In Silico Design of a Multi Epitope Vaccine for Epithelial Cancer Immunotherapy-a comparative study

By

ZUBAIR BIN RAFIQUE

21146019

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons.)

School of Pharmacy

BRAC University

May 2026

© 2026. BRAC University

All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing my 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:

_____

ZUBAIR BIN RAFIQUE

21146019

Approval

The thesis titled “In Silico Design of a Multi Epitope Vaccine for Epithelial Cancer Immunotherapy-a comparative study” submitted by Zubair Bin Rafique (ID: 21146019) of Summer, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on May, 2026.

Supervised By:

Supervisor:

_____

Mohammad Kawsar Sharif Siam

Senior Lecturer, School of Pharmacy

BRAC University

Approved By:

Chairperson:

_____

Professor Sharmind Neelotpol, PhD

Chairperson, School of Pharmacy

BRAC University

Ethics Statement

This thesis was completed in full compliance with required ethical standards, with no involvement of any human participants or animal trials. Research integrity and ethical principles were strictly maintained throughout the study.

Abstract

Epithelial cancers remain a leading cause of morbidity and mortality worldwide, and many patients either fail or cannot tolerate conventional therapies. Cancer immunotherapy, particularly multi-epitope peptide and mRNA vaccines, has emerged as a promising strategy to induce precise, durable anti-tumor immune responses while limiting systemic toxicity. Leveraging tumor-associated and cancer-testis antigens as well as oncogenic drivers enables rational vaccine design tailored to epithelial malignancies such as breast, lung, colorectal and cervical cancers. This study presents a comparative analysis of in silico based multi epitope vaccine construction against epithelial cancer targeting 4 different proteins. Each protein was evaluated for vaccine antigenicity, physicochemical properties and structural stability using immune informative approach to identify the most suitable vaccine candidate. Various tools were used to identify the different epitopes of HTL, CTL, Bcell, which were connected using suitable linkers. Physicochemical properties were also check for the prepared vaccine through using Protparam. Molecular Docking was done with TLR-3 for checking the residual interaction. Among the proteins investigated, fascin1 emerged as the most promising candidate demonstrating superior vaccine antigenicity, favorable physical properties and structural stability. Collectively, this approach is expected to yield a structurally stable, antigenic, and globally applicable multi-epitope vaccine candidate capable of eliciting coordinated humoral and cellular immunity against epithelial cancers, warranting subsequent in vitro and in vivo validation.

Keywords

Immunoinformatics, Multi Epitope Vaccine, CTL Epitope, HTL Epitope, 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 **Pages**

Declaration ..... **ii**

Approval ..... **iii**

Ethics Statement ..... **iv**

Abstract ..... **v**

Acknowledgement ..... **vi**

List of Tables ..... **ix**

List of Figures ..... **x**

List of Acronyms ..... **x**

Chapter 1 Introduction **1**

1.1The Global Burden of Epithelial Cancers……..1

1.2 Epidemiology of Epithelial Cancers in Bangladesh……..1

1.3 Age Distribution and High-Risk Groups in Bangladesh……2

1.4 Limitations of Conventional Cancer Treatments……...2

1.5 Emergence of Immunotherapy and Cancer Vaccines……3

1.6 Concept and Advantages of Multi-Epitope Vaccines……3

1.7 Rationale for a Comparative In Silico Study in Epithelial Cancer……4

Chapter 2 Methodology **5**

2.1 Protein Retrieval and Antigen Selection.....6

2.2 CTL Epitope Prediction and Filtering.....6

2.3 HTL Epitope Prediction and Analysis..... 6

2.4 B-cell Epitope Prediction and Characterization ……. ……..... 7

2.5 Multi Epitope Vaccine Design……..... 7

2.6 Physicochemical and Immunological Profiling…….....7

2.7 3D Structure Modelling and Validation…….....8

2.8 Molecular Docking with Immune Receptors……8

2.9 Immune Response Simulation..... 8

Chapter 3 Results 9

3.1 Antigenicity and protein selection……9

3.2 Selection of CTL epitopes……...12

3.3 Selection of HTL epitopes……..16

3.4 Selection of B-cell epitopes…….18

3.5 Construction of the final vaccine…….21

3.6 Biochemical analysis of the constructed vaccine……23

3.7 3D model of the constructed vaccine……..25

3.8 Analysis of Z-score…….26

3.9 Ramachandran Plotting……..28

3.10 Molecular Docking of the Constructed Vaccine……..30

3.11 Immune Simulation of the Constructed Vaccine…….32

Chapter 4 Discussion 36

Chapter 5 Conclusion 39

References 40

List of Tables

Table 1: Protein Based on Antigenicity scores……..11

Table 2: Summary of predicted CTL epitopes showing sequence identifiers, peptide length, eluted ligand score (EL), percentile ranks for eluted ligand and binding affinity, and MHC class I allele binding affinity as determined using the NetMHCpan 4.1 server……..14

Table 3: Antigenicity, Allergenicity, Toxicity determination of the Selected CTL epitopes….15

Table 4. Helper T-lymphocyte (HTL) epitopes identified from the selected protein candidates and their predicted cytokine-inducing potential……17

Table 5: Antigenicity, Allergenicity, Toxicity determination of the Selected HTL epitopes…17

Table 6: Antigenicity, Allergenicity, toxicity profile assessment of selected B-Cell epitopes ..21

Table 7: Assessment of the constructed vaccines…….23

Table 8: Physico-chemical parameters of the constructed vaccines using Protparam Tool….24

Table 9: Coverage and Confidence determination of the Constructed Vaccines using PHYRE-2……25

Table 10: Z-score evaluation of the predicted vaccine constructs obtained using the ProSA-web structural validation tool……...27

Table 11: Ramachandran Plot Prediction for all the vaccines…….27

Table 12: Summary table of the all the parameters assessed for the vaccine candidates ...….29

List of Figures

Figure 1: Step-by-Step Schematic overview of the entire process involved in constructing an In-Silico Vaccine…….....5

Figure 2: Determining antigenicity of the Protein no: 1 9

Figure 3: Determining antigenicity of the Protein no: 2 10

Figure 4: Determining antigenicity of the Protein no: 3 10

Figure 5: Determining antigenicity of the Protein no: 4 11

Figure 6: CTL epitopes generated by NetCTL1.2 …….13

Figure 7 (a-d): Position vs. score graph and B-Cell epitopes for all proteins 18

Figure 8: Protparam Results for Vaccine ……..25

Figure 9: Phyre 2.2 generated 3D model with 82% residue coverage at 100% confidence…. .26

Figure 10: Z-score graph and local model quality of vaccine i by ProSa Web…….27

Figure 11: Ramachandran plot generation along with scores and model by Swiss Model Expasy……..28

Figure 12: Model energy scores achieved by CLUSPRO…….31

Figure 13: C-ImmSim generated Immune Simulation ……33

Figure 14: Cytokine and Interleukin Concentration graph containing IFN gamma, Interleukin 4 and 10 ……34

Figure 15: B cell Lymphocyte population graph ……..34

Figure 16: CD4 T helper Lymphocyte population graph ……..35

Figure 17: CD8 cytotoxic T Lymphocyte population graph …….35

List of Acronyms

CTL: Cytotoxic T- Lymphocyte

GRAVY: The Grand Average of Hydropathicity

HTL: Helper T-Lymphocyte

MHC: Major Histocompatibility complex

Chapter 1

Introduction

1.1 The Global Burden of Epithelial Cancers

Epithelial malignancies, including cancers of the breast, lung, colorectum, stomach, cervix, skin and oral cavity, account for the majority of incident cancers and cancer deaths worldwide. In 2020, an estimated 19.3 million new cancer cases and 10.0 million deaths were reported globally, with female breast, lung, colorectal, prostate and stomach cancers dominating incidence and mortality profiles (Sung et al., 2021).Despite improvements in early detection and treatment in high-income settings, many epithelial cancers remain associated with substantial disability and premature mortality, particularly in transitioning countries (Sharma et al., 2024).Demographic ageing, population growth and increasing exposure to modifiable risk factors (tobacco, obesity, ambient air pollution, infections) are projected to drive a 47% increase in global cancer incidence by 2040 (Huang et al., 2022).These trends underscore the need for innovative preventive and therapeutic strategies targeting epithelial tumors.

1.2 Epidemiology of Epithelial Cancers in Bangladesh

Bangladesh is a developing Asian country characterized by a relatively lower rate of incident cancers; however, it experiences high mortality and burden of cancer deaths (Sharma et al., 2024). The standardized cancer incidence rate in Bangladesh in 2019 was 99.2 per 100,000 people, making it one of the lowest recorded in Asia. However, the overall burden in terms of deaths and disability-adjusted life years is still on the rise ("Burden and Trends of Cancer in Bangladesh," 2023). According to hospital records and regional incidences, epithelial cancers are widespread across Bangladesh (Kundu et al., 2025). Some of the most common types include breast cancer, colorectal cancer, stomach cancer, cervical cancer, and lung cancer. Breast cancer has been identified as the most common type among Bangladeshi women, where the standardized incidence rate is 22.5 per 100,000 with highest cases of 15-44 year old women (Nessa et al., 2018). Oral cancer, an epithelial cancer involving the lips and mouth, is among the cancers with the highest incidence and mortality rates in South Asia. The standardized rate for oral cancer in Bangladeshi men in 2022 was 13.61 per 100,000 with high mortality rates of 8.07 per 100,000 people, and a consistent upward trend (Warnakulasuriya & Filho, 2025). From national analyses, cancer is currently the fourth-leading cause of death in Bangladesh ("Burden and Trends of Cancer in Bangladesh," 2023).

1.3 Age Distribution and High-Risk Groups in Bangladesh

Age distribution of epithelial cancers in Bangladesh highlights an increased burden on middle-aged and older people; however, there is also an increasing trend towards younger patients in certain types. According to regional statistics from two tertiary institutions in northern Bangladesh between 2019-2021, the 41-60 years age range has the highest prevalence of cancer, followed by the age group between 61-80 years old (Kundu et al., 2025).Breast cancer, the commonest malignancy among this age range, especially afflicts middle-aged and early-adult women (Kundu et al., 2025).According to national and institutional reports, in Bangladeshi women aged 15-44 years, breast cancer is the most common type (19.3 per 100,000), and also, the mean age of breast cancer patients has been decreasing in recent decades (Nessa et al., 2018).Cervical cancer also has a high incidence among young patients, with more than 8,000 new diagnoses and almost 5,000 deaths being noted in 2020, while early childbirth was one of the main factors that lead to this problem ("Burden and Trends of Cancer in Bangladesh," 2023).Lung and liver cancers, both mainly epithelial in nature, have increased prevalence, especially in older men with a high tobacco consumption history ("Burden and Trends of Cancer in Bangladesh," 2023).Consequently, it could be stated that there is an increased burden of epithelial cancers on economically productive age groups in Bangladesh (Sharma et al., 2024).

1.4 Limitations of Conventional Cancer Treatments

Surgical treatment, chemotherapy, and radiotherapy form the primary mode of cancer management, and despite their effectiveness for treating epithelial cancers, they possess notable limitations, including lack of efficacy, especially in poor and developing nations. These interventions may cure certain types of early cancers; however, most of them do not treat advanced or recurrent disease effectively. Moreover, they pose notable risks such as side effects, associated illnesses, and reduced quality of life (Igarashi & Sasada, 2020). Additionally, cancer therapy drugs such as radiotherapy and chemotherapy are not specific to cancerous cells, posing threats to normal tissues and causing adverse reactions, thereby reducing tolerability (Kumar et al., 2021). Furthermore, resistance to cancer therapy is a significant problem since small quantities of cancer cells may remain in the body even after complete surgery, and individual differences between tumors are possible causes of recurrence (Lahiri et al., 2023). Many Bangladeshi patients delay their cancer diagnosis, while lack of access to surgical

intervention, radiotherapy services, and systemic treatment options limit the effectiveness of conventional methods (Sharma et al., 2024).

1.5 Emergence of Immunotherapy and Cancer Vaccines

The current trend in cancer immunotherapy is that cancer immunotherapy has become a crucial modality in the treatment of cancer along with surgery, radiotherapy, chemotherapy, and targeted therapy (Esfahani et al., 2020). Immune checkpoint inhibitors, adoptive cell therapies, monoclonal antibodies, and cytokines are all types of cancer immunotherapies that leverage the host's immune system against cancer cells and have been effective against various epithelial malignancies, including lung, colorectal, gastric, and breast cancers (Yang et al., 2023). Nevertheless, immune checkpoint inhibitors are beneficial in certain patients, and most tumors lack high levels of antigens, T cells, and possess immunosuppressive properties (Igarashi & Sasada, 2020). In contrast, therapeutic cancer vaccines are an alternative that triggers the development or enhancement of anti-tumor immune response. By exposing the immune system to either tumor-associated or novel tumor antigens, therapeutic cancer vaccines target to develop lasting T and B cell-mediated immunity without damaging healthy tissue (Urban et al., 2021). Preclinical trials on stomach and lung cancers reveal that peptide, vector, and cellular vaccines increase patient survival rates and are safe, particularly in combination with chemotherapy and immune checkpoint inhibitors (Lahiri et al., 2023).

1.6 Concept and Advantages of Multi-Epitope Vaccines

Multi-epitope vaccines (MEVs) represent a novel type of cancer vaccines that consist of multiple components capable of activating cytotoxic T cells, helper T cells, and B cells derived from multiple proteins associated with cancer progression. MEVs usually include immune modulating adjuvants and smart linkers (Mahdevar et al., 2022). Multi-epitopes allow for broadening the scope of activation of immune system cells, decreasing the risk of evasion from the immune reaction, and inducing cellular and humoral responses (Zhou et al., 2025).Computer-designed MEVs targeting various forms of breast cancer, colorectal cancer, and other solid tumors possess high antigenic properties, do not produce allergic reactions, target a broad spectrum of HLA alleles (exceeding 90 percent of certain types of HLA), and have a stable bond with toll-like receptors in simulations (Wang et al., 2025). In addition,

immune simulations indicate significant production of cytokines and development of immunological memory with persistent antitumor activity. Such a combination suggests that MEVs may be superior to single-epitope vaccines in terms of broad and long-lasting immunity (Yaseen et al., 2025).This property is particularly relevant for heterogeneous epithelial carcinomas, characterized by multiple oncogenic pathways and tumor cell populations.

1.7 Rationale for a Comparative In Silico Study in Epithelial Cancer

Although there have been many developments in the field of MEVs, relatively little research has been conducted to analyze the various antigens, epitope combinations, and vaccines. The majority of the in silico and preclinical studies tend to focus on a single tumor antigen or construct without comparing alternative designs and approaches (Mahdevar et al., 2022). There has been an urgent need for identifying the best combination of immunogens that could improve the immunogenicity, HLA coverage, and safety of the vaccine and also be compatible with local population genetics and disease burden in Bangladesh for treating common epithelial cancers including breast, lung, colorectal, gastric, and oral cancers (Sung et al., 2021). A comparative immunoinformatics analysis of multiple constructs of MEVs based on the selected epitope targets for epithelial cancers can demonstrate the effects of design parameters such as origin, adjuvant, linkers, and administration mode on the predicted stability, binding affinity to cell surface receptors, and the immune response generated (Wang et al., 2025). In other words, this strategy can aid in selecting potential constructs that should be further analyzed through laboratory testing and clinical applications for low and middle-income populations. Hence, this study employs in silico techniques to conduct a comparative analysis of MEV constructs for epithelial cancer immunotherapy.

