In-Silico Based Multiepitope Vaccine Construction of Envelope Glycoprotein against Hantavirus

By

Samira Siddika
ID: 20146058

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
April 2024

© 2024 Brac University
All rights reserved.

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Samira Siddika
Student ID: 20146058

Approval

The thesis titled “In-Silico Based Multiepitope Vaccine Construction of Envelope Glycoprotein against Hantavirus” submitted by Samira Siddika (ID: 20146058) of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (B.pharm) on 02.05.2024.

Supervised By:

Supervisor:

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
Brac University

Approved By:

Assistant Dean:

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

Departmental Head:
(Dean)

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

Ethics Statement

The completion of this thesis was maintained all the ethical standard that were required. No involvement of any human or animal trials.

Abstract

Hantavirus is a zoonotic virus which is responsible for causing hemorrhagic fever with renal syndrome & hantavirus pulmonary syndrome. But still, there is no stable vaccine developed for this virus. Developing a multiepitope vaccine candidate against this virus was the main concern of this study by using immune-informative approach. Different computational tools were used to identify the CTL, HTL & B cell epitopes for envelope surface glycoprotein. Various linkers were used to connect the epitopes with the adjuvant and construct the candidate vaccine. Physicochemical properties like - molecular weight, instability index, aliphatic index, GRAVY was also checked for the prepared vaccine. Molecular docking was done with toll-like receptor 4 (TLR-4) for checking the residual interaction. And lastly, C-ImmSim was used to evaluate the immune stimulation. Further investigation is needed to check the safety and efficacy of this vaccine candidate because checking safety & efficacy is not a part of in-silico study.

Keywords: Hantavirus; Epitope; CTL; HTL; B cell; In-silico; Envelope glycoprotein

Acknowledgement

I would like to thank my respected supervisor Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University. He guided me and advised me throughout my research work. The research work would not be completed successfully without the guidance and valuable feedback of my supervisor.

Samira Siddika
ID: 20146058
School of Pharmacy
Brac University

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Hantavirus belongs to the Bunyaviridae family of viruses, but it has more than 50 other species. All strains of hantavirus do not cause diseases. But the diseases caused by hantavirus is very serious and now it is a worldwide concern like the coronavirus. The outbreak of hantavirus which was called “Field Nephritis” starts from World War 1 in both German & Allied troops and the outbreak of field nephritis again visualized in World War 2. In 1951, more than 3000 Korean soldiers were affected with hantavirus during the Korean War, and they suffered from hemorrhagic fever. However, they named the virus as hantavirus because this virus was found in the Hantaan River of Korea. The hantavirus has two types – one is Old World Hantavirus, and another is New World Hantavirus. Old World Hantavirus is responsible for causing the disease Hemorrhagic Fever with Renal Syndrome (HFRS) whereas the New World Hantavirus is responsible for causing Hantavirus Pulmonary Syndrome (HPS) (Ghafoor et al., 2021). In Europe and Asia, the disease caused by hantavirus is Old World Hantavirus which includes Seoul, Puumala, and Dobrava. The targeted organ is the kidney and the mortality rate is 1-15% (Ghafoor et al., 2021). However, New World Hantavirus includes – Andes virus & Sin Nombre virus were found in four corners of America in 1993 (Ghafoor et al., 2021). The targeted organ is the lung and the mortality rate is 40%. The pathology is almost similar for both HPS & HFRS like – vascular leakage and thrombocytopenia (Ghafoor et al., 2021). The rate is increasing day by day of affecting hantavirus in China and Europe. It occurs more in males compared to females (Abdulla et al., 2021). Hantavirus is generally associated with rodents and humans are affected by inhaling the aerosols of rodent waste like urine, feces, and saliva. (Joshi et al., 2022). However, person to person contact is very rare. There is no proper vaccine available for hantavirus. Hantavax is the vaccine that is used in Korea, and it is not authorized by FDA. Not

all the country use this vaccine other than Korea because the formulation is derived from rodent brains. (Joshi et al., 2022).

The immune response is stimulated by the vaccine when it produces antibodies and memory cells. Immuno-informative approaches are used in in-silico method to determine to select and screen the suitable multiepitope vaccine construction (Ghafoor et al., 2021)

1.1 Genomic Structure

Hantavirus is a negative sense RNA virus & it is single-stranded. The virions of hantavirus have spherical shapes and the size of the virions is 80-120nm. The genome of hantavirus is divided into three different segments which are S segment, M segment, L segment. Based on their size they are considered as S, L & M segments. The S portion encodes nucleoprotein, the M portion encodes glycoprotein (Gn & Gc), and the L portion encodes RNA-dependent RNA polymerase (Muyangwa et al., 2015, Ghafoor et al., 2021).

1.2 Life Cycle of Hantavirus

Replication of hantavirus occurs in the cytoplasm of the host cell showed in figure 01. Through endocytosis, the virions of the hantavirus attach to the cellular receptors and enter the host cell (Muyangwa et al., 2015). During that time an interaction occurs between the glycoproteins and cellular receptors. Then pH-mediated fusion occurs, and the virions become uncoated & release of viral genome which are ribonucleoproteins (RNPs) to the cytoplasm. Transcription occurs in the endoplasmic reticulum which is the compartment of Golgi apparatus, and a capped primer is required which is formed by the cleavage of mRNAs to initiate the transcription and the RdRp participates in this cleavage (Muyangwa et al., 2015). Replication occurs in the cytoplasm and produces cRNA during final elongation which works as a template to produce huge amount of full-length negative sense vRNA(viral RNA) and assembly with glycoprotein that transported to the Golgi apparatus (Muyangwa et al., 2015). Translation occurs by

producing nucleoprotein, glycoprotein and RNA dependent RNA polymerase where viral mRNAs are translated (Muyangwa et al., 2015). In the plasma membrane egress occurs through the fusion of Golgi vessel.

