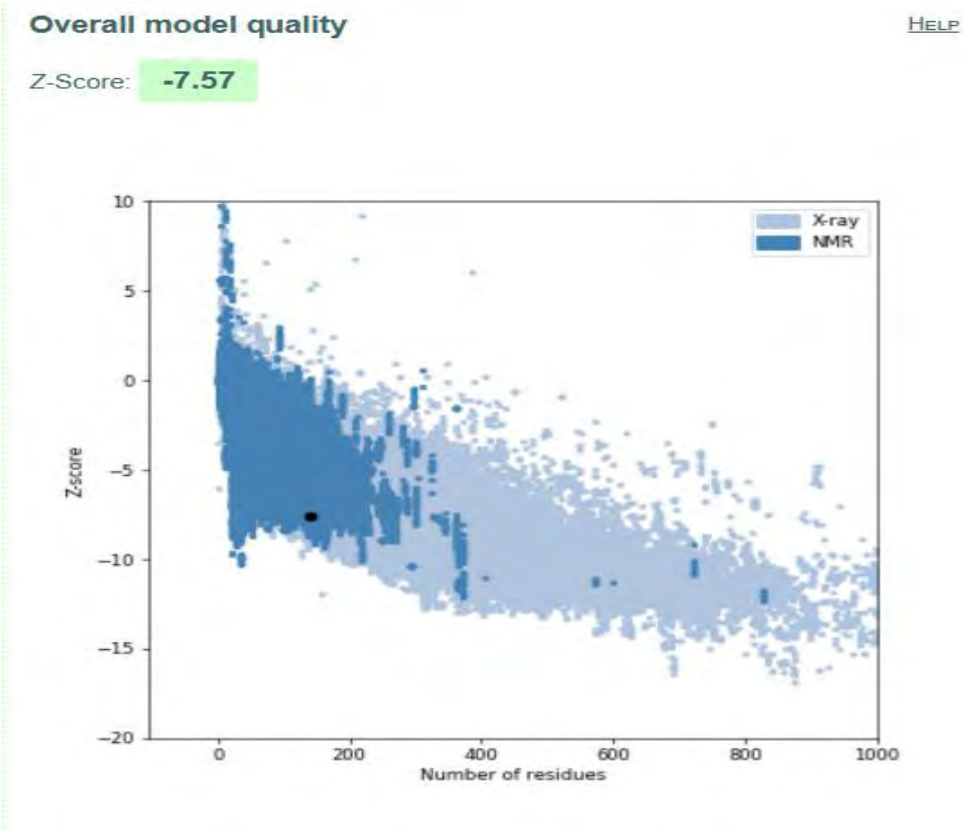


Figure 12: Overall model quality of prepared vaccine

3.11 Molecular Docking

To evaluate the binding interactions between the designed and a human toll-like receptor-3 (TLR-3), molecular docking was performed using the ClusPro server (https://cluspro.bu.edu/). The TLR's PDB file was retrieved from the RCSB Protein Data Bank, while the vaccine's PDB file was uploaded as the ligand. ClusPro generated ten docked complexes, and the complex with the highest binding energy score was chosen for further examination (Kozakov et al., 2017). This resulted in 10 docked complexes out of which the first docked complex was the most stable docked complex. The score of highest docked score was (-964). The best binding Cluster was selected which had both high cluster population and reduced energy.

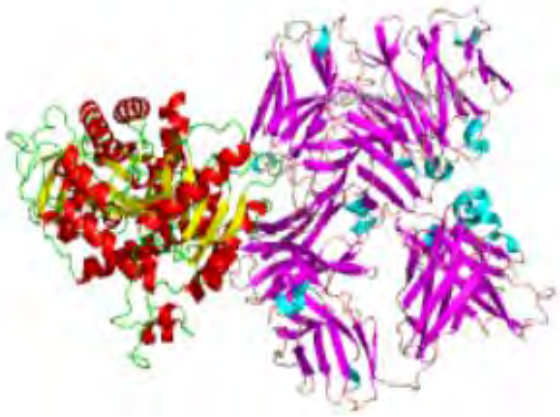


Figure 13 : Molecular docking with suitable TLR-3

Cluster	Members	Representative	Weighted Score
0	64	Center	-851.3
		Lowest Energy	-964.5

Figure 14: Highest stable docked complex from ClusPro

3.12 Immune Simulation

I utilized c-Immsim to analyze the immune response simulation of the vaccine. After entering the sequence, the server successfully completed the process and provided graphical outputs, as shown in Figure 15. The graph illustrates a rise in antigen levels following vaccination. Additionally, it reveals that IgG and IgM levels reached their peak at 28 days, with antibody concentrations increasing significantly by day 60. This suggests that the vaccine effectively stimulated an immune response. Furthermore, lymphocytes were observed to produce both antibodies and memory cells. These memory cells play a crucial role in retaining immune memory, enabling a faster response if the Human Papillomavirus re-enters the body (Rapin et al., 2011).