Chapter 2

Methodology

The major steps are as followed below:

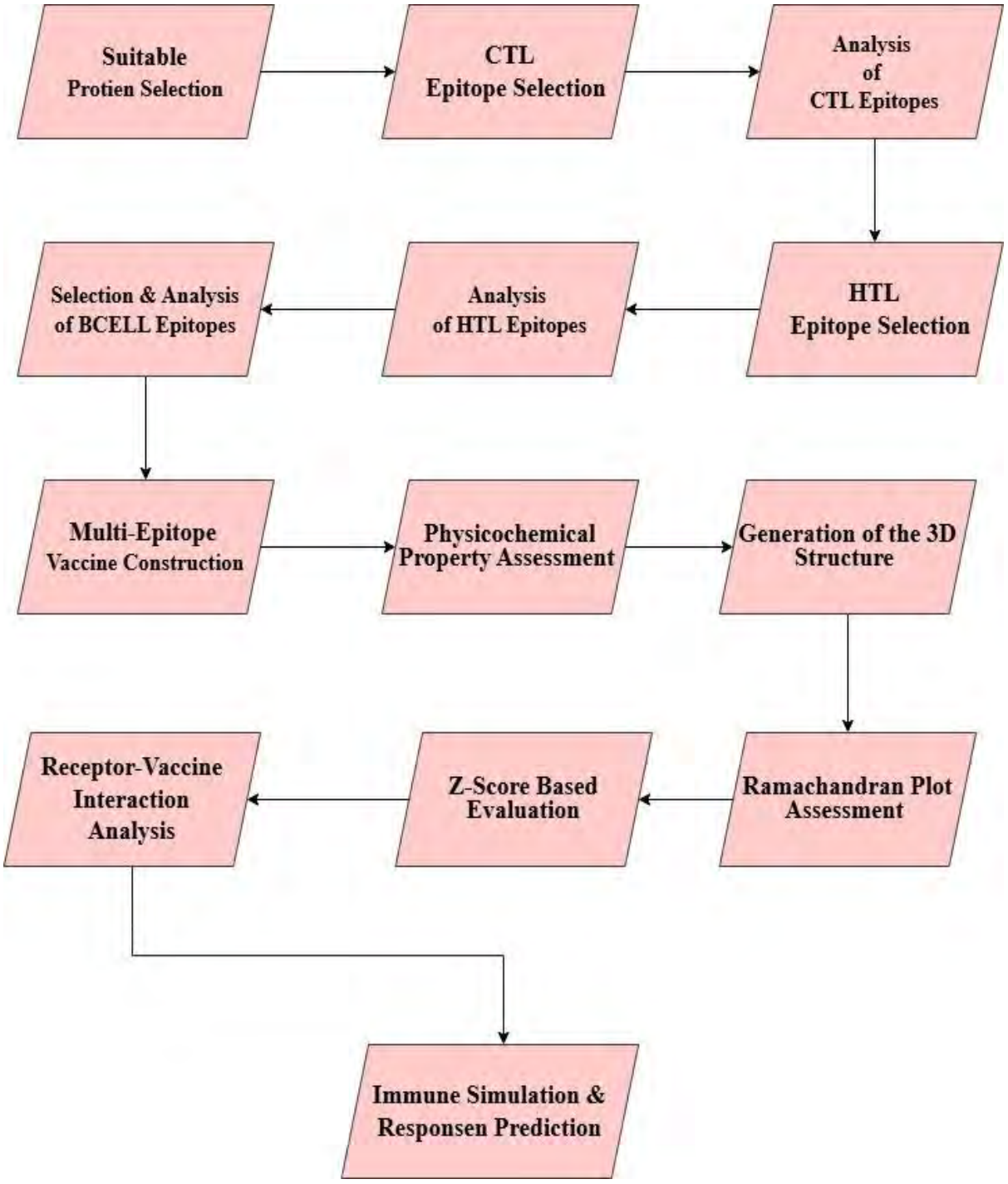


Figure 1: Step-by-Step Schematic overview of the entire process involved in constructing an In-Silico Vaccine

2.1 Protein Retrieval and Antigen Selection

The sequences of the protein for the epithelial cancer-associated antigens were retrieved from UniProt Knowledgebase in FASTA format. We prioritized experimentally validated and reviewed Swiss-Prot sequences (Long et al., 2024). The criteria for choosing the potential antigens include overexpression in epithelial cancers, being cell surface or secreted antigens, having roles in tumor progression, and being dissimilar to essential human housekeeping genes to prevent any possible side effects (Shehri, 2025). Their antigenicity was verified using VaxiJen software; those with high tumor-associated antigenicity prediction values were selected for epitope mapping analysis (Sethi et al., 2025).

2.2 CTL Epitope Prediction and Filtering

Peptide sequences (9-11 amino acids) of cytotoxic T lymphocytes (CTL) epitopes were predicted from the selected antigens via NetCTL and IEDB MHC class I binder prediction tools. Both of these tools take into account the factors such as peptide-MHC interaction, cleavage by proteasomes, and translocation of the resulting peptide by TAP to predict the probability of presentation of the epitopes (Dar et al., 2022). Selected epitopes had high binding affinity to common HLA class I alleles and had higher scoring overall. The selected epitopes underwent antigenicity evaluation via VaxiJen/ANTIGENpro tools, followed by allergenicity and toxicity tests using AllerTOP and ToxinPred tools; peptides selected for further analysis were antigenic but not allergenic or toxic (Elshafei et al., 2024). . Immunogenicity scores were obtained from IEDB class I immunogenicity tools to prioritize CTL epitopes capable of robust T-cell activation (Motamedi et al., 2023) .

2.3 HTL Epitope Prediction and Analysis

Predictions for Helper T Lymphocyte (HTL) epitopes, which ranged between 15-20 amino acid residues, were generated using MHC class II binding tools from the IEDB website. These predictions were validated against a database of common HLA DR, DP, and DQ alleles, and those with high and moderate binding potentials were retained on the basis of percentile rank and IC50 predictions (Long et al., 2024). The HTL epitopes were subjected to antigenicity, allergenicity, and toxicity assessments similar to those done for CTL epitopes, and epitopes with favorable profiles were retained (Sethi et al., 2025). For analysis of their potential immunomodulatory effects, predictions of their capacity to induce the production of IFN-γ, IL-2, and IL-4 were performed by cytokine-specific servers such as IFNepitope, giving priority to epitopes that were associated with Th1-biased responses (Firuzpour et al., 2025).

2.4 B-cell Epitope Prediction and Characterization

Linear B cell epitopes were selected for target antigens by employing ABCpred and BepiPred algorithms. Such computational tools apply machine learning and physicochemical attributes to identify exposed and flexible sites that can be recognized by antibodies (Ayyagari et al., 2022). The chosen epitopes underwent further evaluation concerning their antigenicity, non-allergenic nature, non-toxicity, and significant dissimilarity to human proteins (Motamedi et al., 2023). To construct the vaccine formulation, linear B cell epitopes overlapping or located in the vicinity of T cell epitopes were used to ensure cooperative immune responses (Long et al., 2024).

2.5 Multi-Epitope Vaccine Design

The CTL, HTL, and B-cell epitopes that have been shortlisted were merged to form the multi-epitope vaccine (MEV) designs, with the inclusion of immune-enhancing linkers and adjuvants. Various designs were formulated for analysis (Zhou et al., 2025). The EAAAK linker is used to isolate the CTL epitopes to facilitate proteosome processing. The GPGPG linker is added between the HTL epitopes to preserve their conformation and promote MHC II presentation. The KK linker is added to B-cell epitopes to promote their flexibility and solvent exposure (Motamedi et al., 2023). Universal helper epitopes (e.g., PADRE) and cell-penetrating/TAT motifs are included in some designs to increase the HLA coverage and improve internalization (Firuzpour et al., 2025).

2.6 Physicochemical and Immunological Profiling

Physicochemical parameters for each MEV were estimated using ProtParam, which included the amino acid composition, molecular weight, isoelectric point (pI), instability index, aliphatic index, and GRAVY score to assess stability, thermostability, and hydrophilicity (Sethi et al., 2025). The antigenicity, allergenicity, and toxicity of these constructs were re-evaluated using VaxiJen and AllerTOP and ToxinPred, respectively, to ensure that the fusion of epitopes did not introduce any undesirable characteristics (Long et al., 2024). Constructs meeting the defined threshold criteria proceeded to the structural analysis stage.

2.7 3D Structure Modelling and Validation

The tertiary structures of MEVs were predicted either using homology-based approaches or de novo modeling programs such as Robetta, Phyre2, SWISS MODEL, or AlphaFold, based on whether there were suitable template structures and the length of the sequence (Zhou et al., 2025). These predicted structures were further optimized using GalaxyRefine or similar servers, where they were optimized for proper side chain prediction and stereochemical properties (Fatahi et al., 2025). The quality of these models was assessed using Ramachandran plot analysis using PROCHECK or MolProbity for evaluating the proportion of amino acid residues that occur in favorable, allowable, and disallowed regions, and Z score assessment using ProSA web for comparing energy distributions of the modeled structure against other experimentally determined proteins with similar lengths (Elshafei et al., 2024).

2.8 Molecular Docking with Immune Receptors

To validate how efficiently the TLR ligands (TLR2, TLR3, TLR4, or TLR8 based on the type of adjuvant) can engage with the MEV ligands, protein-protein docking was done using various software such as ClusPro, PatchDock/FireDock, HDOCK, and other similar programs (Salahlou et al., 2024). Prior to the docking, the receptor and ligand molecules were refined through the removal of waters and other unnecessary components and addition of missing hydrogens. Default or balanced scoring schemes were considered for docking simulations, and the complexes were selected on the basis of clusters' number, binding energies, and surface complementarities (Long et al., 2024). Additionally, important intermolecular interactions, including hydrogen bonding and salt bridges, were observed to evaluate the strength and authenticity of the TLR engagements (Fatahi et al., 2025).

2.9 Immune Response Simulation

Computational simulation of the immune reaction was conducted using the C ImmSim server to identify the manner by which each of the MEVs would elicit a humoral and cellular response during the prime boost regime (Salahlou et al., 2024). Each construct's amino acid sequence was used as an input, with default values for random seed, simulation volume, and cell number, where applicable. Immunization was conducted at specific times as a means of mimicking a three-shot vaccination process. From the outputs generated, there was antigen clearance, the class of antibodies that formed, variations in the populations of CTL and HTL, and generation of memory cells as well as production of certain cytokines like IFN-γ and IL-2 (Motamedi et al., 2023).

Chapter 3

Results

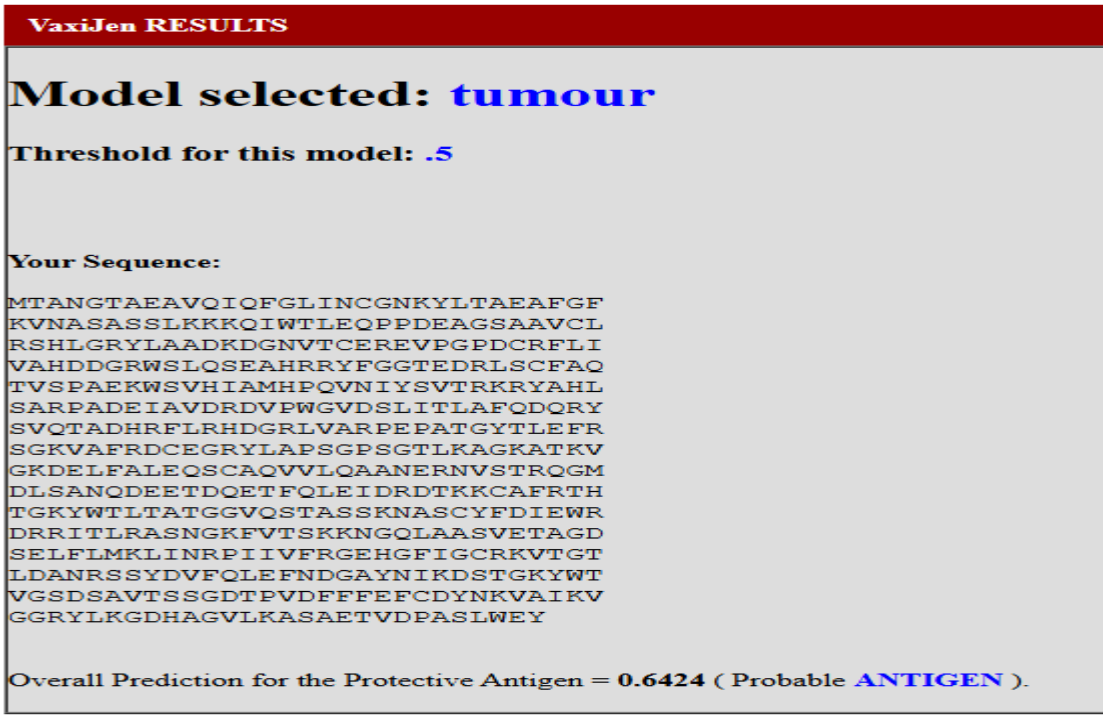
3.1 Antigenicity and protein selection

4 proteins were selected and downloaded from the Uniprot database. Afterwards, their antigenic potential was checked by Vaxijen. Antigenicity is an important aspect because it determines whether the protein has the capacity to show immune response. Vaxijen 2.0 is an essential web based tool which anticipates whether protein sequences are antigenic or not based on the threshold and organism that has been set. For our initial antigenicity screening our chosen protein passed as it passed as antigenic when the 0.5 threshold was applied and the target organism we selected for this is "Tumour" as we are working with Cancer. The protein sequences that are retrieved from Uniprot are given below:

Protein 1

Protein Name: **Fascin**

Antigenicity: 0.6424


VaxiJen RESULTS

Model selected: tumour

Threshold for this model: .5

Your Sequence:

MTANGTAEAVQIQFGLINCGNKYLTAEAFGF
KVNASASSLKKKQIWTLEQPPDEAGSAAVCL
RSHLGRYLAADKDGNVTCEREVPGPDCRFLI
VAHDDGRWSLQSEAHRRYFGGTEDRLSCFAQ
TVSPAEKWSVHIAMHPQVNIYSVTRKRYAHL
SARPADEIAVDRDVPWGVDSLITLAFQDQRY
SVQTADHRFLRHDGRLVARPEPATGYTLEFR
SGKVAFRDCEGRYLAPSGPSGTLKAGKATKV
GKDELFALEQSCAQVVLQAANERNVSTRQGM
DLSANQDEETDQETFQLEIDRDTKKCAFRTH
TGKYWTLTATGGVQSTASSKNASCYFDIEWR
DRRITLRASNGKFVTSKKNGQLAASVETAGD
SELFLMKLINRPIIVFRGEHGFIGCRKVTGT
LDANRSSYDVFQLEFNDGAYNIKDSTGKYWT
VGSDSAVTSSGDTPVDFFFEFCDYNKVAIKV
GGRYLKGDHAGVLKASAETVDPASLWEY

Overall Prediction for the Protective Antigen = **0.6424** (Probable ANTIGEN).