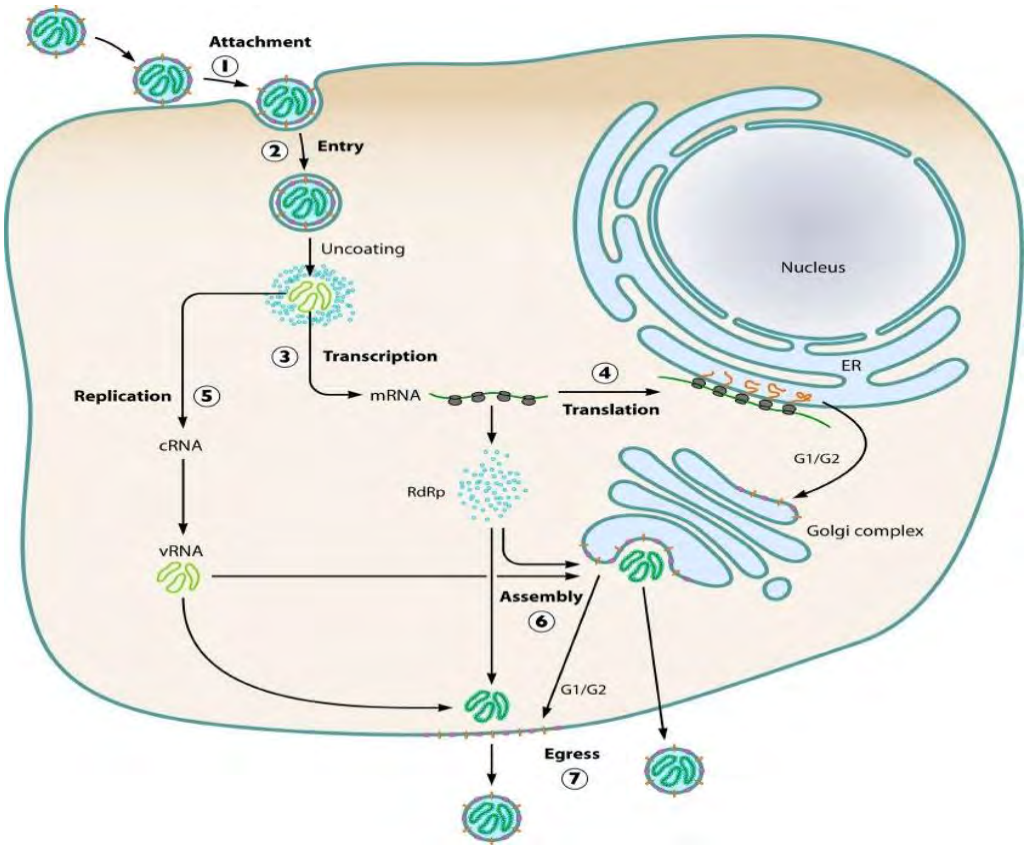


Figure 01: Life cycle of hantavirus (Muyangwa et al., 2015)

Chapter 2

Methodology

The mentioned steps were maintained to generate the final vaccine

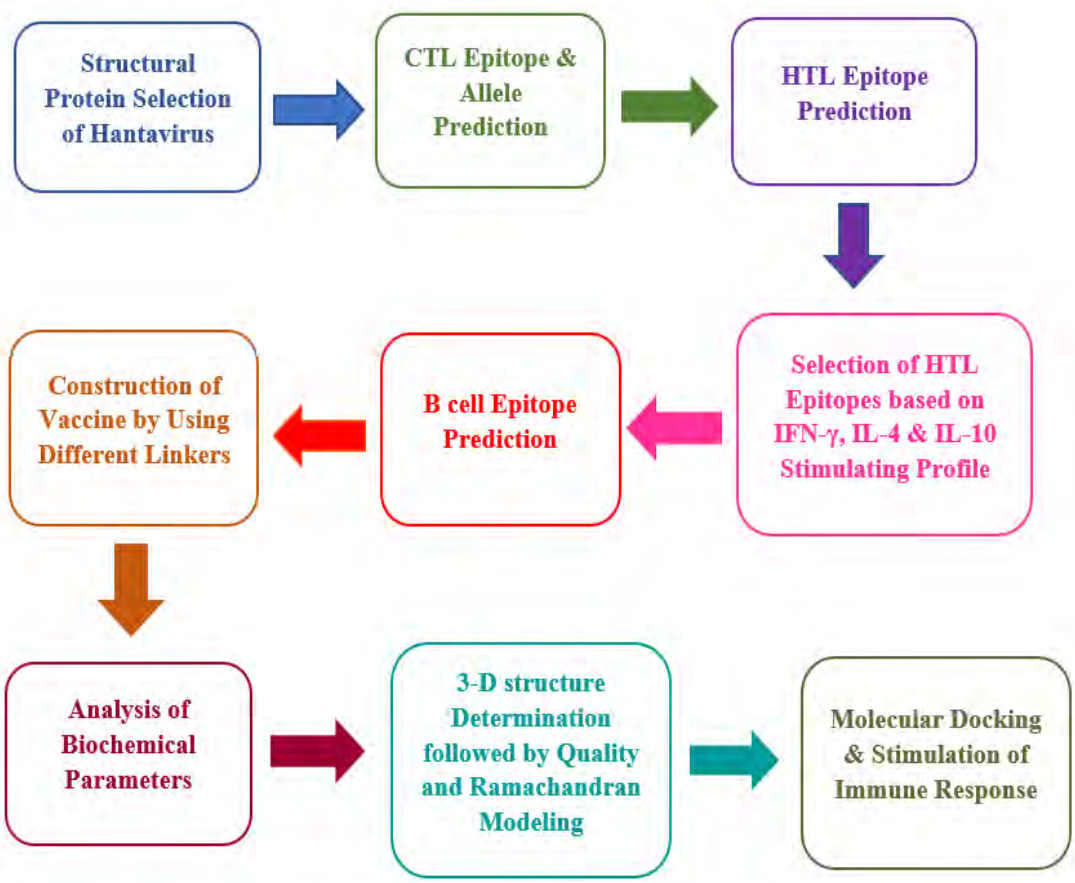


Figure 02: Schematic presentation of the whole process used for in-silico vaccine construction

2.1 Selection of protein Sequence

For the selection of protein sequence, UniProt protein database was used at https://www.uniprot.org/ (Bateman et al., 2023). From this server, different protein sequences were collected. This protein database contains 9016 protein sequences for hantavirus where 63 sets are reviewed (Swiss-Prot) and 8953 sets are unreviewed (TrEMBL). The antigenicity was checked to select the protein sequence and that's why the protein sequences were downloaded in fasta format. For checking the antigenicity, Vaxijen v.2.0 server was used (Doytchinova & Flower, 2007) where targeted organism virus was selected & 0.5 threshold was maintained

otherwise there is a chance to get the false positive result. Protein sequence was selected based on highest score of antigenicity for continued the further steps and the protein was a M polyprotein (Glycoprotein).

2.2 Determination of CTL Epitopes:

For inducing immune response in host cells CTL epitopes are important. To determine these CTL epitopes NetCTL1.2 was used (Larsen et al., 2007). Three parameters of this tool are TAP transport efficiency, MHC-I binding peptides and C-terminal cleavage. The proteasomal C-terminal cleavage was predicted by artificial neural networks whereas the weight matrix was used to predict the TAP transport efficiency (Larsen et al., 2007). The server NetCTL1.2 is able to detect CTL epitopes up to 12 major histocompatibility complex class I subtype (MHC-I) where in FASTA format the sequence of protein was given & 0.75 threshold was maintained for the selection of epitope.

2.3 Evaluation of CTL Epitopes:

NetMHCpan4.1 was used to evaluate the MHC-I binding alleles that are specific to CTL epitopes (Reynisson et al., 2020). This server uses an artificial neural network to predict the MHC molecules (Reynisson et al., 2020). For peptide length 9 mer peptide was selected and all the HLA were selected. For strong protein binding the threshold 0.5 was maintained & for weak protein binding the threshold 2 was selected. The server selected BA (Binding Affinity) predict & sort by prediction score. The output page showed the results that fell within the threshold. But the peptides which had strong protein binding were selected. Then these peptides were checked for antigenicity, allergenicity and toxicity. To check antigenicity server Vaxijen v.2.0 was (Doytchinova & Flower, 2007). To determine the allergenicity, an AllerTop v.2.0 server was used (Dimitrov et al., 2014) and for toxicity the server ToxinPred was used (Gupta et al., 2013).