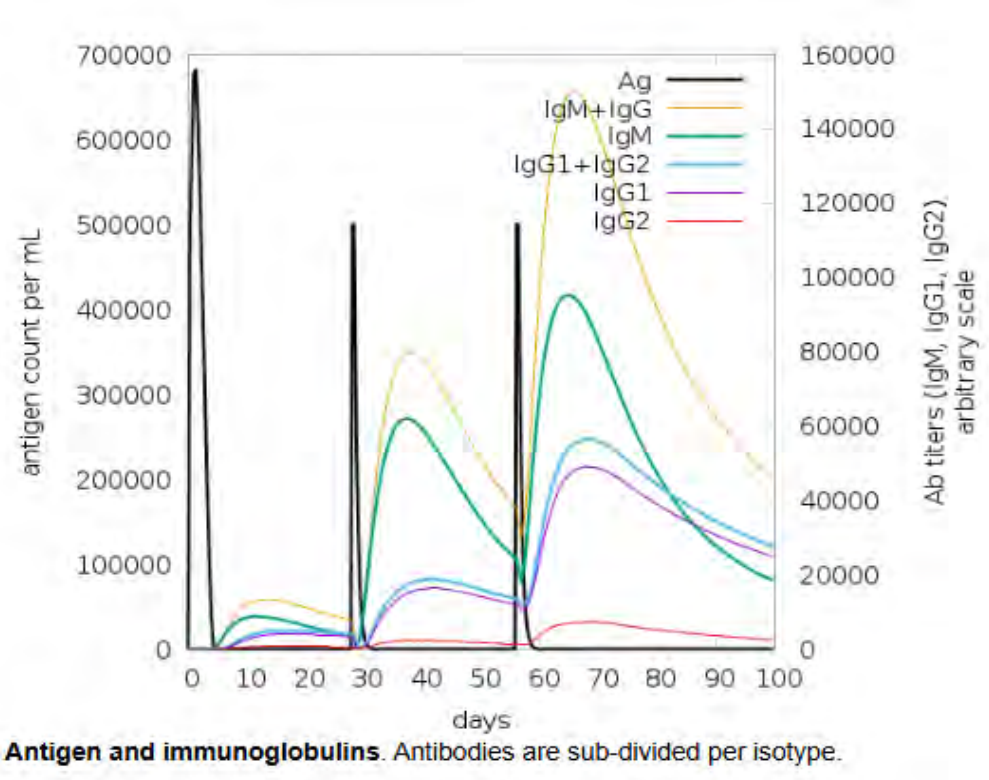


Figure 15: Immune Simulation of vaccine illustrated graphically via C-immsim

Chapter 4

Discussion

Human Papillomavirus particularly strain 16, is a leading cause of oropharyngeal cancer, mainly affecting the tonsils and base of the tongue. It spreads through oral contact, including kissing and oral sex, leading to persistent infections that may develop into cancer. Unlike traditional head and neck cancers linked to smoking and alcohol, HPV-related cases are more common in younger, non-smoking individuals. The virus disrupts key tumor suppressor genes, promoting uncontrolled cell growth. Symptoms include a sore throat, difficulty swallowing, voice changes, and swollen lymph nodes. Diagnosis involves clinical exams, biopsies, HPV testing, and imaging scans. Treatment includes surgery, radiation, chemotherapy, targeted therapy, and immunotherapy, with HPV-positive cases showing better outcomes. Early-stage survival rates exceed 80%. Prevention includes HPV vaccination, safe sexual practices, and routine screenings. Increasing awareness and vaccination efforts are key to reducing cases.

A novel vaccine for Human Papillomavirus (HPV) was designed by selecting a primary protein based on antigenicity using the VaxiJen server. The protein's sequence was analyzed to identify one cytotoxic T lymphocyte (CTL) epitope, one helper T lymphocyte (HTL) epitope, and four B-cell epitopes through various computational tools. These epitopes were carefully screened based on multiple parameters. After evaluating different epitope combinations, six vaccine candidates were developed, and the final formulation included one CTL, one HTL, and one B-cell epitope, as it met all the essential criteria. The antigenicity of this final construct was assessed and found to be superior to that of the initial protein sequence.

To determine the stability of the designed vaccine, the ProtParam tool was utilized, confirming its stability. The amino acid composition was within the desired range, and the Grand Average of Hydropathicity (GRAVY) score was negative, indicating improved solubility and absorption. The 3D structural model was generated using Phyre2, yielding a model with 84% coverage at 100% confidence, which was deemed satisfactory. Further safety evaluations were performed using AllergenOnline and T3DB, confirming that the vaccine was non-allergenic and non-toxic, two fundamental requirements of an ideal vaccine to ensure it does not induce harmful reactions upon administration. The structural quality of the 3D model was validated using ProsaWeb, where the Z-score indicated an acceptable model with proper energy distribution. Additionally, Ramachandran plot analysis was conducted through Swiss-Model Expasy, demonstrating that the structural parameters were within an acceptable range. For molecular docking, human Toll-like receptor 3 (TLR3) was selected as the suitable receptor for HPV. Ten docked complexes were generated, and the most stable one was chosen based on binding energy, with a value of -841.0, indicating a strong and stable interaction. Finally, immune response simulations were conducted using C-IMMSIM, which produced favorable results. Considering all evaluations, this study concludes that the developed vaccine holds potential as a novel preventive measure against Human Papillomavirus.