Figure 2: Determining antigenicity of the Protein no: 1 (Doytchinova & Flower, 2007)

Protein 2

Protein Name: **Tubulin gamma-1 chain**

Antigenicity: 0.5473

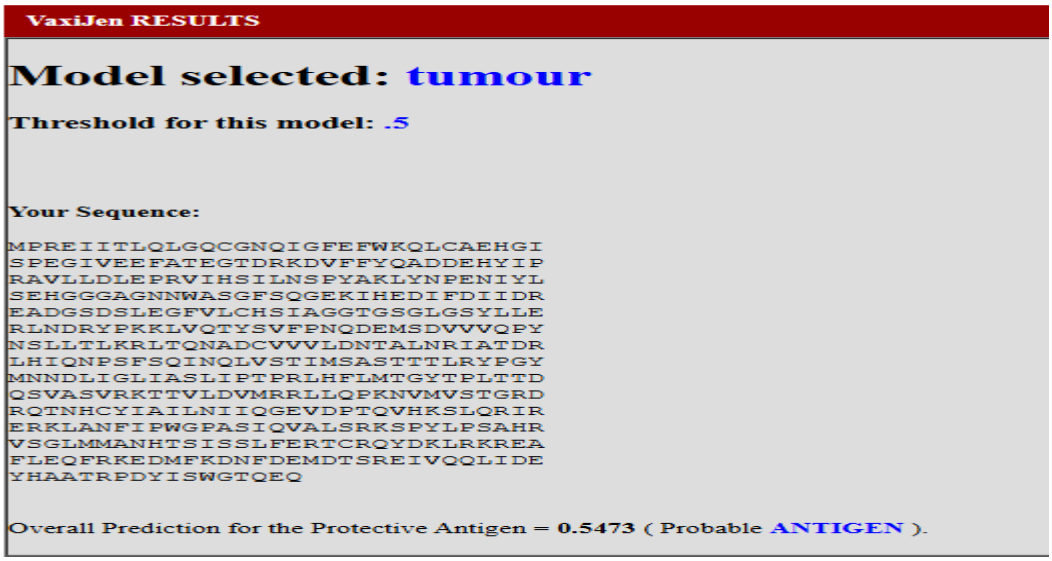


Figure 3: Determining antigenicity of the Protein no: 2 (Doytchinova & Flower, 2007)

Protein 3

Protein Name: **Interstitial collagenase**

Antigenicity: 0.5232

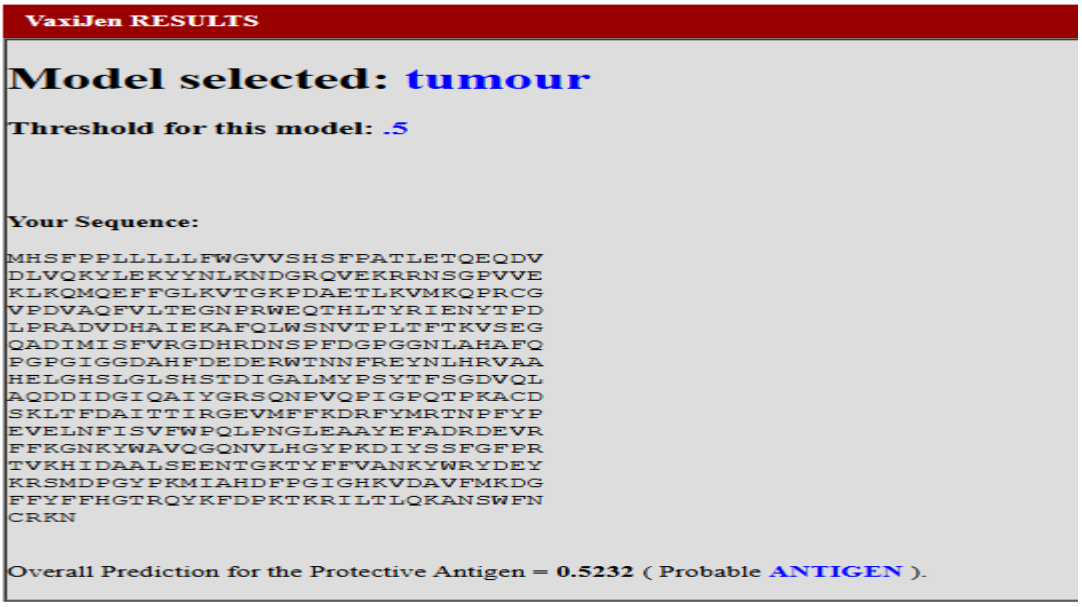


Figure 4: Determining antigenicity of the Protein no: 3 (Doytchinova & Flower, 2007)

Protein 4

Protein Name: **Dynein regulatory complex subunit 4**

Antigenicity: 0.6220

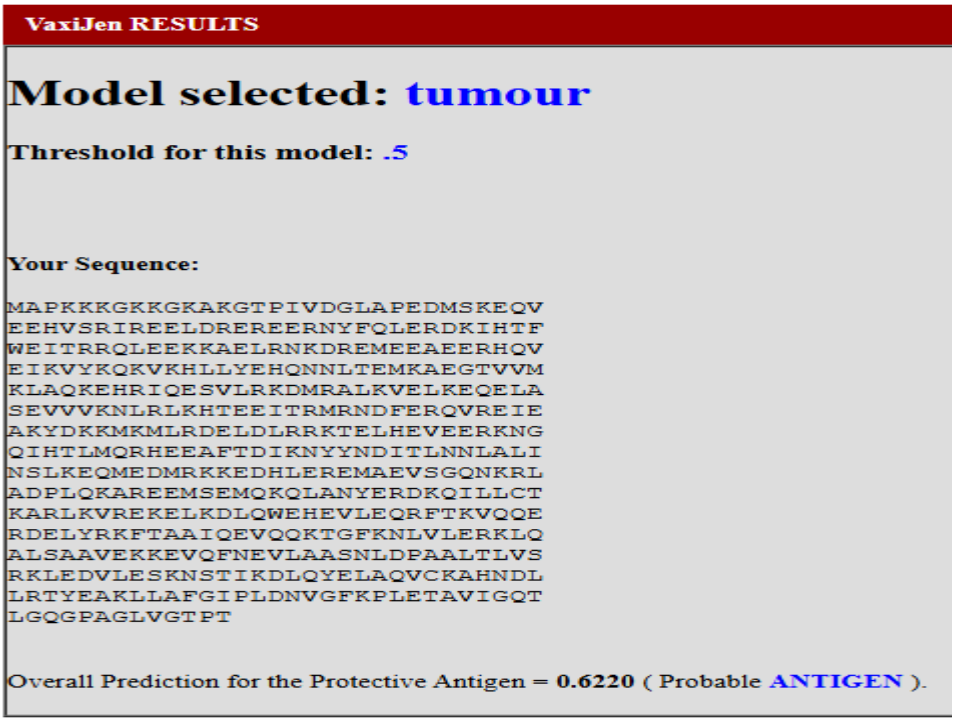


Figure 5: Determining antigenicity of the Protein no: 4 (Doytchinova & Flower, 2007)

Protein no.	Antigenicity Score	Comments
Protein 01	0.6424	Highest Antigenicity , strong potential
Protein 02	0.5473	High antigenicity, Good Potential
Protein 03	0.5232	Low antigenicity, less preferred
Protein 04	0.6220	High antigenicity, Good Potential

Table 1: Protein Based on Antigenicity scores

Antigenicity analysis revealed that Protein 1 (FSCN1) and Protein 4 (Dynein regulatory complex subunit 4) exhibited the highest scores and were ranked as top-priority candidates. Proteins 2 (Tubulin gamma-1 chain) showed moderate antigenicity and were assigned high priority, while Protein 3 (Interstitial collagenase) displayed comparatively lower antigenicity and was considered low priority for further analysis.

3.2 Selection of CTL epitopes

At first we implemented NetCTL 1.2 tool to select CTL epitopes as it helps to precisely predict these sequences considering different criteria. For MHC class I selection A1 supertype was selected along with C Terminal cleavage weight being 0.15, TAP transport efficiency weight being 0.05 and lastly epitope identification threshold being 0.75 (Larsen et al., 2007). We further manually selected "combined score" on sort by score option which generated us with results that have seventeen epitopes with "<-E" assigned to them. This "<-E" represents best predicted epitopes from our provided sequence along with their combined scores. These were accounted for further evaluation. These were then arranged sequentially and placed in NetMHCpan 4.1 immunoinformatic tool to find out compatible binding alleles and gave us 9-mer peptide sequences across every HLA alleles of MHC I class. The threshold for strong binders and weak binders was kept to 0.5 and 2 respectively and binding affinity prediction was selected. The output gave us eleven CTL epitopes after removing redundancy of recurring epitopes.

From the table 1, we can see the assessment of our 9-mer epitopes based on antigenicity using Vaxijen 2.0, allergenicity using AllerTOP v2.1, and toxicity with ToxinPred, all of which are essential bioinformatics tools. We only selected the ones which met the criteria of being antigenic, non-allergenic, and nontoxic.

```
NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

 327 ID fasta pep ASSKNASCY aff   0.5502 aff_rescale   2.3362 cle 0.5813 tap   3.1150 COMB   2.5792 <-E
 403 ID fasta pep TLDANRSSY aff   0.5344 aff_rescale   2.2692 cle 0.9542 tap   2.7170 COMB   2.5481 <-E
 415 ID fasta pep QLEFNDGAY aff   0.4236 aff_rescale   1.7985 cle 0.9458 tap   2.7080 COMB   2.0758 <-E
 178 ID fasta pep TLAFQDQRY aff   0.2819 aff_rescale   1.1967 cle 0.9321 tap   2.9490 COMB   1.4840 <-E
 137 ID fasta pep AMHPQVNIY aff   0.2622 aff_rescale   1.1134 cle 0.9753 tap   3.2290 COMB   1.4211 <-E
 288 ID fasta pep ETDQETFQL aff   0.2623 aff_rescale   1.1137 cle 0.9761 tap   0.3860 COMB   1.2794 <-E
 424 ID fasta pep NIKDSTGKY aff   0.1560 aff_rescale   0.6624 cle 0.9711 tap   3.0360 COMB   0.9599 <-E
 103 ID fasta pep LQSEAHRRY aff   0.1534 aff_rescale   0.6515 cle 0.9540 tap   3.0540 COMB   0.9473 <-E
 410 ID fasta pep SYDVFQLEF aff   0.1579 aff_rescale   0.6703 cle 0.8861 tap   2.7300 COMB   0.9397 <-E
 461 ID fasta pep VAIKVGGRY aff   0.1466 aff_rescale   0.6225 cle 0.9557 tap   3.2090 COMB   0.9263 <-E
 190 ID fasta pep TADHRFLRH aff   0.1858 aff_rescale   0.7889 cle 0.4598 tap  -0.7290 COMB   0.8214 <-E
 368 ID fasta pep ETAGDSELF aff   0.1473 aff_rescale   0.6256 cle 0.4721 tap   2.3730 COMB   0.8150 <-E
 114 ID fasta pep GTEDRLSCF aff   0.1271 aff_rescale   0.5395 cle 0.7725 tap   2.1600 COMB   0.7634 <-E
 355 ID fasta pep VTSKKNGQL aff   0.1336 aff_rescale   0.5673 cle 0.9564 tap   0.8860 COMB   0.7551 <-E


NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

 178 ID fasta pep MSDVVVQPY aff   0.7908 aff_rescale   3.3574 cle 0.9155 tap   2.6690 COMB   3.6282 <-E
 237 ID fasta pep SASTTTLRY aff   0.6076 aff_rescale   2.5798 cle 0.9543 tap   3.0620 COMB   2.8761 <-E
 153 ID fasta pep LLERLNDRY aff   0.4069 aff_rescale   1.7276 cle 0.5403 tap   2.7520 COMB   1.9463 <-E
 277 ID fasta pep TTDQSVASV aff   0.3759 aff_rescale   1.5961 cle 0.9044 tap   0.1440 COMB   1.7389 <-E
 240 ID fasta pep TTTLRYPGY aff   0.3021 aff_rescale   1.2828 cle 0.8971 tap   2.9000 COMB   1.5624 <-E
  44 ID fasta pep GTDRKDVFF aff   0.3151 aff_rescale   1.3379 cle 0.8304 tap   1.9910 COMB   1.5620 <-E
 144 ID fasta pep GTGSGLGSY aff   0.2668 aff_rescale   1.1328 cle 0.9328 tap   2.5310 COMB   1.3993 <-E
 128 ID fasta pep GSDSLEGFV aff   0.2790 aff_rescale   1.1848 cle 0.2844 tap  -0.1390 COMB   1.2205 <-E
 427 ID fasta pep IVQQLIDEY aff   0.2033 aff_rescale   0.8631 cle 0.9637 tap   2.8870 COMB   1.1520 <-E
 265 ID fasta pep RLHFLMTGY aff   0.1490 aff_rescale   0.6325 cle 0.9377 tap   3.0730 COMB   0.9268 <-E
 381 ID fasta pep HTSISSLFE aff   0.2346 aff_rescale   0.9962 cle 0.0324 tap  -1.5410 COMB   0.9240 <-E
  78 ID fasta pep LNSPYAKLY aff   0.1271 aff_rescale   0.5397 cle 0.7258 tap   2.7810 COMB   0.7876 <-E
```

```
NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

 229 ID fasta pep STDIGALMY aff   0.8319 aff_rescale   3.5320 cle 0.9558 tap   2.7570 COMB   3.8133 <-E
 392 ID fasta pep FVANKYWRY aff   0.6323 aff_rescale   2.6845 cle 0.9567 tap   2.8800 COMB   2.9720 <-E
 252 ID fasta pep DIDGIQAIY aff   0.3996 aff_rescale   1.6967 cle 0.9774 tap   2.4220 COMB   1.9644 <-E
 141 ID fasta pep WSNVTPLTF aff   0.2809 aff_rescale   1.1929 cle 0.8292 tap   2.5020 COMB   1.4423 <-E
 113 ID fasta pep HLTYRIENY aff   0.2228 aff_rescale   0.9459 cle 0.9323 tap   2.8270 COMB   1.2271 <-E
 156 ID fasta pep QADIMISFV aff   0.2534 aff_rescale   1.0757 cle 0.8762 tap   0.1740 COMB   1.2159 <-E
  33 ID fasta pep LVQKYLEKY aff   0.2035 aff_rescale   0.8640 cle 0.9536 tap   2.8960 COMB   1.1519 <-E
 108 ID fasta pep RWEQTHLTY aff   0.1583 aff_rescale   0.6720 cle 0.9682 tap   3.1260 COMB   0.9736 <-E
  37 ID fasta pep YLEKYYNLK aff   0.1841 aff_rescale   0.7816 cle 0.8689 tap   0.4210 COMB   0.9329 <-E


NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

 414 ID fasta pep NSTIKDLQY aff   0.5114 aff_rescale   2.1715 cle 0.9661 tap   3.0670 COMB   2.4698 <-E
 230 ID fasta pep FTDIKNYYN aff   0.3300 aff_rescale   1.4010 cle 0.0414 tap  -1.4540 COMB   1.3345 <-E
 393 ID fasta pep NLDPAALTL aff   0.1696 aff_rescale   0.7199 cle 0.9768 tap   0.8890 COMB   0.9109 <-E
  99 ID fasta pep KQKVKHLLY aff   0.1370 aff_rescale   0.5815 cle 0.9585 tap   3.0530 COMB   0.8779 <-E
 338 ID fasta pep VQQERDELY aff   0.1534 aff_rescale   0.6515 cle 0.3273 tap   3.1310 COMB   0.8571 <-E
 292 ID fasta pep EMQKQLANY aff   0.1292 aff_rescale   0.5485 cle 0.9650 tap   2.8640 COMB   0.8364 <-E
```

Figure 6: CTL epitopes generated by NetCTL1.2 (Larsen et al., 2007)

Using the NetMHCpan 4.1 platform, high-affinity MHC class I alleles corresponding to the predicted CTL epitopes were identified for further analysis. The selected alleles are summarized in Table 3 and represent the dominant supertypes derived from the CTL epitope predictions generated by NetCTL 1.2, as listed in Table 2. Based on binding affinity scores, epitopes were classified into two categories: strong binders, defined by binding scores ≤0.500, and weak binders, characterized by scores ≥2.000. This classification strategy enabled the prioritization of CTL epitopes with superior MHC class I binding potential, thereby facilitating the selection of the most immunologically relevant candidates for inclusion in the epithelial cancer vaccine design.