2.4 Identification of HTL Epitopes:

HTL Epitopes are important for the adaptive immune system. To identify the HTL epitopes NetMHC-IIpan server was used (Reynisson et al., 2020). This server predicts the HTL epitopes by using artificial neural networks. The protein sequence was given in Fasta format. 15 mer peptide was chosen for peptide length and selected maximum 20 alleles per submission. This process continued until the alleles were selected. For strong binder the threshold was 1% & for weak binder the threshold was 5%. The result page showed the alleles which water falls within this threshold and chose those alleles which had high binding affinity.

2.5 Evaluation of HTL Epitopes

IFN- γ, IL-4 and IL-10 these three cytokines are released from HTL. There are other cytokines which are released from helper T lymphocytes but these three are very important for the body's immune system. They can determine whether the body produces the immune response after giving the vaccine or not. The interferon gamma activates MHC molecules. To determine whether the HTL epitopes were IFN-gamma inducer or not, server IFNepitope was used (Kupani et al., 2023). The interleukin 4 helps the B cell maturation and proliferation. To check HTL epitopes were IL-4 inducer or not, the server IL-4 pred was used (Dhanda et al., 2013). Interleukin 10 is important for maintaining the homeostasis and to check HTL epitopes were IL-10 inducer or not, the server IL-10 pred was used (Singh et al., 2022). Those HTL epitopes that maintain all the criteria were further checked for antigenicity, allergenicity and toxicity. After checking all these parameters, HTL epitopes were IFN- γ inducer, IL4 positive and IL10 positive & antigen, non-allergen, non-toxin selected.

2.6 Prediction and Evaluation of B cell epitopes

To predict the B cell epitopes IEDB Resource Analysis server was used (Sun et al., 2013) “Bepipred Linear prediction 2.0” this method was selected for the prediction of B cell epitopes (Sun et al., 2013). The protein sequence was given, and the output page showed a graph along

with a table based on antigenicity. For B cell prediction 0.5 threshold was maintained. Those epitopes that fall within this range were selected. After that these epitopes were checked for antigenicity, allergenicity and toxicity like the previous one.

2.7 Constitution of Vaccine and Evaluate Antigenicity

Figure 03: Vaccine construction sequence using different linkers

In-silico multi epitope vaccine was prepared by using different linkers with the primary protein sequence. Usually, to prepare a vaccine four different linkers are used. But here three linkers were used instead of four. To connect the primary protein sequence (adjuvant) to CTL epitopes EAAAK linker was used. Similarly, to connect CTL to HTL GPGPG linker was used. And to connect HTL to B cell epitopes KK linker was used. After constructing the vaccine, antigenicity of the prepared vaccine was checked by using VaxiJen v.2.0 server (Doytchinova & Flower, 2007).

2.8 Biochemical Analysis of the Constructed Vaccine

To check the physicochemical properties of the constructed vaccine Protparam server was used (Garg et al., 2016). This server is able to provide information about the number of amino acids, molecular weight, instability index, GRAVY and so on.

2.9 Evaluation of Allergenicity & Toxicity of the Constructed vaccine

To check the allergenicity of the constructed vaccine, an allergen online server was used (Goodman et al., 2016) where the constructed vaccine sequence was given in plain format. However, for checking whether the constructed vaccine was toxic or nontoxic T3DB server was used (Wishart et al., 2015).

2.10 Creation of 3D modeling of the Constructed vaccine

To generate a homology or 3D modeling of the constructed vaccine Phyre2 server was used (Kelley et al., 2015) where the sequence was given in a plain format. Primary and tertiary structure was determined by this server. The main concern is to determine the confidence and coverage of the constructed vaccine. The result was sent via an email where a PDB file was present. Pymol software was used to access the PDB file to see the 3D structure.

2.11 Generation and Evaluation of Ramachandran Plots

To check the 3D model validity Ramachandran plot was generated by using Swiss Model Expasy (Schwede et al., 2003) in structure assessment method. The PDB file was uploaded which was sent via an email. Phi and psi were the angels of the Ramachandran plot which were plotted against each other. For a particular protein it determines the polypeptide chain (Schwede et al., 2003).

2.12 Generation and Evaluation of Z-Score

To determine the z-score of the constructed vaccine, ProSA-web was used (Wiederstein & Sippl, 2007). This server was able to access the 3D model of quality of the vaccine and energy distribution throughout the model. The PDB file was uploaded to the server. Z score, z-score vs number of residues graph, a knowledge-based energy vs sequence position graphs were shown in the output page.

2.13 Molecular Docking of the Constructed Vaccine

Molecular docking was performed to check the binding affinity of the constructed vaccine and the suitable human toll-like receptor. ClusPro was used at https://cluspro.bu.edu/ (Comeau et al., 2004) to check the binding affinity. The PDB file for toll-like receptor was downloaded from RCSB database and uploaded in the receptor molecule & PDB which was obtained via

mail was uploaded in the ligand molecule. The result page showed 10 docked complexes where the highest energy scoring docked complex was downloaded.

2.14 Immune Response Stimulation

C-ImmSim server was used to generate several graphs that provide graphical information about the levels of antigen, lymphocytes, immunoglobulins and so on. This server basically provides the graphical details about the humoral & cellular response of the immune system after giving the vaccine doses. In this server, the vaccine sequence was given in fasta format, and 3 injections were added where the stimulation step was 300 selected and time step of injections were 1, 84 & 168 respectively. One day time step equals eight hours.

2.15 Interpose on Methodology

Development of vaccine using in-silico approach was the main of this research where screening, prediction and different analysis were performed by using online tools. This technique is quite effective but unable to measure the parameters of efficacy and safety of the developed vaccine. So that further investigation is necessary to know whether this vaccine is a suitable candidate against hantavirus or not.

Chapter 3

Result

3.1 Protein Sequence selection based on Antigenicity

M polyprotein (Envelope Surface Glycoprotein) of hantavirus was selected based on antigenicity because it showed highest antigenicity during primary screening. The sequence of the selected protein was given below:

GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT
SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVT
IKYTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTC
QFGDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFT
MNKLEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFI
TAIKACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTA
SDNHVYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRG

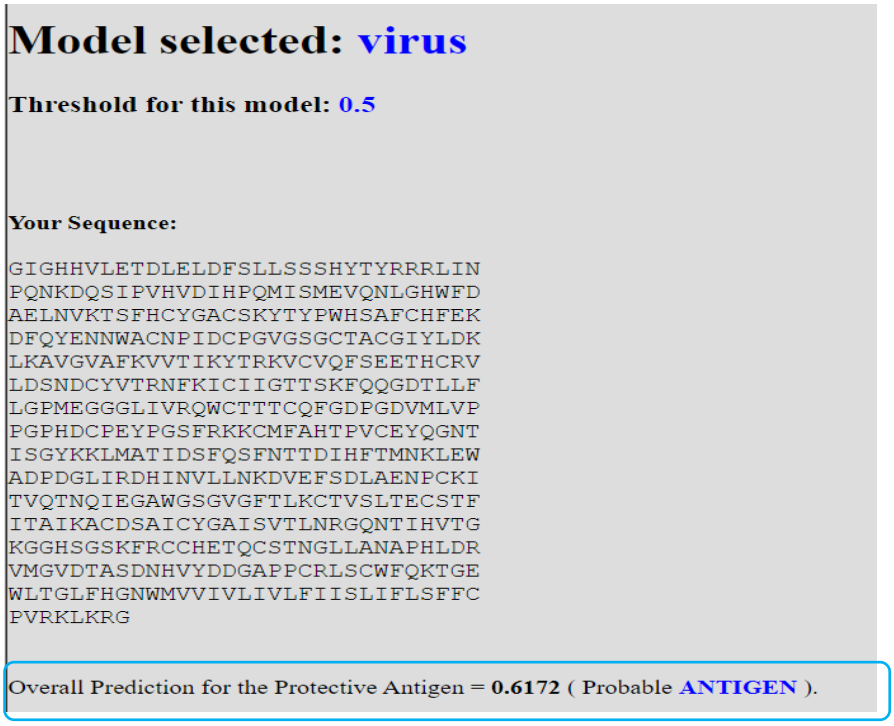


Figure 04: Antigenicity of the protein sequence (Doytchinova & Flower, 2007)

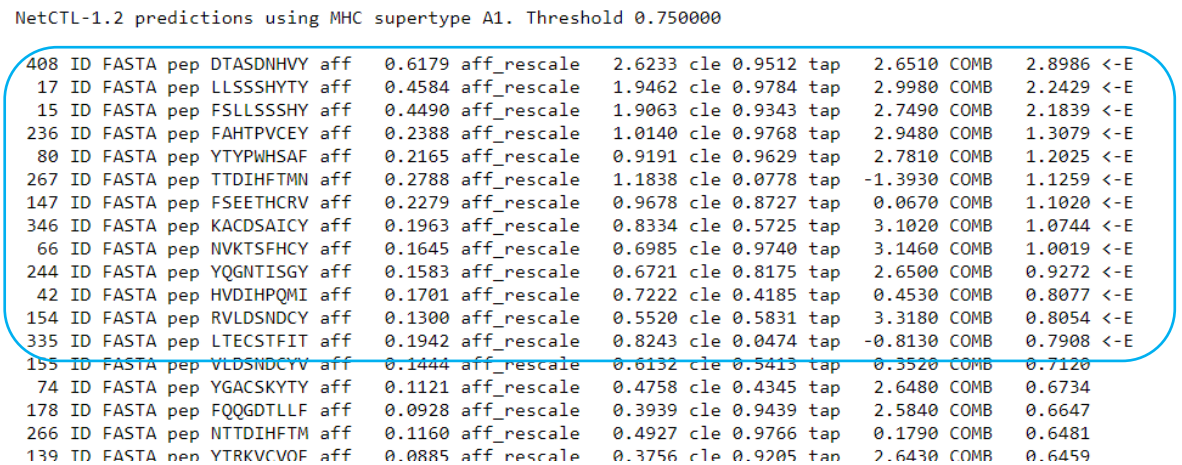
The antigenicity of the selected protein sequence was 0.6172 (Probable Antigen).

3.2 Identification of CTL epitopes

NetCTL1.2 server was used to identify the CTL epitopes where this server using MHC-I supertype A1 and here the threshold maintained was 0.75. A total of 13 CTL epitopes were found that are shown in table 01.

Table 01: CTL epitopes with combined score

CTL Epitopes	Combined Scores
DTASDNHVY	2.8986
LLSSSHYTY	2.2429
FSLLSSSHY	2.1839
FAHTPVCEY	1.3079
YTYPWHSAF	1.2025
TTDIHFTMN	1.1259
FSEETHCRV	1.102
KACDSAICY	1.0744
NVKTSFHCY	1.0019
YQGNTISGY	0.9272
HVDIHPQMI	0.8077
RVLDSNDCY	0.8054
LTECSTFIT	0.7908