Chapter 5

Conclusion

In conclusion, Human Papillomavirus (HPV) remains a global concern, but no reported cases have emerged in Bangladesh thus far. However, there is no certainty that the virus will never pose a threat in the country. Despite extensive research efforts, an effective vaccine for HPV is still unavailable. This study utilized computational tools to screen and evaluate various parameters for the development of a novel vaccine candidate. Additionally, immune response stimulation was analyzed, yielding promising results.

The primary objective of this research was to design a multi-epitope vaccine using an in-silico approach. However, a key limitation is that this study focuses solely on the computational modeling of the vaccine without assessing its safety or efficacy. For a vaccine to be considered for real-world application, it must undergo rigorous in-vitro and in-vivo testing to confirm its safety and effectiveness. These experimental validations are essential steps before advancing towards clinical trials and commercialization.

References

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... & Yeh, L. S. L. (2005). The universal protein resource (UniProt). *Nucleic acids research*, *33*(suppl_1), D154-D159.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, *8*, 1-7.

Williams, M. A., & Bevan, M. J. (2007). Effector and memory CTL differentiation. *Annu. Rev. Immunol.*, *25*(1), 171-192.

Abbas, A. K., Murphy, K. M., & Sher, A. (1996). Functional diversity of helper T lymphocytes. *Nature*, *383*(6603), 787-793.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic acids research*, *48*(W1), W449-W454.

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, *8*, 1-12.

Wong, C. K., Ho, C. Y., Ko, F. W. S., Chan, C. H. S., Ho, A. S. S., Hui, D. S. C., & Lam, C. W. K. (2001). Proinflammatory cytokines (IL-17, IL-6, IL-18 and IL-12) and Th cytokines (IFN-γ, IL-4, IL-10 and IL-13) in patients with allergic asthma. *Clinical & Experimental Immunology*, *125*(2), 177-183.

Ponomarenko, J., Papangelopoulos, N., Zajonc, D. M., Peters, B., Sette, A., & Bourne, P. E. (2010). IEDB-3D: structural data within the immune epitope database. *Nucleic acids research*, *39*(suppl_1), D1164-D1170.

Garg, V. K., Avashthi, H., Tiwari, A., Jain, P. A., Ramkete, P. W., Kayastha, A. M., & Singh, V. K. (2016). MFPPI–multi FASTA ProtParam interface. *Bioinformation*, *12*(2), 74.

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., ... & Rappaport, S. M. (2015). T3DB: the toxic exposome database. *Nucleic acids research*, *43*(D1), D928-D934.

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., ... & Taylor, S. L. (2016). AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular nutrition & food research*, *60*(5), 1183-1198.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic acids research*, *37*(suppl_1), D387-D392.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(suppl_2), W407-W410.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., ... & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature protocols*, *12*(2), 255-278.

Rapin, N., Lund, O., & Castiglione, F. (2011). Immune system simulation online. *Bioinformatics*, *27*(14), 2013-2014.

Burd, E. M., & Dean, C. L. (2016). Human papillomavirus. *Diagnostic Microbiology of the Immunocompromised Host*, 177-195.

Brentjens, M. H., Yeung-Yue, K. A., Lee, P. C., & Tyring, S. K. (2002). Human papillomavirus: a review. *Dermatologic clinics*, *20*(2), 315-331.

Munger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., ... & Huh, K. (2004). Mechanisms of human papillomavirus-induced oncogenesis. *Journal of virology*, *78*(21), 11451-11460.

Moody, C. A., & Laimins, L. A. (2010). Human papillomavirus oncoproteins: pathways to transformation. *Nature reviews cancer*, *10*(8), 550-560.

Taberna, M., Mena, M., Pavón, M. A., Alemany, L., Gillison, M. L., & Mesía, R. (2017). Human papillomavirus-related oropharyngeal cancer. *Annals of Oncology*, *28*(10), 2386-2398.

Stein, A. P., Saha, S., Kraninger, J. L., Swick, A. D., Yu, M., Lambert, P. F., & Kimple, R. J. (2015). Prevalence of human papillomavirus in oropharyngeal cancer: a systematic review. *The Cancer Journal*, *21*(3), 138-146.

Tanaka, T. I., & Alawi, F. (2018). Human papillomavirus and oropharyngeal cancer. *Dental Clinics*, *62*(1), 111-120.

Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tân, P. F., ... & Gillison, M. L. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *New England Journal of Medicine*, *363*(1), 24-35.

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., ... & Gillison, M. L. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology*, *29*(32), 4294-4301.