Protein no:1				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*03:01	AMHPQVNIY	0.459	1.214	533.91
HLA-A*26:01	NIKDSTGKY	0.027	0.085	181.07
HLA-A*26:01	ETAGDSELF	0.049	0.045	86.14
HLA-A*01:01	ETDQETFQL	0.262	0.514	1163.20

Protein no:2				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	LLERLNDRY	0.258	0.323	609.46
HLA-A*01:01	GTDRKDVFF	0.358	0.913	2619.05
HLA-A*26:01	RLHFLMTGY	0.225	0.653	456.46
HLA-A*26:01	GTGSGLGSY	0.152	0.245	567.54

Protein no:3				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	DIDGIQAIY	0.143	0.162	225.91
HLA-A*01:01	WSNVTPLTF	0.211	0.202	332.11

Protein no:4				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	NSTIKDLQY	0.040	0.107	133.86
HLA-A*01:01	KQKVKHLLY	0.365	0.252	415.23

Table 2: Summary of predicted CTL epitopes showing sequence identifiers, peptide length, eluted ligand score (EL), percentile ranks for eluted ligand and binding affinity, and MHC class I allele binding affinity as determined using the NetMHCpan 4.1 server.

After confirming the strong binding affinity of the CTL epitopes by NetMHCpan 4.1, their antigenicity, allergenicity, toxicity profile were thoroughly screened using Vaxijen v.2.0 server, Allertop v.2.0 server, Toxinpred. This study used Vaxijen v.2.0 server to predict the immunogenicity of all identified epitopes resulting in the discovery of 2 antigenic epitopes for Protein no. 1, 4 antigenic epitopes for Protein no.2, 2 antigenic epitopes for Protein no.3, 2 antigenic epitopes for Protein no. 4. Moreover, 2 antigenic epitopes for Protein no.1, 4 antigenic epitopes for Protein no.2, 2 antigenic epitopes for Protein no.3, 2 antigenic epitopes for Protein no.4 were found to be non-allergenic predicted by Allertop v.2.0 and non-toxin predicted by Toxinpred. Afterwards, CTL epitopes that satisfied every requirement were selected to be included in the creation of vaccines.

No.	**Selected CTL Epitopes**	**Antigenicity**	**Allergenicity**	**Toxicity**
Protein no:1	AMHPQVNIY	Antigen	Non-Allergen	Non-Toxin
	NIKDSTGKY	Antigen	Non-Allergen	Non-Toxin
Protein no:2	LLERLNDRY	Antigen	Non-Allergen	Non-Toxin
	GTDRKDVFF	Antigen	Non-Allergen	Non-Toxin
	RLHFLMTGY	Antigen	Non-Allergen	Non-Toxin
	GTGSGLGSY	Antigen	Non-Allergen	Non-Toxin
Protein no:3	DIDGIQAIY	Antigen	Non-Allergen	Non-Toxin
	WSNVTPLTF	Antigen	Non-Allergen	Non-Toxin
Protein no:4	NSTIKDLQY	Antigen	Non-Allergen	Non-Toxin
	KQKVKHLLY	Antigen	Non-Allergen	Non-Toxin

Table 3: Antigenicity, Allergenicity, Toxicity determination of the Selected CTL epitopes

3.3 Selection of HTL epitopes

Helper T Lymphocytes not only recognizes antigen displayed on MHC CLASS II molecules but also play an indispensable role in immune response activation by cytokine release like interferon gamma (macrophage activation and antigen presentation), interleukin-4 (induce B cell response promotion) and interleukin-10 (inflammation regulation). NetMHC-IIpan 4.0 tool was employed for achieving HTL epitopes across all alleles where threshold for strong binder was kept 1 (% Rank) and weak binders had 5 (% Rank) and we also incorporated "BA" or binding affinity predictions. The generated results showed us a vast list of 15-mer peptides out of which we chose the "<=SB" designated ones after omitting redundancy. After taking the epitopes, we further screened them for antigenicity, allergenicity and toxicity along with ensuring whether they were interferon gamma positive and IL-4 and IL-10 inducer or not.

Protein No: 1			
HTL Epitopes	IFN	IL-4 Prediction	IL-10 Prediction
TRKRYAHLSARPADE	Positive	IL-4 Inducer	IL-10 Inducer
RKRYAHLSARPADEI	Positive	IL-4 Inducer	IL-10 Inducer
VTRKRYAHLSARPAD	Positive	IL-4 Inducer	IL-10 Inducer

Protein No: 2			
HTL Epitopes	IFN	IL-4 Prediction	IL-10 Prediction
IDEYHAATRPDYISW	Positive	IL-4 Inducer	IL-10 Inducer

Protein No: 3			
HTL Epitopes	IFN	IL-4 Prediction	IL-10 Prediction
QMQEFFGLKVTGKPD	Positive	IL-4 Inducer	IL-10 Inducer
MQEFFGLKVTGKPDA	Positive	IL-4 Inducer	IL-10 Inducer

Protein No: 4			
HTL Epitopes	IFN	IL-4 Prediction	IL-10 Prediction
RQLEEKKAELRNKDR	Positive	IL-4 Inducer	IL-10 Inducer
KHLLYEHQNNLTEMK	Positive	IL-4 Inducer	IL-10 Inducer

Table 4. Helper T-lymphocyte (HTL) epitopes identified from the selected protein candidates and their predicted cytokine-inducing potential.

These HTL epitopes were filtered for antigenicity using Vaxijen 2.0 with threshold 0.5, allergenicity was checked by AllerTOP v2.1 and toxicity by ToxinPred. Additionally, these were screened for its ability to be IFN, IL-4 and IL-10 inducers through the tools IFN epitope server, IL4Pred and IL10Pred respectively. The epitopes which perfectly met these six parameters were chosen for further evaluation.

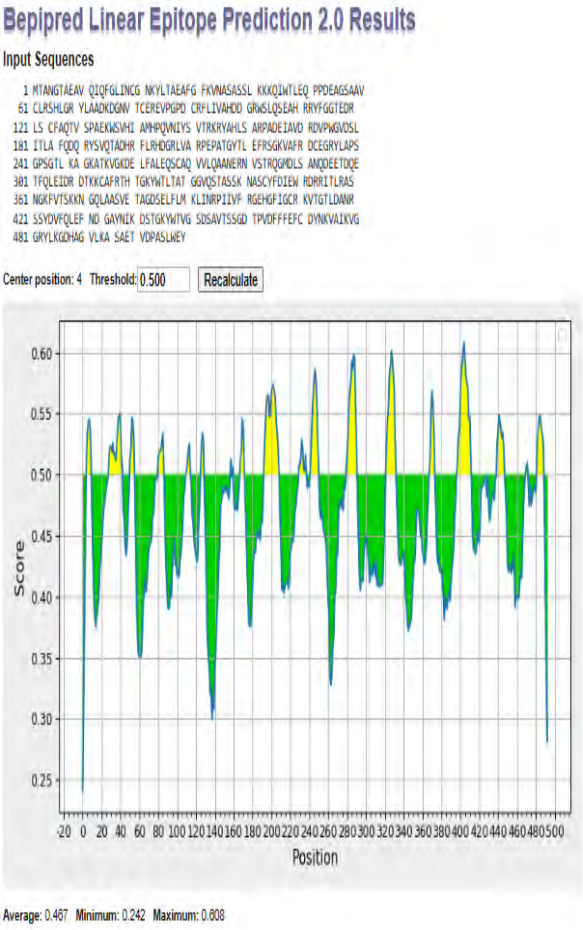
No.	**Selected CTL Epitopes**	**Antigenicity**	**Allergenicity**	**Toxicity**
Protein no:1	VTRKRYAHLSARPAD	Antigen	Non-Allergen	Non-Toxin
Protein no:2	IDEYHAATRPDYISW	Antigen	Non-Allergen	Non-Toxin
Protein no:3	QMQEFFGLKVTGKPD	Antigen	Non-Allergen	Non-Toxin
	MQEFFGLKVTGKPDA	Antigen	Non-Allergen	Non-Toxin
Protein no:4	RQLEEKKAELRNKDR	Antigen	Non-Allergen	Non-Toxin
	KHLLYEHQNNLTEMK	Antigen	Non-Allergen	Non-Toxin

Table 5: Antigenicity, Allergenicity, Toxicity determination of the Selected HTL epitopes

3.4 Selection of B-cell epitopes

B cells or B lymphocytes play the primary role in the generation of humoral immune response as they are involved in antibody production against pathogens. The epitopes of B cells from our sequence were achieved using IEDB Analysis Resource where we submitted our sequence and selected "Bepipred Linear Prediction 2.0" approach (Fleri et al., 2017). The result that we got included a graphical representation of a score versus position graph, where the yellow spikes indicated potential epitopes and green downward spikes indicated non-antigenic epitopes. It also displayed a data box of predicted peptides from which we chose seven to thirty amino acid sequences as this is the optimal range for antibody binding and antigenic stability. Additional assessment was done for assuring antigenicity through Vaxijen 2.0, nonallergenicity through AllerTOP v2.1 and non-toxicity through ToxinPred and only those peptides were taken as B cell epitopes which met all three of these criteria which is shown in the tabular figure below.

Protein 1

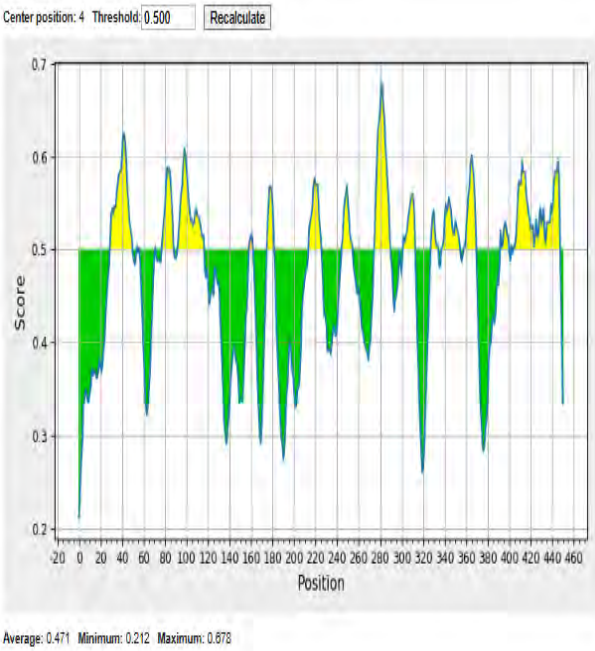


Predicted peptides:

No.	Start	End	Peptide	Length
1	5	10	GTAEAV	6
2	29	43	FGFKVNASASSLKKK	15
3	51	55	PPDEA	5
4	81	87	EREVPGP	7
5	110	115	RYFGGT	6
6	126	129	VSPA	4
7	158	158	R	1
8	160	160	A	1
9	168	172	DVPWG	5
10	193	209	HRFLRHDGRLVARPEPA	17
11	230	238	YLAPSGPSG	9
12	242	251	AGKATKVGKD	10
13	281	291	LSANQDEETDQ	11
14	322	333	GVQSTASSKNAS	12
15	368	373	ETAGDS	6
16	398	412	RKVTGTLDANRSSYD	15
17	440	448	AVTSSGDTP	9
18	470	472	LKG	3
19	482	489	AETVDPAS	8

Figure 7 (a)

Protein 2

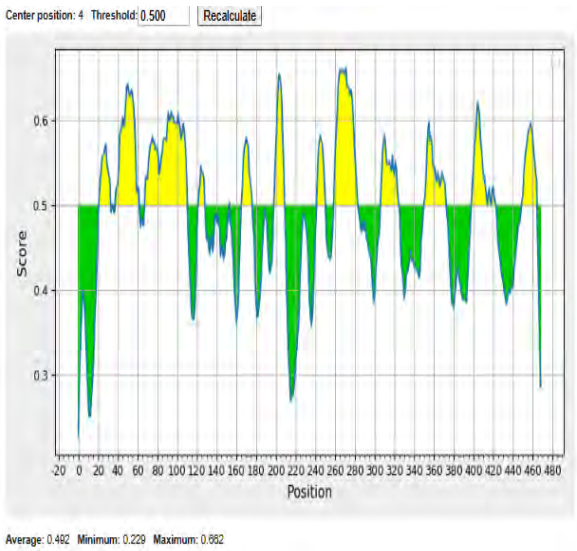


Predicted peptides:

No.	Start	End	Peptide	Length
1	30	50	GISPEGIVEEFATEGTDRKDV	21
2	55	55	A	1
3	57	57	D	1
4	72	72	R	1
5	78	88	LNSPYAKLYNP	11
6	93	117	LSEHGGGAGNNWASGFSQGEKIHED	25
7	159	162	DRYP	4
8	176	182	DEMSDVV	7
9	215	226	TDRLHIQNPSFS	12
10	246	255	PGYMNNDLIG	10
11	276	290	LTTDQSVASVRKTTV	15
12	303	313	VMVSTGRDRQT	11
13	328	335	VDPTQVHK	8
14	338	355	QRIRERKLANFIPWGPAS	18
15	359	370	ALSRKSPYLPSA	12
16	393	400	RQYDKLRK	8
17	403	403	A	1
18	405	448	LEQFRKEDMFKDNFDEMDTSREIVQQLIDEYHAATRPDYISWGT	44