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

408 ID FASTA pep DTASDNHVY aff   0.6179 aff_rescale   2.6233 cle 0.9512 tap   2.6510 COMB   2.8986 <-E
 17 ID FASTA pep LLSSSHYTY aff   0.4584 aff_rescale   1.9462 cle 0.9784 tap   2.9980 COMB   2.2429 <-E
 15 ID FASTA pep FSLLSSSHY aff   0.4490 aff_rescale   1.9063 cle 0.9343 tap   2.7490 COMB   2.1839 <-E
236 ID FASTA pep FAHTPVCEY aff   0.2388 aff_rescale   1.0140 cle 0.9768 tap   2.9480 COMB   1.3079 <-E
 80 ID FASTA pep YTYPWHSAF aff   0.2165 aff_rescale   0.9191 cle 0.9629 tap   2.7810 COMB   1.2025 <-E
267 ID FASTA pep TTDIHFTMN aff   0.2788 aff_rescale   1.1838 cle 0.0778 tap  -1.3930 COMB   1.1259 <-E
147 ID FASTA pep FSEETHCRV aff   0.2279 aff_rescale   0.9678 cle 0.8727 tap   0.0670 COMB   1.1020 <-E
346 ID FASTA pep KACDSAICY aff   0.1963 aff_rescale   0.8334 cle 0.5725 tap   3.1020 COMB   1.0744 <-E
 66 ID FASTA pep NVKTSFHCY aff   0.1645 aff_rescale   0.6985 cle 0.9740 tap   3.1460 COMB   1.0019 <-E
244 ID FASTA pep YQGNTISGY aff   0.1583 aff_rescale   0.6721 cle 0.8175 tap   2.6500 COMB   0.9272 <-E
 42 ID FASTA pep HVDIHPQMI aff   0.1701 aff_rescale   0.7222 cle 0.4185 tap   0.4530 COMB   0.8077 <-E
154 ID FASTA pep RVLDSNDCY aff   0.1300 aff_rescale   0.5520 cle 0.5831 tap   3.3180 COMB   0.8054 <-E
335 ID FASTA pep LTECSTFIT aff   0.1942 aff_rescale   0.8243 cle 0.0474 tap  -0.8130 COMB   0.7908 <-E
155 ID FASTA pep VLDSNDCYV aff   0.1444 aff_rescale   0.6132 cle 0.5413 tap   0.3520 COMB   0.7120
 74 ID FASTA pep YGACSKYTY aff   0.1121 aff_rescale   0.4758 cle 0.4345 tap   2.6480 COMB   0.6734
178 ID FASTA pep FQQGDTLLF aff   0.0928 aff_rescale   0.3939 cle 0.9439 tap   2.5840 COMB   0.6647
266 ID FASTA pep NTTDIHFTM aff   0.1160 aff_rescale   0.4927 cle 0.9766 tap   0.1790 COMB   0.6481
139 ID FASTA pep YTRKVCVQF aff   0.0885 aff_rescale   0.3756 cle 0.9205 tap   2.6430 COMB   0.6459
```

Figure 05: CTL epitopes that were found from NetCTL1.2 server (Larsen et al., 2007)

NetMHCpan-4.1 server was used to identify strong binding MHC-I alleles specific to CTL epitopes. Out of 13 CTL epitopes, 7 CTL epitopes were found that were specific to CTL epitopes that are shown in table 02.

Table 02: Strong Binding CTL Epitopes

Binding Allele	Peptide	% Rank EL	% Rank BA	Aff(nM)
HLA-A*01:01	DTASDNHVY	0.087	0.222	349.04
HLA-A*01:01	LLSSSHYTY	0.266	0.288	520.07
HLA-A*01:01	FSLLSSSHY	0.151	0.173	251.09
HLA-A*26:01	YTYPWHSAF	0.015	0.010	19.78
HLA-A*26:01	FAHTPVCEY	0.168	0.350	1013.02
HLA-A*26:01	NVKTSFHCY	0.165	0.456	1374.53
HLA-B*15:01	YQGNTISGY	0.058	0.080	18.14

For checking whether the CTL epitopes were antigen or not, VaxiJen v.2.0 server was used. After checking the antigenicity individually for each of the CTL epitopes, it was found that only one CTL epitope was antigen out of seven CTL epitopes & other six CTL epitopes were non-antigen. After checking the antigenicity, AllerTop v.2.0 server was used to check the allergenicity of the CTL epitopes where six CTL epitopes were probable non-allergen and one CTL epitope showed allergenicity. Then toxicity was checked by using ToxinPred server where one CTL epitope was toxin and other six were non-toxin. Only one CTL epitope maintained all the criteria and it was selected for vaccine construction. The antigenicity, allergenicity and toxicity of the CTL epitopes are shown in the following table 03.

Table 03: Antigenicity, Allergenicity & Toxicity of the selected CTL Epitopes

Epitopes	Antigenicity	Allergenicity	Toxicity
DTASDNHVY	Non-antigen	Probable non-allergen	Non-toxin
LLSSSHYTY	Antigen	Probable non-allergen	Non-toxin
FSLLSSSHY	Antigen	Probable allergen	Non-toxin
YTYPWHSAF	Non-antigen	Probable non-allergen	Non-toxin
FAHTPVCEY	Non-antigen	Probable non-allergen	Non-toxin
NVKTSFHCY	Antigen	Probable non-allergen	Toxin
YQGNTISGY	Non-antigen	Probable non-allergen	Non-toxin

3.3 Identification of HTL Epitopes

NetMHCpanII 4.0 server was used to identify HTL epitopes which were strong binder. 39 strong binder HTL epitopes were found after using this server. These HTL epitopes were further checked based on their IFN- γ, IL-4 & IL-10 stimulating profile that are shown in table 04. IFNepitope server was used to check IFN- γ, IL4pred server was used to check IL-4 inducer & IL10pred server was to checked IL-10 inducer. Eight HTL epitopes were maintained with all these criteria.

Table 04: HTL Epitopes with IFN- γ, IL-4 & IL-10 stimulating ability

Serial No.	HTL Epitopes	IFN-γ Induction	IL-4 Induction	IL-10 Induction
1	ADPDGLIRDHINVLL	Negative	Inducer	Inducer
2	AICYGAISVTLNRGQ	Positive	Inducer	Inducer