Figure 7 (b)

Protein 3



Predicted peptides:

No.	Start	End	Peptide	Length
1	22	33	ATLETQEQDVDL	12
2	36	36	K	1
3	39	62	EKYYNLKNDGRQVEKRRNSGPVVE	24
4	68	111	QEFFGLKVTGKPDAETLKVMKQPRCGVPDVAQFVLTEGNPRWEQ	44
5	123	129	PDLPRAD	7
6	154	154	E	1
7	166	177	GDHRDNSPFDGP	12
8	199	209	EDERWTNNFRE	11
9	243	251	SGDVQLAQD	9
10	260	283	YGRSQNPVQPIGPQTPKACDSKLT	24
11	307	325	PFYPEVELNFISVFWPQLP	19
12	351	373	VQGQNVLHGYPKDIYSSFGFPRT	23
13	399	423	RYDEYKRSMDPGYPKMIAHDFPGIG	25
14	450	465	KTKRILTLQKANSWFN	16

Figure 7 (c)

Protein 4

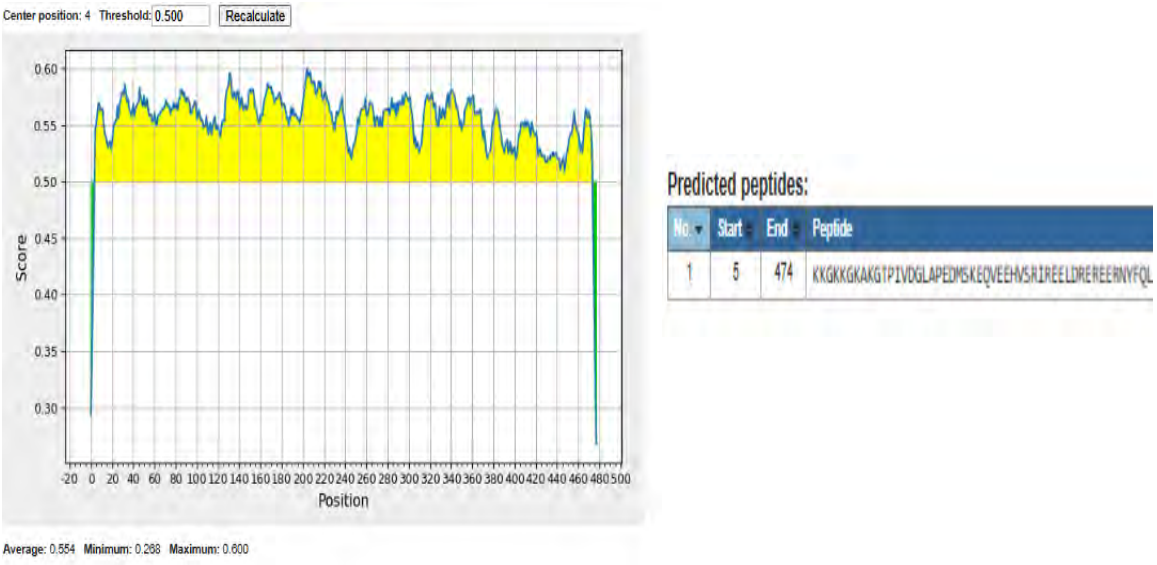


Figure 7 (d)

By analyzing the IEDB Analysis Resource- with Bepipred Linear Epitope Prediction 2.0 technique resulting in the identification of 10 peptides for Protein no. 1, 12 peptides for Protein no.2, 11 peptides for Protein no.3, 0 peptides for Protein no.4 as B cell epitopes. Considering that they have fewer possibilities to stimulate an immunological response, epitopes with incredibly short lengths were removed from the initial evaluation

No.	**Selected CTL Epitopes**	**Antigenicity**	**Allergenicity**	**Toxicity**
Protein no:1	AGKATKVGKD	Antigen	Non-Allergen	Non-Toxin
	AVTSSGDTP	Antigen	Non-Allergen	Non-Toxin
	GVQSTASSKNAS	Antigen	Non-Allergen	Non-Toxin
	AETVDPAS	Antigen	Non-Allergen	Non-Toxin
Protein no:2	GISPEGIVEEFATEGTDRKDV	Antigen	Non-Allergen	Non-Toxin

	LSEHGGGAGNNWASGFSQGEKIHED	Antigen	Non-Allergen	Non-Toxin
	RQYDKLRKKKVMVSTGRDRQT	Antigen	Non-Allergen	Non-Toxin
	QRIRERKLANFIPWGPAS	Antigen	Non-Allergen	Non-Toxin
Protein no:3	ATLETQEQDVDL	Antigen	Non-Allergen	Non-Toxin
	EKYYNLKNDGRQVEKRRNSGPVVE	Antigen	Non-Allergen	Non-Toxin
	YGRSQNPVQPIGPQTPKACDSKLT	Antigen	Non-Allergen	Non-Toxin
	PFYPEVELNFISVFWPQLP	Antigen	Non-Allergen	Non-Toxin
Protein no:4	No Epitopes	NA	NA	NA

Table 6: Antigenicity, Allergenicity, toxicity profile assessment of selected B-Cell epitopes

3.5 Construction of the final vaccine

After getting all the desired CTL, HTL and B cell epitopes which meet all the required criteria, we combined them with our primary sequence with appropriate linkers. Linkers are crucial when we are working with multi epitope vaccine construction as they aid the presentation of each epitope distinctly and retain their structure restoring immunogenicity. This ensures proper folding; hence, structural integrity can be preserved preventing steric hindrance among epitopes. We combined the CTL epitopes with “EAAAK” linker, which is known for its rigidity, HTL epitopes with the flexible linker “GPGPG” which efficiently promoted HTL presentation and lastly “KK” polar linker was used for connecting B cell epitopes which ensured solubility. All of these elements play a key role in making our vaccine construct more stable and effective immunologically (Ayyagari et al., 2022). We took the epitopes with the highest scoring antigenicity to boost our construct’s antigenicity which led us to use one cytotoxic T lymphocyte epitope, one helper T lymphocyte epitope and two B lymphocyte epitopes. The final vaccine design is displayed below.

FINAL VACCINE CONSTRUCTION:

Name	Fasta Sequence
Vaccine Candidate 1	MTANGTAEAVQIQFGLINCGNKYLTAEAFGFKVNASASSLKKKQIWT LEQPPDEAGSAAVCLRSHLGRYLAADKDGNVTCEREVPGPDCRFLI VAHDDGRWSLQSEAHRRYFGGTEDRLSCFAQTVSPAEKWSVHIAMH PQVNIYSVTRKRYAHLSARPADEIAVDRDVPWGVDSLITLAFQDQRY SVQTADHRFLRHDGRLVARPEPATGYTLEFRSGKVAFRDCEGRYLAP SGPSGTLKAGKATKVGKDELFALEQSCAQVVLQAANERNVSTRQG MDLSANQDEETDQETFQLEIDRDTKKCAFRTHTGKYWTLTATGGVQ STASSKNASCYFDIEWRDRRITLRASNGKFVTSKKNGQLAASVETAG DSELFLMKLINRPIIVFRGEHGFIGCRKVTGTLDANRSSYDVFQLEFN DGAYNIKDSTGKYWTVGSDSAVTSSGDTPVDFFFEFCDYNKVAIKV GGRYLKGDHAGVLKASAETVDPASLWEYEAAAKNIKDSTGKYEAA AKAMHPQVNIYGPGPGVTRKRYAHLSARPADKKAGKATKVGKDKK AVTSSGDTPKKGVQSTASSKNASKKAETVDPAS
Vaccine Candidate 2	MPREIITLQLGQCGNQIGFEFWKQLCAEHGISPEGIVEEFATEGTDRKDVF FYQADDEHYIPRAVLLDLEPRVIHSILNSPYAKLYNPENIYLSEHGGGAGN NWASGFSQGEKIHEDIFDIIDREADGSDSLEGFVLCHSIAGGTGSGLGSYL LERLNDRYPKKLVQTYSVFPNQDEMSDVVVQPYNSLLTLKRLTQNADCV VVLDNTALNRIATDRLHIQNPSFSQINQLVSTIMSASTTTLRYPGYMNNDLI GLIASLIPTPRLHFLMTGYTPLTTDQSVASVRKTTVLDVMRRLLQPKNVMV STGRDRQTNHCYIAILNIIQGEVDPTQVHKSLQRIRERKLANFIPWGPASIQ VALSRKSPYLPSAHRVSGLMMANHTSISSLFERTCRQYDKLRKREAFLEQ FRKEDMFKDNFDEMDTSREIVQQLIDEYHAATRPDYISWGTQEQEAAAKL LERLNDRYEAAAKGTDRKDVFFEAAAKRLHFLMTGYEAAAKGTGSGLGS YGPGPGIDEYHAATRPDYISWKKGISPEGIVEEFATEGTDRKDVKKLSEH GGGAGNNWASGFSQGEKIHEDKKRQYDKLRKKKVMVSTGRDRQTKKQ RIRERKLANFIPWGPAS
Vaccine Candidate 3	MHSFPPLLLLLFWGVVSHSFPATLETQEQDVDLVQKYLEKYYNLKN DGRQVEKRRNSGPVVEKLKQMQEFFGLKVTGKPDAETLKVMKQPR CGVPDVAQFVLTEGNPRWEQTHLTYRIENYTPDLPRADVDHAIEKA FQLWSNVTPLTFTKVSEGQADIMISFVRGDHRDNSPFDGPGGNLAH AFQPGPGIGGDAHFDEDERWTNNFREYNLHRVAAHELGHSLGLSHS TDIGALMYPSYTFSGDVQLAQDDIDGIQAIYGRSQNPVQPIGPQTPK ACDSKLTFDAITTIRGEVMFFKDRFYMRTNPFYPEVELNFISVFWPQ LPNGLEAAYEFADRDEVRFFKGNKYWAVQGQNVLHGYPKDIYSSF GFPRTVKHIDAALSEENTGKTYFFVANKYWRYDEYKRSMDPGYPK MIAHDFPGIGHKVDAVFMKDGFFYFFHGTRQYKFDPKTKRILTLQK ANSWFNCRKNEAAAKDIDGIQAIYEAAAKWSNVTPLTFGPGPGQM QEFFGLKVTGKPDGPGPGMQEFFGLKVTGKPDAKKATLETQEQDV DLKKEKYYNLKNDGRQVEKRRNSGPVVEKKYGRSQNPVQPIGPQT PKACDSKLTKKPFYPEVELNFISVFWPQLP

Vaccine Candidate 4	MAPKKKGKKGKAKGTPIVDGLAPEDMSKEQVEEHVSRIREELDRER EERNYFQLERDKIHTFWEITRRQLEEKKAELRNKDREMEEAEERHQ VEIKVYKQKVKHLLYEHQNNLTEMKAEGTVVMKLAQKEHRIQESV LRKDMRALKVELKEQELASEVVVKNLRLKHTEEITRMRNDFERQV REIEAKYDKKMKMLRDELDLRRKTELHEVEERKNGQIHTLMQRHE EAFTDIKNYYNDITLNNLALINSLKEQMEDMRKKEDHLEREMAEVS GQNKRLADPLQKAREEMSEMQKQLANYERDKQILLCTKARLKVRE KELKDLQWEHEVLEQRFTKVQQERDELYRKFTAAIQEVQQKTGFK NLVLERKLQALSAAVEKKEVQFNEVLAASNLDPAALTLVSRKLEDV LESKNSTIKDLQYELAQVCKAHNDLLRTYEAKLLAFGIPLDNVGFKP LETAVIGQTLGQGPAGLVGTPTEAAAKNSTIKDLQYEAAAKKQKVK HLLYGPGPGRQLEEKKAELRNKDRGPGPGKHLLYEHQNNLTEMK

After preparation of the constructed vaccine, all of them were assessed for their antigenicity, allergenicity and Toxicity to observe whether they demonstrate immunogenic properties. The antigenicity of vaccine candidates verified using Vaxijen, allergenicity by Allergen online server & toxicity by T3DB server.

No	Antigenecity	Presence of allergen	Presence of Toxicity
Vaccine Candidate 1	0.6968	Number of sequence with hits:0	Your search returned no results
Vaccine Candidate 2	0.6347	Number of sequence with hits:4	Your search returned no results
Vaccine Candidate 3	0.5636	Number of sequence with hits:0	1 Toxic
Vaccine Candidate 4	0.6538	Number of sequence with hits:0	Your search returned no results

Table 7: Assessment of the constructed vaccines

3.6 Biochemical analysis of the constructed vaccine

To assess the biochemical characteristics of the vaccines that have been developed, an online server called Protparam was used. This server provides information whether the vaccine is

eligible or not by giving results of physicochemical parameters like: molecular weight, molecular formula, instability index, GRAVY ,amino acid number, molecular weight, therapeutic pI.

Parameters	Vaccine Candidate 1	Vaccine Candidate 2	Vaccine Candidate 3	Vaccine Candidate 4
Molecular Weight	64305.05	69895.99	71173.47	64004.29
Theoretical PI	8.93	7.00	7.27	8.76
Instability Index	32.71	28.29	34.60	36.57
Aliphatic Index	66.45	78.84	64.05	81.47
GRAVY (Grand Average of Hydropathicity)	-0.534	-0.532	-0.626	-0.982

Table 10: Physico-chemical parameters of the constructed vaccines using Protparam Tool.

ProtParam - Results

User-provided sequence:

```
        10         20         30         40         50         60
MTANGTAEAV QIQFGLINCG NKYLTAEAFG FKVNASASSL KKKQIWTLEQ PPDEAGSAAV

        70         80         90        100        110        120
CLRSHLGRYL AADKDGNVTC EREVPGPDCR FLIVAHDDGR WSLQSEAHRR YFGGTEDRLS

       130        140        150        160        170        180
CFAQTVSPAE KWSVHIAMHP QVNIYSVTRK RYAHLSARPA DEIAVDRDVP WGVDSLITLA

       190        200        210        220        230        240
FQDQRYSVQT ADHRFLRHDG RLVARPEPAT GYTLEFRSGK VAFRDCEGRY LAPSGPSGTL

       250        260        270        280        290        300
KAGKATKVGK DELFALEQSC AQVVLQAANE RNVSTRQGMD LSANQDEETD QETFQLEIDR

       310        320        330        340        350        360
DTKKCAFRTH TGKYWTLTAT GGVQSTASSK NASCYFDIEW RDRRITLRAS NGKFVTSKKN

       370        380        390        400        410        420
GQLAASVETA GDSELFLMKL INRPIIVFRG EHGFIGCRKV TGTLDANRSS YDVFQLEFND

       430        440        450        460        470        480
GAYNIKDSTG KYWTVGSDSA VTSSGDTPVD FFFEFCDYNK VAIKVGGRYL KGDHAGVLKA

       490        500        510        520        530        540
SAETVDPASL WEYEAAAKNI KDSTGKYEAA AKAMHPQVNI YGPGPGVTRK RYAHLSARPA

       550        560        570        580
DKKAGKATKV GKDKKAVTSS GDTPKKGVQS TASSKNASKK AETVDPAS
```

Number of amino acids: 588
Theoretical pI: 8.93
Molecular weight: 64307.05

Formula: $C_{2819}H_{4425}N_{817}O_{877}S_{16}$
Total number of atoms: 8954

Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 72935
Abs 0.1% (=1 g/l) 1.134, assuming all pairs of Cys residues form cystines

Ext. coefficient 72310
Abs 0.1% (=1 g/l) 1.124, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 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 32.71
This classifies the protein as stable.

Aliphatic index: 66.45
Grand average of hydropathicity (GRAVY):-0.534

Figure 8: Biochemical Property Analysis values by ProtParam tool (Wilkins et al., 1999)

3.7 3D model of the constructed vaccine

For homology modelling of our sequence and translating it into a 3D structure we used the immunoinformatic tool Phyre 2.2 where we pasted our sequence and it generated us a 3D model in PDB format. For assuring reliability, it also has two important parameters which are confidence percentage and coverage percentage. Structural visualization was done by the software PyMOL. The model it generated based on our input had a coverage score of 82% (80% or more is desired) where 483 residues of our sequence had been modelled with 100% confidence further validating our constructed multi epitope vaccine**.**

Parameters	**Vaccine Candidate 1**	**Vaccine Candidate 2**	**Vaccine Candidate 3**	**Vaccine Candidate 4**
Confidence	100%	100%	100%	99.7%
Coverage	82%	70%	58%	83%

Table 11: Coverage and Confidence determination of the Constructed Vaccines using PHYRE-2

From the table we can see, that Vaccine candidate 1 showed coverage around 82% and vaccine 4 shows 83% but its confidence is lower than 100. Rest of the other vaccine candidates showed results less than this (less than 80%) which was not desirable for this study. The confidence and Coverage for vaccine candidate 1 is shown below:

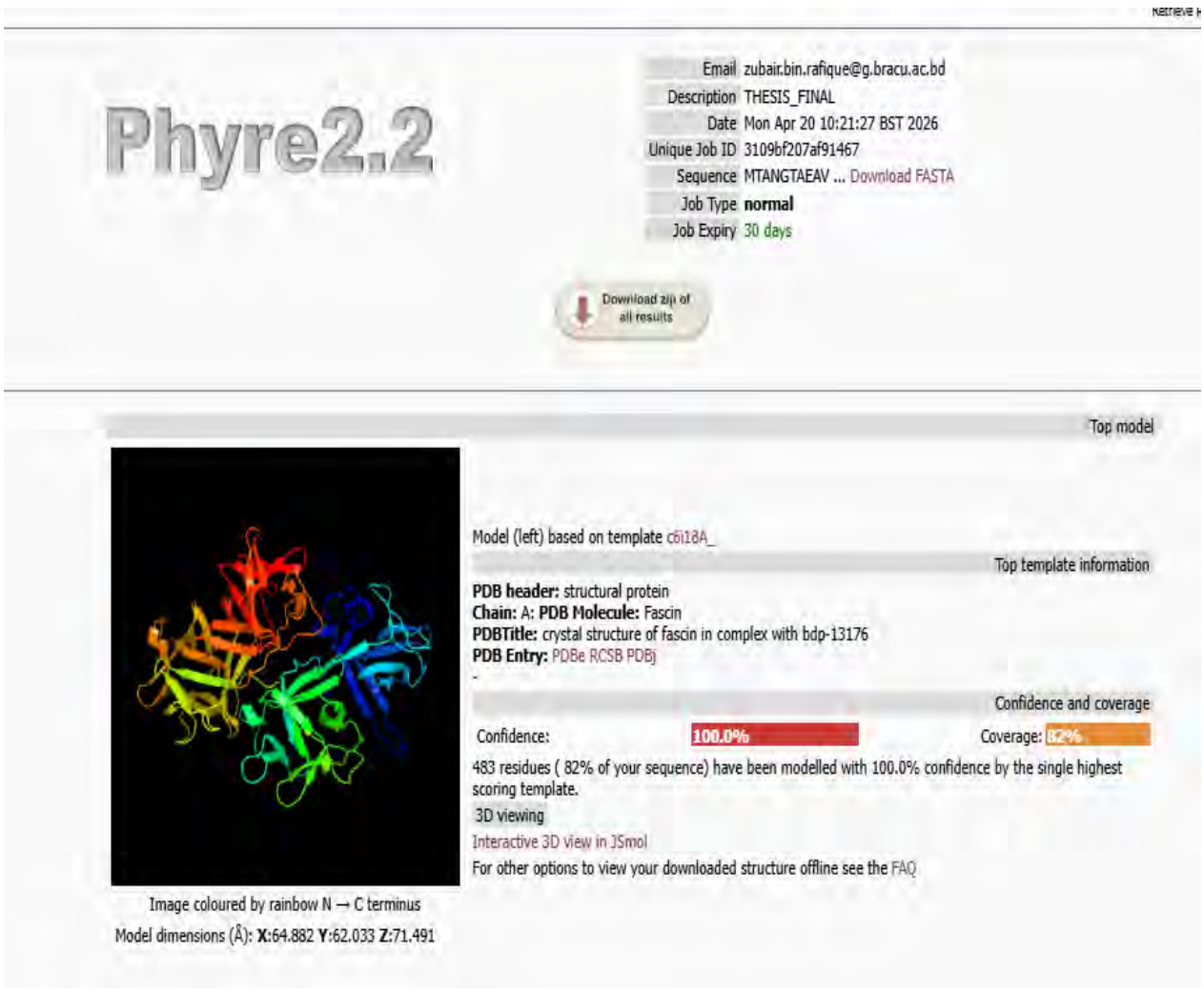


Figure 9: Phyre 2.2 generated 3D model with 82% residue coverage at 100% confidence (Kelley et al., 2015).

3.8 Analysis of Z-score

Determination of Z-score was done by ProSa Web which stands for protein structure analysis, compares our PDB formatted model with real and experimentally validated structures. It has two distinctly colored regions for X-ray crystallography and NMR spectroscopy which tells us which approach our structure falls. The graph is plotted against residue numbers versus Z-score (Wiederstein & Sippl, 2007). After inputting our Phyre's PDB structure in ProSa web, the results generated Z-score which falls within the acceptable range of -5 to -10. Our model also falls in the light blue x ray region which aids to validate high resolution structure. Local model quality graph (knowledge-based energy versus sequence position) is also generated where most of the residues take place in the negative region further reiterating our Z-score validation.

Z Scores	Vaccine Candidate 1	Vaccine Candidate 2	Vaccine Candidate 3	Vaccine Candidate 4
	-7.52	-9.48	-7.09	-10.71

Table 12: Z-score evaluation of the predicted vaccine constructs obtained using the ProSA-web structural validation tool

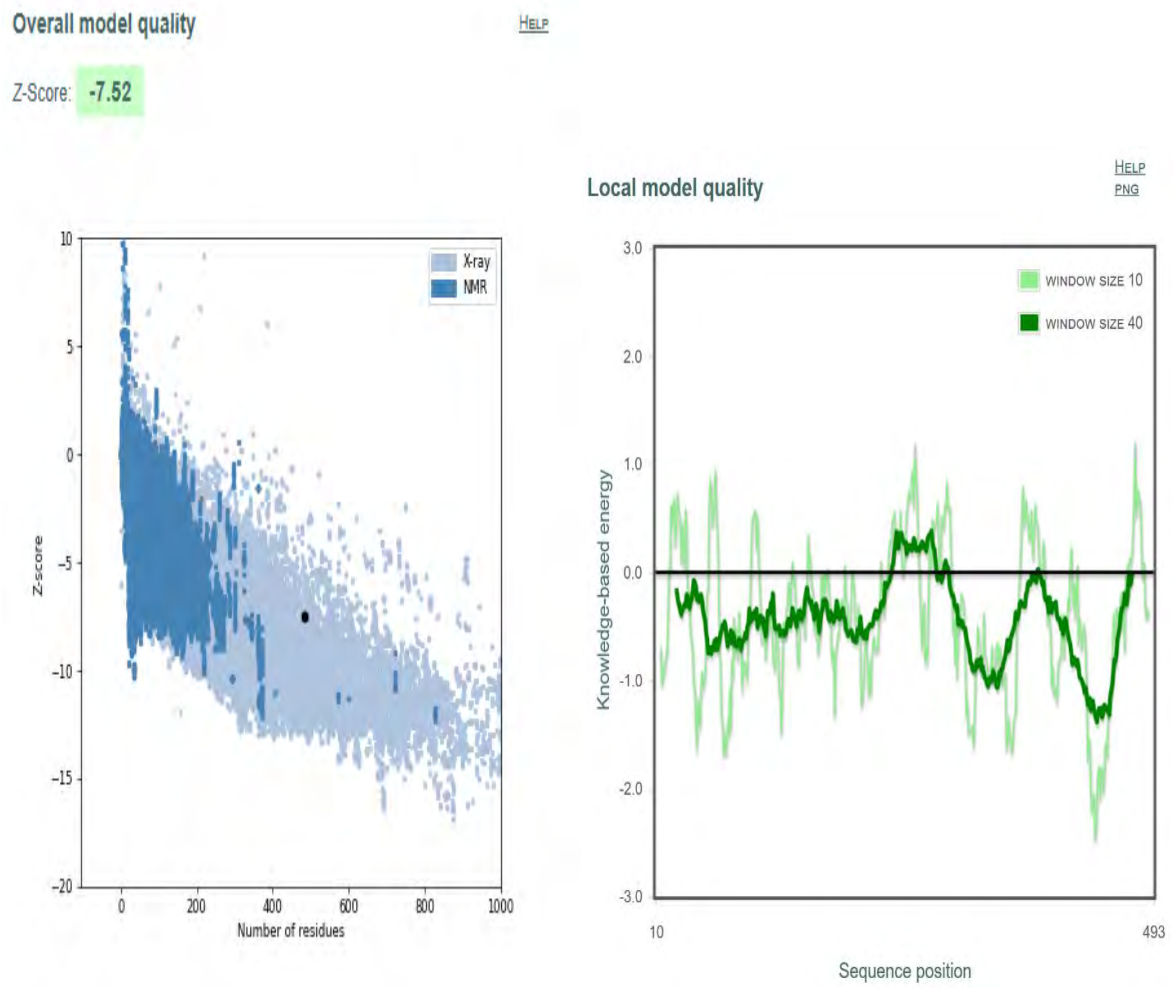
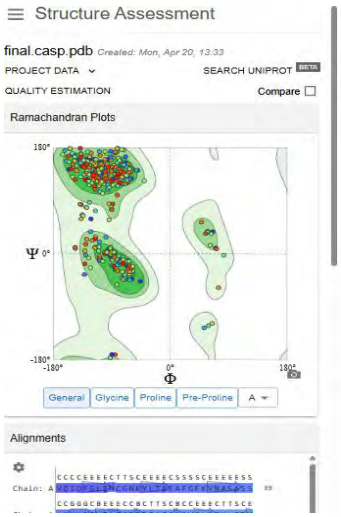


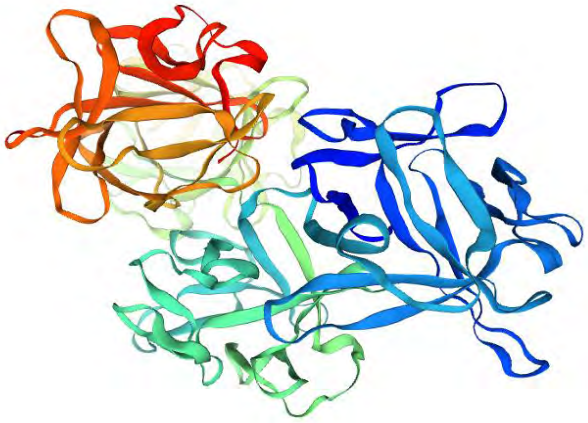
Figure 10: Z-score graph and local model quality of vaccine Candidate *I by ProSa Web (Wiederstein & Sippl, 2007)*

3.9 Ramachandran Plotting

Employment of Swiss Model Expasy server was done in order to estimate the stereochemistry of our generated PDB model obtained from Phyre 2.2. This is required to ensure the protein's

correct folding and overall structural reliability. The data achieved from Swiss Model Expasy is given below:





MolProbity Results

MolProbity Score	2.32
Clash Score	37.32

(A169 VAL-A171 TRP), (A51 PRO-A59 ALA), (A41 LYS-A44 GLN), (A167 ARG-A169 VAL), (A154 HIS-A166 ASP), (A74 LYS-A108 HIS), (A43 LYS-A65 HIS), (A173 VAL-A381 ILE), (A171 TRP-A310 HIS), (A162 GLU-A239 THR), (A294 PHE-A315 TRP), (A356 THR-A367 VAL), (A31 PHE-A68 ARG), (A40 LEU-A44 GLN), (A171 TRP-A309 THR), (A197 ARG-A199 ASP), (A179 LEU-A411 TYR), (A145 TYR-A150 LYS), (A397 CYS-A404 LEU), (A298 ILE-A457 ASP), (A193 HIS-A385 ILE), (A281 LEU-A315 TRP), (A141 GLN-A383 ARG), (A296 LEU-A307 PHE), (A418 PHE-A421 GLY), (A13 GLN-A45 ILE), (A280 ASP-A281 LEU), (A261 ALA-A493 TYR), (A414 PHE-A433 TRP), (A485 VAL-A490 LEU), (A215 GLU-A248 VAL), (A386 ILE-A388 PHE), (A190 THR-A192 ASP), (A425 ILE-A433 TRP), (A340 TRP-A453 PHE), (A69 TYR-A83 GLU), (A259 SER-A383 ARG), (A396 GLY-A408 ARG), (A37 ALA-A39 SER), (A374 GLU-A374 GLU), (A225 ASP-A225 ASP), (A18 ASN-A35 ALA), (A305 CYS-A338 ILE), (A149 ARG-A238 GLY), (A468 ARG-A480 ALA), (A387 VAL-A411 TYR), (A303 LYS-A457 ASP)

Ramachandran Favoured	96.06%
Ramachandran Outliers	0.41%

Figure 11: Ramachandran plot generation for vaccine Candidate 1 along with scores and model by Swiss Model Expasy (Kiefer et al., 2009)