3	ATIDSFQSFNTTDIH	Negative	Inducer	Non-Inducer
4	CSTNGLLANAPHLDR	Negative	Inducer	Non-Inducer
5	CYGAISVTLNRGQNT	Positive	Inducer	Inducer
6	DGLIRDHINVLLNKD	Positive	Inducer	Inducer
7	DIHFTMNKLEWADPD	Positive	Inducer	Non-Inducer
9	DQSIPVHVDIHPQMI	Negative	Inducer	Non-Inducer
10	EKDFQYENNWACNPI	Positive	Inducer	Non-Inducer
11	ELDFSLLSSSHYTYR	Positive	Inducer	Inducer
12	ENPCKITVQTNQIEG	Negative	Inducer	Non-Inducer
13	FEKDFQYENNWACNP	Negative	Inducer	Non-Inducer
14	GIYLDKLKAVGVAFK	Positive	Inducer	Inducer
15	GNTISGYKKLMATID	Positive	Inducer	Non-Inducer
16	HVDIHPQMISMEVQN	Positive	Inducer	Inducer
17	ISGYKKLMATIDSFQ	Negative	Inducer	Non-Inducer
18	LDRVMGVDTASDNHV	Positive	Inducer	Non-Inducer
19	LIRDHINVLLNKDVE	Positive	Inducer	Non-Inducer
20	LMATIDSFQSFNTTD	Negative	Inducer	Non-Inducer
21	MATIDSFQSFNTTDI	Negative	Inducer	Non-Inducer
22	NPCKITVQTNQIEGA	Negative	Inducer	Non-Inducer
23	NTISGYKKLMATIDS	Positive	Inducer	Non-Inducer
24	PCKITVQTNQIEGAW	Negative	Inducer	Non-Inducer
25	PDGLIRDHINVLLNK	Positive	Inducer	Inducer
26	QGNTISGYKKLMATI	Positive	Inducer	Non-Inducer
27	QNTIHVTGKGGHSGS	Negative	Inducer	Inducer
28	QSIPVHVDIHPQMIS	Negative	Inducer	Inducer
29	RDHINVLLNKDVEFS	Negative	Inducer	Non-Inducer
30	RLINPQNKDQSIPVH	Negative	Inducer	Non-Inducer
31	RRLINPQNKDQSIPV	Negative	Inducer	Non-Inducer
32	RRRLINPQNKDQSIP	Negative	Inducer	Non-Inducer
33	SIPVHVDIHPQMISM	Negative	Inducer	Inducer
34	STNGLLANAPHLDRV	Negative	Inducer	Non-Inducer
35	TDIHFTMNKLEWADP	Negative	Inducer	Non-Inducer
36	TNGLLANAPHLDRVM	Negative	Inducer	Non-Inducer
37	YGAISVTLNRGQNTI	Positive	Inducer	Inducer
38	YQGNTISGYKKLMAT	Positive	Inducer	Non-Inducer
39	YRRRLINPQNKDQSI	Negative	Inducer	Non-Inducer

The eight HTL epitopes were further checked based on three parameters which were - antigenicity, allergenicity and toxicity (shown in table 05). Two HTL epitopes were found which satisfy these criteria. These two HTL epitopes were selected for vaccine construction.

Table 05: HTL epitopes with antigenicity, allergenicity and toxicity stimulating profile

Epitopes	Antigenicity	Allergenicity	Toxicity
AICYGAISVTLNRGQ	Antigen	Allergen	Non-toxic
CYGAISVTLNRGQNT	Antigen	Allergen	Non-toxic
DGLIRDHINVLLNKD	Non-Antigen	Allergen	Non-toxic
ELDFSLLSSSHYTYR	Antigen	Non-Allergen	Non-toxic
HVDIHPQMISMEVQN	Antigen	Allergen	Non-toxic
PDGLIRDHINVLLNK	Non-Antigen	Non-Allergen	Non-toxic
YGAISVTLNRGQNTI	Antigen	Allergen	Non-toxic
GIYLDKLKAVGVAFK	Antigen	Non-Allergen	Non-toxic

3.4 Identification of B cell Epitopes

B cell prediction tool - IEDB Analysis Resource was used to determine the B cell epitopes where different methods are available but Bepipred Linear Epitope Prediction 2.0 method was chosen as it is the most advanced method compared to the others. A score vs position graph was visualized where threshold 0.5 was maintained. Here, the green color indicates they were not the desired B cell epitopes as their threshold was less than 0.5 whereas the yellow color represents the desired B epitopes as their threshold above 0.5.

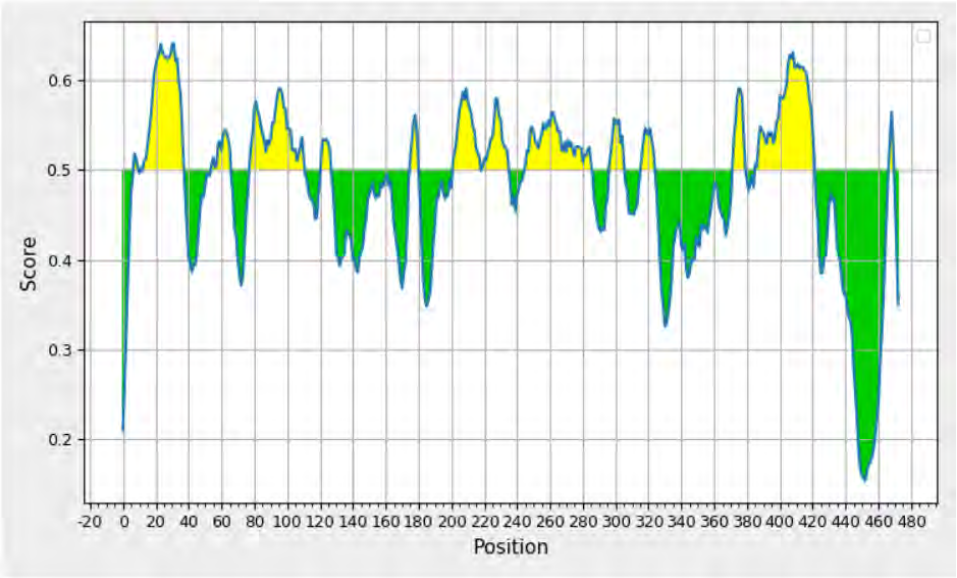


Figure 06: Score vs Position graph of B cell epitopes found from IEDB Analysis Resources (Sun et al., 2013)

By using this server, total 15 epitopes were found where 7 B cell epitopes fall within the desired threshold range (Figure 07). These B cell epitopes were further checked based on three parameters – antigenicity, allergenicity and toxicity (showed in table 06).

No.	Start	End	Peptide	Length
1	8	10	ETD	3
2	12	12	E	1
3	14	38	DFSLLSSSHYTYRRRLINPQNKDQS	25
4	55	66	QNLGHWFDAELN	12
5	78	111	SKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGV	34
6	122	127	LDKLKA	6
7	176	181	SKFQQG	6
8	202	218	TTTCQFGDPGDVMLVPP	17
9	220	235	PHDCPEYPGSFRKKCM	16
10	246	286	GNTISGYKKLMATIDSFQSFNTTDIHFTMNKLEWADPDGLI	41
11	297	306	DVEFSDLAEN	10
12	317	324	IEGAWGSG	8
13	373	380	KGGHSGSK	8
14	386	422	ETQCSTNGLLANAPHLDRVMGVDTASDNHVYDDGAPP	37
15	467	470	VRKL	4

Figure 07: B cell Epitopes found from IEDB Resource by using Bepipred Linear Epitope Prediction 2.0 Method (Sun et al., 2013)

After tested the B cell epitopes based on antigenicity, allergenicity and toxicity stimulating profile it was found that four –B cell epitopes satisfy all the criteria, and these four B cell epitopes were suitable for constructing the vaccine.

Table 06: B cell Epitopes based on antigenicity, allergenicity and toxicity parameters

B cell Epitopes	Antigenicity	Allergenicity	Toxicity
DFSLLSSSHYTYRRRLINPQNKDQS	Antigen	Non-Allergen	Non-Toxin
QNLGHWFDAELN	Non-Antigen	Non-Allergen	Non-Toxin
TTTCQFGDPGDVMLVPP	Antigen	Non-Allergen	Non-Toxin
PHDCPEYPGSFRKKCM	Non-Antigen	Allergen	Non-Toxin
DVEFSDLAEN	Antigen	Non-Allergen	Non-Toxin
IEGAWGSG	Non-Antigen	Non-Allergen	Non-Toxin
KGGHSGSK	Antigen	Non-Allergen	Non-Toxin

3.5 Construction of Vaccine

Six vaccines were constructed by using different epitope combinations where different linkers were used to prepare the vaccine. To connect adjuvant to CTL epitope, EAAK linker was used, to connect CTL to HTL, GPGPG was used and to connect HTL to B cell epitope, KK linker was used. Vaccine (I) was prepared by using 1 CTL, 2 HTL and 4 B cell epitopes. Vaccine (II) was prepared by 1CTL, 1 HTL, 3 B cell epitopes, Vaccine (III), (IV), (V) & (VI) were construed by 1 CTL, 1 HTL and 1 B cell epitopes but the combination were different for each of these vaccines. Vaccine (VI) was the final vaccine, and the sequence was generated by one CTL, one HTL and one B cell epitope.

Table 07: Construction of Vaccines using different combination of epitopes

	Construction of Vaccine
Vaccine I	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGELDFSLLSSSHYTYRGPGPGGIYLDKLKAVGVAFKKKDFSLLSSSHYTYRRRLINPQNKD QSKKTTTCQFGDPGDVMLVPPKKDVEFSDLAENKKKGGHSGSK
Vaccine II	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYIYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGELDFSLLSSSHYTYRGPGPGGIYLDKLKAVGVAFKKKTTTCQFGDPGDVMLVPPKKDVEF SDLAENKKKGGHSGSK
Vaccine III	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGGIYLDKLKAVGVAFKKKKTTTCQFGDPGDVMLVPP
Vaccine IV	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGELDFSLLSSSHYTYRKKDVEFSDLAEN
Vaccine V	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGGIYLDKLKAVGVAFKKKDVEFSDLAEN
Vaccine VI (Final Vaccine)	GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHWFDAELNVKT SFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGIYLDKLKAVGVAFKVVTIK YTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQGDTLLFLGPMEGGGLIVRQWCTTTCQF GDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTPVCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNK LEWADPDGLIRDHINVLLNKDVEFSDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAI KACDSAICYGAISVTLNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNH VYDDGAPPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAAKLLSSS HYTYGPGPGGIYLDKLKAVGVAFKKKGGHSGSK

3.6 Biochemical Analysis of Prepared Vaccines

The biochemical analysis of six constructed vaccine were shown table 08:

Table 08: Biochemical analysis of six candidate vaccines

Parameters	Vaccine I	Vaccine II	Vaccine III	Vaccine IV	Vaccine V	Vaccine VI
Antigenicity	0.6790	0.6643	0.6200	0.6461	0.6412	0.6216
Therapeutic pI	7.74	6.92	6.88	6.62	6.50	7.54
Instability Index	28.20	24.69	24.67	26.13	24.45	24.58
Gravy	-0.230	-0.171	-0.103	-0.143	-0.111	-0.127
Allergenicity	0 Allergen	0 Allergen	0 Allergen	0 Allergen	0 Allergen	0 Allergen
Toxicity	Non-toxin	Non-toxin	Non-toxin	Non-toxin	Non-toxin	Non-toxin
3D model Coverage & Confidence	100% & 68%	100% & 71%	100% & 76%	100% & 77%	100% & 77%	100% & 78%
Z-score	-6.62	-6.62	-6.62	-6.62	-6.62	-6.62
Ramachandran Region	95.99%	95.99%	95.99%	95.99%	95.99%	95.99%
Ramachandran outlier	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%

The six-candidate vaccine satisfy the acceptable criteria – antigenicity (>0.5), instability index (<40), GRAVY (negative value), allergenicity (0 allergen), toxicity (non-toxin), Z-score (plotted in the region that is acceptable), Ramachandran favored region (>90%), Ramachandran outlier (<2%). Vaccine (I) was not chosen because for vaccine (I), the confidence and coverage of the 3D modeling were 100% and 68%. But the desired range is with 100% confidence the coverage should be greater than 70%. Vaccine (III)-(IV) were also not chosen because their therapeutic pI was <7. For therapeutic pI, the acceptable range is (7-9). However, Vaccine (VI) properly satisfy all the criteria. That's why, vaccine (VI) proposed as final vaccine.

Here in figure 08, it is showed that, for vaccine (VI) which is the final vaccine, the number of amino acid was 516 and molecular weight was 57211.52Da. The therapeutic pI was 7.54. However, this Protparam server also showed the composition of the amino acids where it is visualized that the highest % of amino acid contain glycine and the % was 9.1%. The figure also represents the negatively charged & positively charged residues which were 47 & 48. This server also represents the instability index, aliphatic index, GRAVY (grand average of hydropathicity). When instability index is less than 40 it indicates the vaccine is stable. And the value for aliphatic index always greater than 60 is acceptable. GRAVY indicates hydropathicity which means whether it is hydrophobic or hydrophilic. And hydrophilic is desirable otherwise there is a chance to increase the toxicity and the absorption is not done properly. Here in figure 09, it is seen that the instability index was 24.58, aliphatic index was 77.23 and GRAVY was –0.127.