We can see that in vaccine 1 Ramachandran favored score is 96.06% which is exceptionally high as its required to be above 90% and Ramachandran outliers is 0.41% which is required to be below 2%. What these basically mean is that more than 96% of our sequences residues falls in the energetically favored regions which are the green regions in the graph indicating good folding. Analyzation of this was done by phi (φ) and psi (ψ) dihedral angles distribution. The outliers meaning disallowed portions are only 0.41% indicating less possibility of any structural

error or strain. C-Beta deviations are also in acceptable range (< 0.25 Å). All the four's results in given below:

Parameters	**Vaccine Candidate 1**	**Vaccine Candidate 2**	**Vaccine Candidate 3**	**Vaccine Candidate 4**
Ramachandran Favored	96.06%	94.24%	87.05%	98.00%
Ramachandran Outlier	0.41%	0.00%	3.58%	0.00%
Rotamer outliers	0.00%	0.00%	0.32%	0.00%
MolProbity	2.32	2.47	2.88	1.59

Table 13 : Ramachandran Plot Prediction for all the vaccines

Parameters	**Server**	**Vaccine Candidate 1**	**Vaccine Candidate 2**	**Vaccine Candidate 3**	**Vaccine Candidate 4**
Antigenicity of vaccine	Vaxijen V.2.0	0.6968	0.6347	0.5636	0.6538
Allergenicity	Allergen Online	None	4	None	None
Toxicity	T3DB	None	None	One	None
Theoretical Isoelectric Point (pI)	Protparam	8.93	7.00	7.27	8.76
Instability Index		32.71	28.29	34.60	36.57
GRAVY		-0.534	-0.532	-0.626	-0.982
Confidence	PHYRE-2	100%	100%	100%	99.7%
Coverage		82%	70%	58%	83%
Ramachandran Favored	Ramachandran SWISS MODEL	96.06%	94.24%	87.05%	98.00%
Ramachandran Outlier		0.41%	0.00%	3.58%	0.00%

Table 14: Summary table of the all the parameters assessed for the vaccine candidates

Based on the summary table, Vaccine Candidate 1 satisfies the required prerequisites. Additionally, the values of the physicochemical parameters were within the acceptable range: the theoretical pI value was within the range of 8–10 (8.93), the instability index (32.71) was lower than 40, and the GRAVY value was negative (-0.534). These results suggest that the vaccine is stable and possesses hydrophilic characteristics. The vaccine also demonstrated strong antigenicity with a score of 0.6968, indicating its potential to induce an immune response. Furthermore, it did not exhibit any allergenicity or toxicity.

The structural validation parameters also supported the reliability of the model. The confidence score was 100% with coverage of 82%. In addition, the Ramachandran favored region exceeded 90% (96.27%) and no Ramachandran outliers were observed (0.41%), suggesting good stereochemical quality of the predicted structure.

In comparison, Vaccine Candidate 2 & Vaccine Candidate 3 also demonstrated high antigenicity and a perfect confidence score (100%), although it exhibited lower coverage (70%,58%). While its physicochemical properties remained acceptable, the presence of predicted toxicity and allergen hits and comparatively reduced structural coverage made it less favorable than Vaccine Candidate 1.

Although Vaccine Candidate 4 also exhibited favorable characteristics, but its confidence is lower than 100% which is not acceptable. Vaccine Candidate 1 demonstrated comparatively better antigenicity and structural reliability among the candidates. The main objective of this study was to evaluate each vaccine construct individually based on antigenicity, physicochemical properties, and structural stability, and then compare them to identify the most suitable vaccine candidate.

Overall, Vaccine candidate 1 showed higher antigenicity and a satisfactory percentage of residues within the favored regions of the Ramachandran plot, indicating that it may serve as a promising candidate for further investigation. Therefore, Vaccine candidate 1 was selected for subsequent analyses to evaluate its immune response potential and binding affinity. To validate this, molecular docking and immune simulation studies were performed to further assess its suitability for vaccine development.

3.10 Molecular Docking of the Constructed Vaccine

Using ClusPro software as a docking tool to establish whether the vaccine candidate construct binds a receptor, we determined whether the vaccine would be able to initiate an immune response through the formation of a docking site between the vaccine and a receptor (Kiflu,

2024).The ClusPro software program contains a receptor box and ligand box. To find a potential target for our vaccine, we chose Toll-like receptor 3 (TLR3). In the PDB format, we obtained the TLR3 structure from the RCSB Protein Data Bank. As a ligand, we used the Phyre2.2 PDB structure and ran a docking procedure (Tombari et al., 2025). As a result, several models were generated from docked structures, each having an associated model score. Model 0 is the one that best illustrates the central pose of our 0 cluster. Cluster 0 is usually very reliable since it has the greatest number of members (106) and the most common docking orientation. The average energy of our central conformation is -655.2, while the most stable and lowest-energy conformation is -735.0. Even though cluster 3 is lower in energy, it was rejected due to its small number of members. Therefore, we chose the 0 cluster because it represents a high binding energy together with many docking opportunities, and thus it is very reliable.

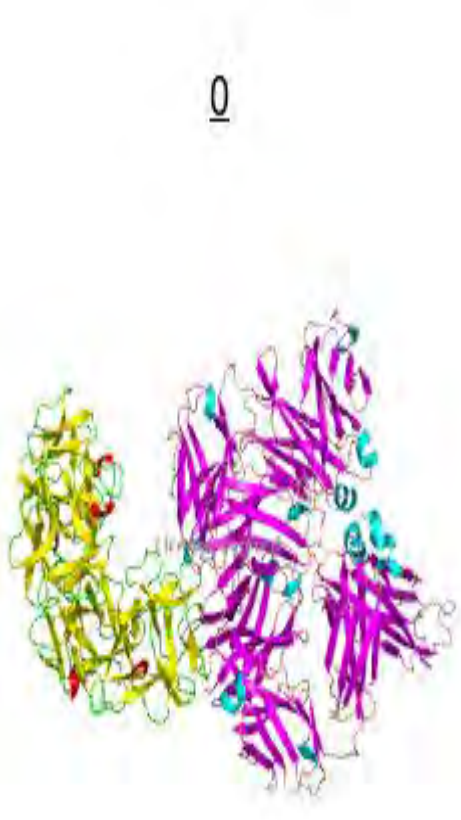


Figure 12: Docked complex achieved by ClusPro (Comeau et al., 2004)

Cluster	Members	Representative	Weighted Score
0	106	Center	-655.2
		Lowest Energy	-735.0
1	87	Center	-787.1
		Lowest Energy	-787.1
2	80	Center	-640.6
		Lowest Energy	-760.9
3	67	Center	-614.7
		Lowest Energy	-772.3
4	56	Center	-612.8
		Lowest Energy	-744.1
5	52	Center	-648.6
		Lowest Energy	-710.0
6	48	Center	-632.1
		Lowest Energy	-710.0
7	27	Center	-724.0
		Lowest Energy	-796.6

Figure 13: Model energy scores achieved by ClusPro (Comeau et al., 2004)

3.11 Immune Simulation of the Constructed Vaccine

Our constructed vaccine sequence was uploaded in an online simulator server titled C-ImmSim 10.1 where simulation parameters were adjusted to 300 and time-steps for three doses. Results obtained from the simulation were presented in graphs that predicted immune responses elicited by our multiepitope vaccine. From the first graph we have 3 black spikes that show introduction of antigens as doses. In the graph provided, days are plotted on the X-axis while antigen/antibody titers are on the Y-axis with black spikes showing introduction of antigens through doses while right Y-axis represents antibody titers. On day 0, we observe a spike where the concentration of antigen peaks while the same is primed for its recurrence. Antibodies developed in response to antigens are IgM, which is represented by green lines, and total antibody responses represented by the yellow line shows a moderate increase indicating that the vaccine elicited primary immune responses. When the second dose is administered on day 28, IgM and total antibody response (represented by the yellow line) increases drastically indicating secondary immune responses have been induced. The high prevalence of IgG guarantees that humoral immunity was sustained. Also, we observe increased levels of IgM (green line) and IgG1+IgG2 (blue line). As a result, we can observe an even greater IgM+IgG response and much faster antigen clearance at days 58–60 after the third dose was given. Significant spikes are also observed in the other immunoglobulins. Antigen levels show a

remarkable and faster decline following each treatment. We may conclude that our rabies vaccination was effective since it produced memory cells for additional protection against re-exposure and triggered a robust immunogenic response. The figure is given below:

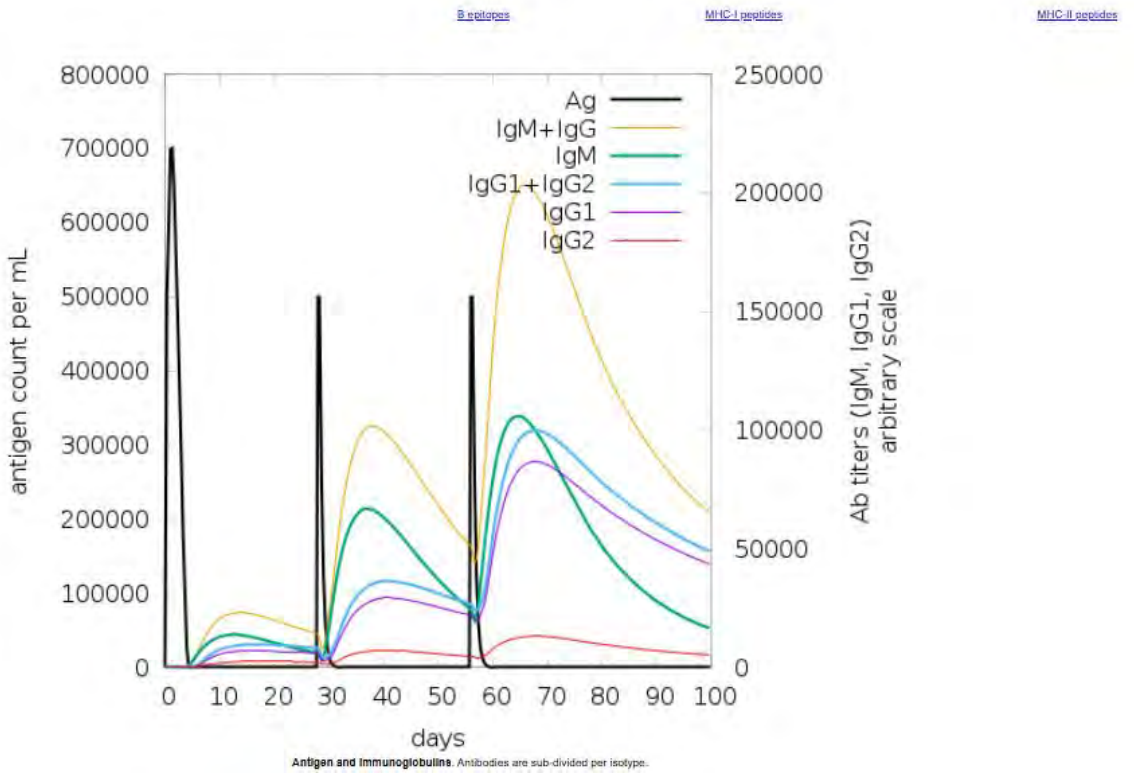


Figure 13: C-ImmSim generated Immune Simulation for vaccine candidate 1 (Rapin et al., 2010)

From the graph of cytokines and interleukins, there are very high levels of interferon gamma and interleukins 4, 10, and 12. This implies that there is a Th1 response involving cellular activity, a Th2 type of humoral response, regulation of inflammation, and promotion of Th1 differentiation to produce interferon gamma. An inset box indicates spikes in interleukin 2 which implies the growth of T cells.

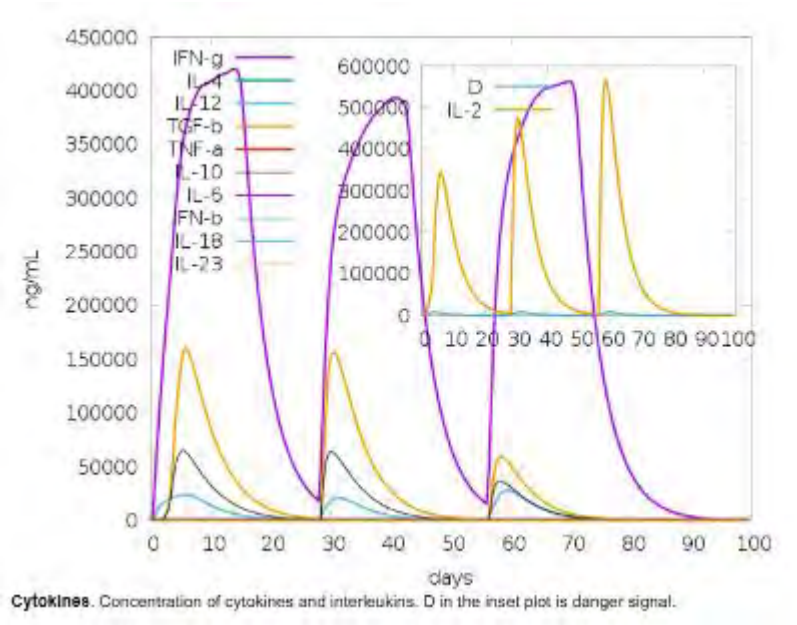


Figure 14: Cytokine and Interleukin Concentration graph containing IFN gamma, Interleukin 4 and 10 (Rapin et al., 2010)

The next graph that we considered was that of B lymphocytes. It shows that the number of B cells increases implying that the vaccine is effective in inducing a humoral immune response to provide long-term immunity through memory cells. Isotype switching demonstrates the maturity of antibodies as follows: IgM increases indicating the beginning of the immune response. Afterwards, a steady increase in IgG occurs indicating the maturation of B cells. Finally, green spikes indicate that memory cells offer protection on re-exposure.

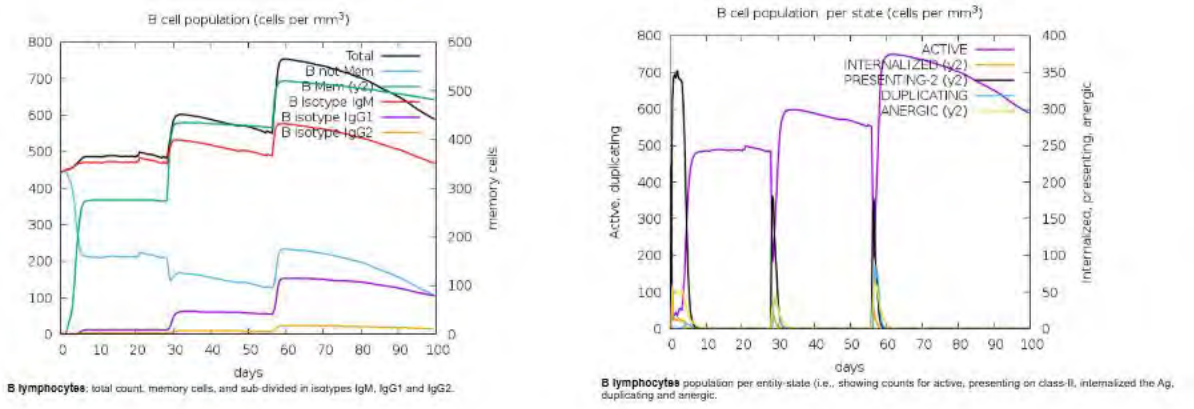


Figure 15: B cell Lymphocyte population graph (Rapin et al., 2010)

The other group of graphs involves CD4+ T helper cells. It clearly indicates that there are rises in the memory, active and dividing T helper cells after the immunization process. This implies that the vaccine can help activate the T helper cells.

.

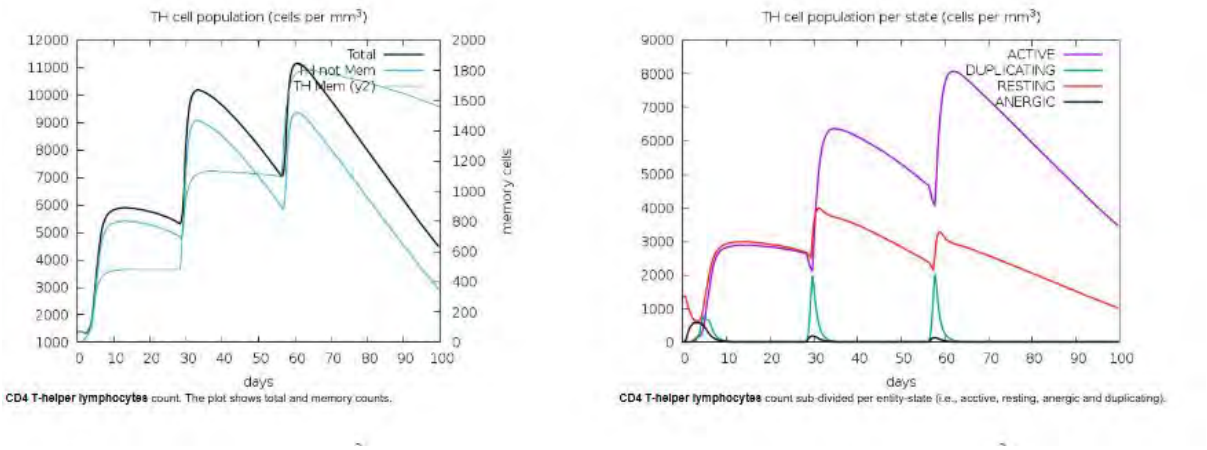


Figure 16: CD4 T helper Lymphocyte population graph (Rapin et al., 2010)

The final graph considers CD8+ cytotoxic lymphocytes. It shows that there is a steady rise in active cells while the resting cells drop gradually implying that it offers a sustainable immune response. A consistent number of anergic cells further confirms a sustainable baseline in the immune response.

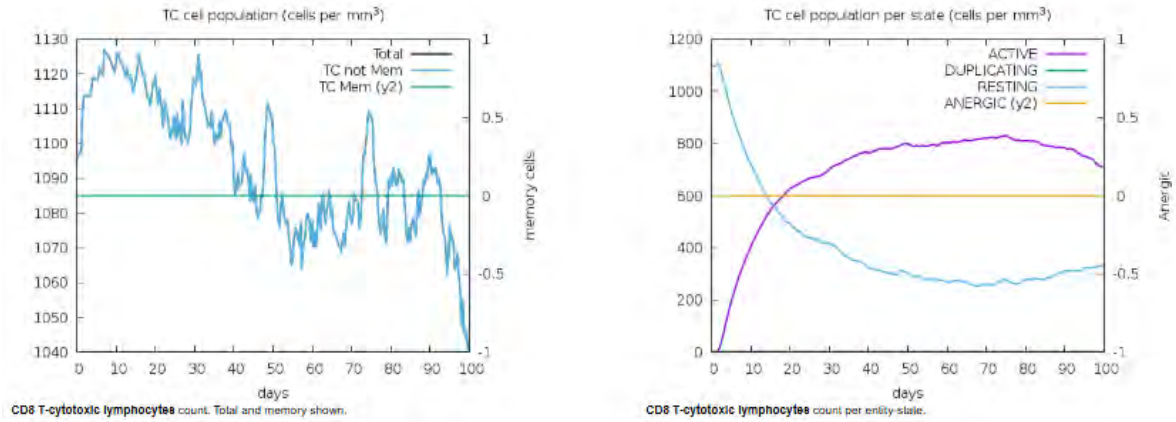


Figure 17: CD8 cytotoxic T Lymphocyte population graph (Rapin et al., 2010)

To summarize, all graphs obtained from C-ImmSim show that the vaccine can induce cellular and humoral immunity over time depending on the vaccine doses. This means that the vaccine helps build immunity to fight against the Rabies virus. The memory cells ensure that long-term immunity is attained. In the combined graph, the levels of the antigens rise three times and clearance of these antigens occurs at increasingly steep peaks meaning that clearance becomes increasingly efficient with time.

Chapter 4

Discussion

Epithelial tumors are one of the most common cancers throughout the world. They include breast, colorectal, lung, and cervical tumors. Most importantly, these types of cancers originate from epithelial tissue, usually diagnosed later, disseminate easily, and are associated with unfavorable prognoses. Moreover, in such countries as Bangladesh, patients suffer not only from a higher prevalence but also have less access to early screening, inadequate healthcare infrastructure, increased exposure to environmental and lifestyle risk factors, and higher incidence rates among young patients. Thus, the development of effective and cost-effective therapy options is crucial for improving public health. In these conditions, cancer vaccines appear to be one of the most perspective treatment strategies due to their ability to induce specific, long-term, and targeted antitumor responses within the host body.Although conventional cancer therapy options such as chemotherapy and radiotherapy could be efficient, they may cause considerable side effects, impact healthy tissues, and induce resistance to treatment. Immunotherapies, including vaccines, represent a promising alternative to the aforementioned treatment strategies. They rely on the ability of the patient's immune system to recognize and attack cancer cells. The integration of computational (in silico) methods into vaccine design revolutionized this area of research and reduced the duration and costs related to identifying, evaluating, and testing potential candidates.An epitope-based vaccine was designed for treatment of epithelial cancers via comparative immunoinformatics. Different proteins involved in the progression and metastasis of cancer were analyzed in order to identify the best protein among them. FSCN1_HUMAN protein scored the highest. This protein is highly expressed in epithelial tumors and promotes tumor growth, migration, and metastasis by facilitating motility and invasion. FSCN1_HUMAN is not a universal housekeeping protein such as tubulin gamma-1, and also, it is not found in any other DRC4 categories.Epitope sequences of the selected proteins were predicted, analyzed, and filtered through several parameters: antigenicity, allergenicity, toxicity, physicochemical properties, and stability. Antigenicity was calculated by the use of VaxiJen 2.0 software and only epitopes with high antigenicity values were included. Evaluation of allergenicity and toxicity was done to ensure safety of the developed vaccine, which is vital in immunotherapy. Analysis results indicated that FSCN1-derived epitopes had superior antigenic properties compared to those derived from other proteins, indicating their high immunogenicity potential.The selection of the best vaccine

design candidate was made based on various parameters. High antigenicity value, optimal epitope coverage rate, and favorable physicochemical properties were observed in FSCN1 compared to other studied proteins. Such parameters as molecular weight, stability index, and hydrophobicity influence stability and solubility of the vaccine, thus having considerable impact on ease of expression, purification, and storage. Stable and soluble vaccines can be easily manufactured on a large scale, making them cost-effective in underdeveloped areas such as Bangladesh. Modeling and validation have been done to establish the appropriateness of the selected vaccine candidate. The three-dimensional structure of the protein and the final vaccine design were predicted through software such as Phyre2, which provided insights into the structure and stability of the protein. Validation was done based on Z-scores and Ramachandran plots, which showed that most of the amino acids fell within favored regions, thereby establishing structural stability. Docking studies were conducted to determine whether the vaccine structure would have binding affinity to immune receptors such as toll-like receptors (TLR). The role of TLRs is significant in initiating the process of innate immune response as well as bridging innate immunity with the adaptive immune system. Docking studies demonstrated that there was good binding affinity between the vaccine and TLRs, indicating that the vaccine had the potential to activate immunological signaling pathways effectively**.** Also, there were simulations conducted using computer programs that predicted the effect of the vaccine. The outcome indicated antibodies such as IgG and IgM, as well as activation of the CD4+ helper T cells and CD8+ killer T cells. This is critical in cancer immunotherapy since it involves direct destruction of the tumor cells and formation of an adaptive immunity against the cancer cells. The graph plotted by the simulations revealed consistent immunity response following multiple vaccinations, hence proving the vaccine's immunogenicity. Conclusively, it is essential to determine the suitable cancer vaccine antigen through appropriate assessment. While several proteins could be used as candidates for an immunogenic cancer vaccine, most lack ideal properties. This finding highlights the need for a careful and systematic evaluation procedure. Additionally, FSCN1's lack of investigation as an epitope vaccine further demonstrates the necessity of filling knowledge gaps in existing cancer immunotherapy protocols for developing new therapies. However, there are limitations worth noting. The study is entirely based on computational analysis. While helpful in understanding biological processes, it cannot provide a complete insight into the complex mechanisms that occur. Such factors as antigen processing, immunological rules, and the tumor microenvironment may influence the vaccine's functionality within the host body. Hence, future investigations should

use cellular and animal experiments to determine its safety and efficiency. Differences among populations in MHC gene frequency should also be considered since they could impact its efficiency. In conclusion, the research reveals the efficiency of designing a multi-epitope vaccine based on immunoinformatics in epithelial cancer immunotherapy. By comparing different antigens, it is possible to identify the ideal candidate for such therapy. In the case under consideration, computational tools allowed for a comprehensive evaluation of numerous factors in a fast and effective way. Although more laboratory studies are required, they will contribute to the existing research concerning cancer vaccines, which demonstrate promising results in developing innovative treatment methods for countries like Bangladesh.

Chapter 5

Conclusion

In summary, this research investigated a novel concept of a vaccine against epithelial cancer based on the comparison of four proteins that consist of more than one antigen. In order to identify a protein as a vaccine target, we used computer-based analysis and selected the sequences of epitopes from each of the target proteins. Next, we assessed each of these candidate proteins concerning vaccine development potential. To do this, we analyzed antigenicity, allergenicity, toxicity, and other characteristics such as physicochemical properties and stability of each candidate. Lastly, we considered each option and chose the best protein for the development of a cancer vaccine. Fascin1 was identified as the most viable vaccine candidate due to its high antigenicity, excellent physicochemical properties, and stability of structure.

Reference:

Ayyagari, V. S., T. C., V., K., A. P., & Srirama, K. (2022). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: An immunoinformatics approach.

Burden and Trends of Cancer in Bangladesh: Insights into Breast, Lung, Cervical, and Liver Cancers and Implications for Public Health Interventions. (2023).

Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: An immunoinformatics approach: Journal of Biomolecular Structure and Dynamics.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines.

Esfahani, K., Roudaia, L., Buhlaiga, N., Del Rincon, S. V., Papneja, N., & Miller, W. H. (2020). A Review of Cancer Immunotherapy: From the Past, to the Present, to the Future.

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design.

Huang, J., Deng, Y., Tin, M. S., Lok, V., Ngai, C. H., Zhang, L., Lucero-Prisno, D. E., Xu, W., Zheng, Z.-J., Elcarte, E., Withers, M., & Wong, M. C. S. (2022). Distribution, Risk Factors, and Temporal Trends for Lung Cancer Incidence and Mortality.

Igarashi, Y., & Sasada, T. (2020). Cancer Vaccines: Toward the Next Breakthrough in Cancer Immunotherapy.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis.

Kiflu, A. B. (2024). The Immune Escape Strategy of Rabies Virus and Its Pathogenicity

Kundu, S., Chakrabarty, S., Chakrabortty, P., Nirob, M. H., & Monjur-E-Elahi, M. (2025). Regional cancer burdens: A retrospective analysis of demographic and diagnostic patterns in Bangladesh.

Lahiri, A., Maji, A., Potdar, P. D., Singh, N., Parikh, P., Bisht, B., Mukherjee, A., & Paul, M. K. (2023). Lung cancer immunotherapy: Progress, pitfalls, and promises.

Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction | BMC Bioinformatics | Springer Nature Link. (n.d.).

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.

Mahdevar, E., Safavi, A., Abiri, A., Kefayat, A., Hejazi, S. H., Miresmaeili, S. M., & Iranpur Mobarakeh, V. (2022). Exploring the cancer-testis antigen BORIS to design a novel multi-epitope vaccine against breast cancer based on immunoinformatics approaches.

Motamedi, H., Ari, M. M., Shahlaei, M., Moradi, S., Farhadikia, P., Alvandi, A., & Abiri, R. (2023). Designing multi-epitope vaccine against important colorectal cancer (CRC) associated pathogens based on immunoinformatics approach.

Nessa, A., Hussain, T., Alam, S. M., Faruk, I., & Jahan, I. (2018). Age distribution pattern of female breast cancer patients in Bangladesh- developing early and presenting late.

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System.

Sharma, R., Abbastabar, H., Abdulah, D. M., Abidi, H., Abolhassani, H., Abrehdari-Tafreshi, Z., Absalan, A., Ali, H. A., Abu-Gharbieh, E., Acuna, J. M., Adib, N., Sakilah Adnani, Q. E., Aghaei, A., Ahmad, A., Ahmad, S., Ahmadi, A., Ahmadi, S., Ahmed, L. A., Ajami, M., … Zoladl, M. (2024). Temporal patterns of cancer burden in Asia, 1990–2019: A systematic examination for the Global Burden of Disease 2019 study.

Shehri, Z. S. A. (2025). Immuno-informatics analyses of important esophageal cancer associated viruses for multi-epitope vaccine design.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.

Tombari, W., Khamessi, O., Othman, H., Kallala, O., Mahjoub, R., Ghedira, K., & Trabelsi, A. (2025). A novel mRNA-based multi-epitope vaccine for rabies virus computationally designed via reverse vaccinology and immunoinformatics.

Urban, K., Mehrmal, S., Uppal, P., Giesey, R. L., & Delost, G. R. (2021). The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study.

Wang, L., Zhou, X., Wei, Y., & Lin, J. (2025). A novel multi-epitope mRNA vaccine against colorectal cancer: In silico design and immune efficacy profiling.

Warnakulasuriya, S., & Filho, A. M. (2025). Oral Cancer in the South and South-East Asia Region, 2022: Incidence and Mortality.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein Identification and Analysis Tools in the ExPASy Server.

Yaseen, A. R., Suleman, M., Alzahrani, K. J., Alsharif, K. F., & Alzahrani, F. M. (2025). Unleashing the future of cancer immunotherapy: In silico design of a multi-epitope and mRNA vaccine duo targeting EWSR1-ATF1, EWSR1-CREB1, and PRAME to conquer clear cell sarcoma using immunoinformatics approaches.

Zhou, P., Shi, X., Xia, J., & Hu, H. (2025). In silico design and in vitro validation of a multi-epitope peptide vaccine targeting triple-negative breast cancer.