Number of amino acids: 516

Molecular weight: 57211.52

Theoretical pI: 7.54

Amino acid composition: CSV format

Ala (A)	24	4.7%
Arg (R)	15	2.9%
Asn (N)	22	4.3%
Asp (D)	28	5.4%
Cys (C)	27	5.2%
Gln (Q)	17	3.3%
Glu (E)	19	3.7%
Gly (G)	47	9.1%
His (H)	22	4.3%
Ile (I)	30	5.8%
Leu (L)	40	7.8%
Lys (K)	33	6.4%
Met (M)	9	1.7%
Phe (F)	28	5.4%
Pro (P)	22	4.3%
Ser (S)	35	6.8%
Thr (T)	37	7.2%
Trp (W)	9	1.7%
Tyr (Y)	17	3.3%
Val (V)	35	6.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 47
Total number of positively charged residues (Arg + Lys): 48

Figure 08: Amino acid number, molecular weight, composition and charged residues of prepared vaccine (Garg et al., 2016)

Atomic composition:

Carbon	C	2554
Hydrogen	H	3918
Nitrogen	N	686
Oxygen	O	739
Sulfur	S	36

Formula: $C_{2554}H_{3918}N_{686}O_{739}S_{36}$
Total number of atoms: 7933

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	76455
Abs 0.1% (=1 g/l)	1.336, assuming all pairs of Cys residues form cystines

Ext. coefficient	74830
Abs 0.1% (=1 g/l)	1.308, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is G (Gly).

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 24.58
This classifies the protein as stable.

Aliphatic index: 77.23

Grand average of hydropathicity (GRAVY): -0.127

Figure 09: Atomic composition, formula, half-life, atom number, extinction coefficient, instability index, aliphatic index & GRAVY of the prepared vaccine (Garg et al., 2016)

3.7 Toxicity & Allergenicity of the constructed vaccine

The prepared vaccine was checked for whether it was toxic or non-toxic. And that's why the constructed vaccine was inserted into the T3DB server which is a database for toxins, and it match toxicity with its database. After completing the matching, the result page showed that the prepared vaccine has no toxicity (Figure 10).

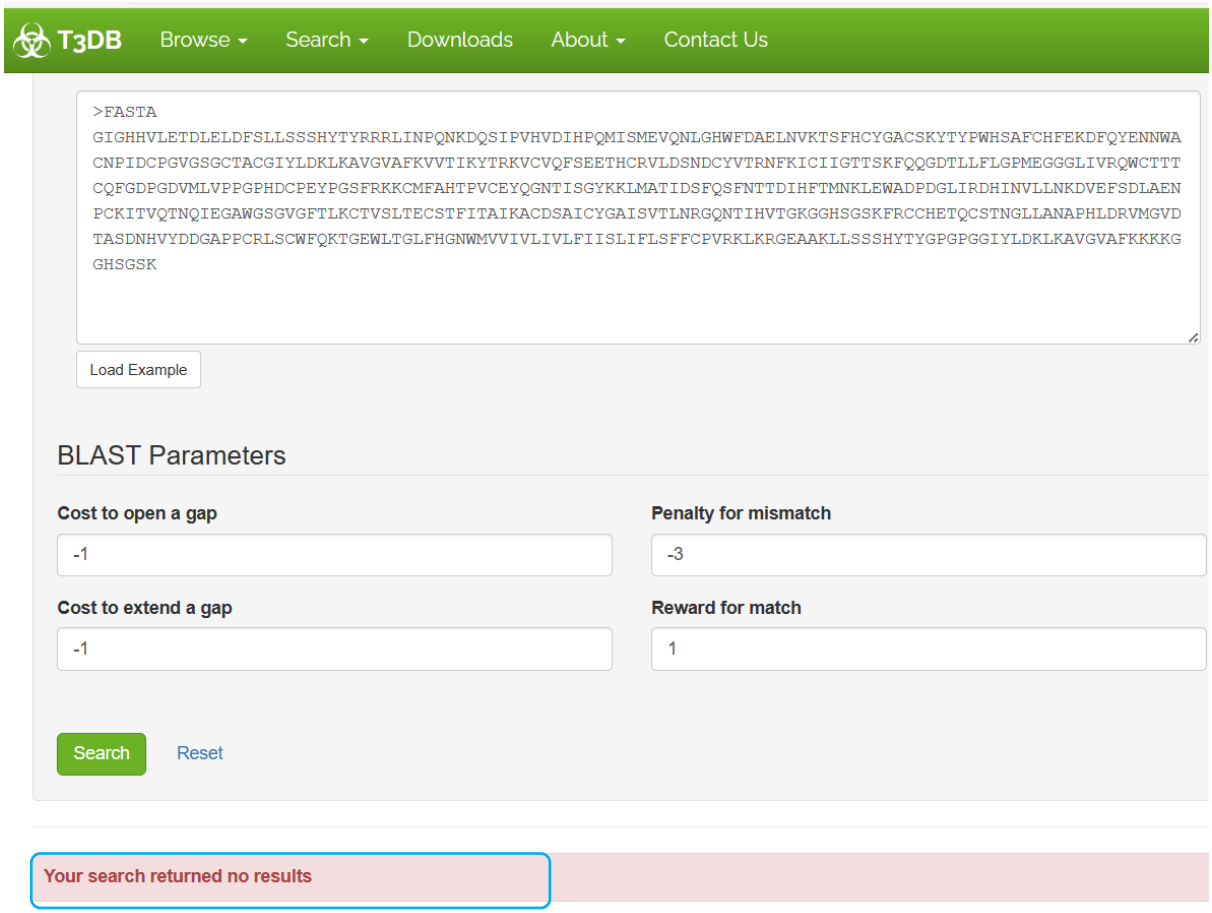


Figure 10: Non-toxin result of construed vaccine predict from T3DB server (Wishart et al., 2015)

The allergenicity of the prepared vaccine was checked by inputted the vaccine sequence into the allergen online server which is a database for allergens, and it matches the allergenicity that is store in its database and show the result. If there is any allergen present then it will how many allergen it contains. For the prepared vaccine, it was found that zero allergen (Figure 11) present in the constructed vaccine which indicated the constructed vaccine was non-allergen.

80mer Sliding Window Search Results

Database	AllergenOnline Database v22 (May 25, 2023)
Input Query	>FASTA GIGHHVLETDLELDFSLLSSSHYTYRRRLINPQNKDQSIPVHVDIHPQMISMEVQNLGHW FDAELNVKTSFHCYGACSKYTYPWHSAFCHFEKDFQYENNWACNPIDCPGVGSGCTACGI YLDKLKAVGVAFKVVTIKYTRKVCVQFSEETHCRVLDSNDCYVTRNFKICIIGTTSKFQQ GDTLLFLGPMEGGGLIVRQWCTTTCQFGDPGDVMLVPPGPHDCPEYPGSFRKKCMFAHTP VCEYQGNTISGYKKLMATIDSFQSFNTTDIHFTMNKLEWADPDGLIRDHINVLLNKDVEF SDLAENPCKITVQTNQIEGAWGSGVGFTLKCTVSLTECSTFITAIKACDSAICYGAISVT LNRGQNTIHVTGKGGHSGSKFRCCHETQCSTNGLLANAPHLDRVMGVDTASDNHVYDDGA PPCRLSCWFQKTGEWLTGLFHGNWMVVIVLIVLFIISLIFLSFFCPVRKLKRGEAAKLLS SSHYTYGPGPGGIYLDKLKAVGVAFKKKKGGHSGSK
Length	516
Number of 80 mers	437
Number of Sequences with hits	0

No Matches of Greater than 35% Identity Found

AllergenOnline Database v22 (May 25, 2023)

Figure 11: 0 allergenicity predict from allergen-online server (Goodman et al., 2016)

3.8 Homology modeling of the constructed vaccine

3D model of the constructed vaccine was found from Phyre2 server that was shown in figure 12. With 100% confidence this homology model predicted that the coverage of the vaccine model was 78% (Figure 13).

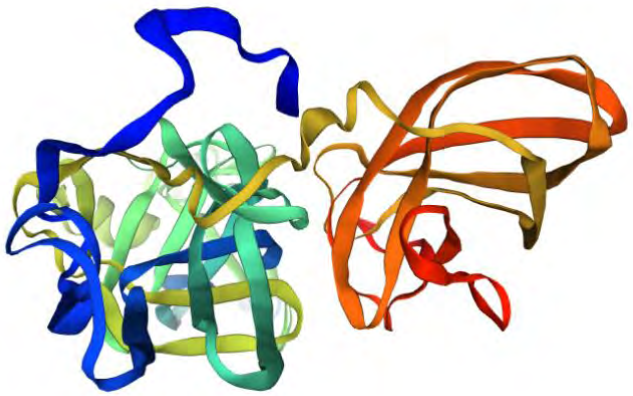


Figure 12: 3D model of the constructed vaccine from Phyre2 (Kelley et al., 2015)

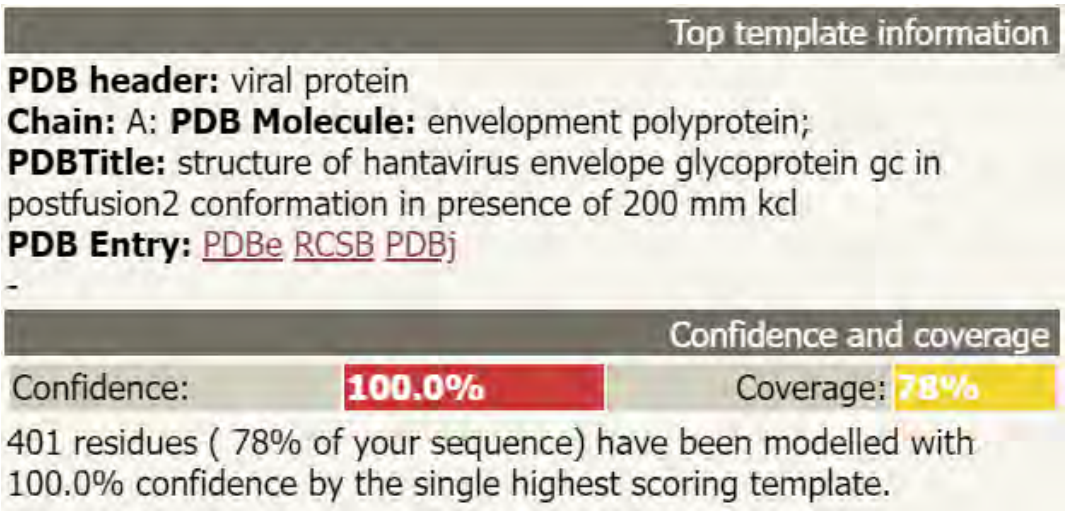


Figure 13: Score of the homology modeling found from Phyre2 (Kelley et al., 2015)

3.9 Z-score Determination

The z-score of the model was -6.62 and from the graph (Figure 14) it was seen that the structure belongs to X-Ray region. Another graph that was basically the local quality graph represents the knowledge-based energy vs sequence position of the model. Here in this graph, some residues rise peaks below the baseline which exhibits negative, and some residues rises peaks above the baseline which exhibits positive energy (Figure 15).

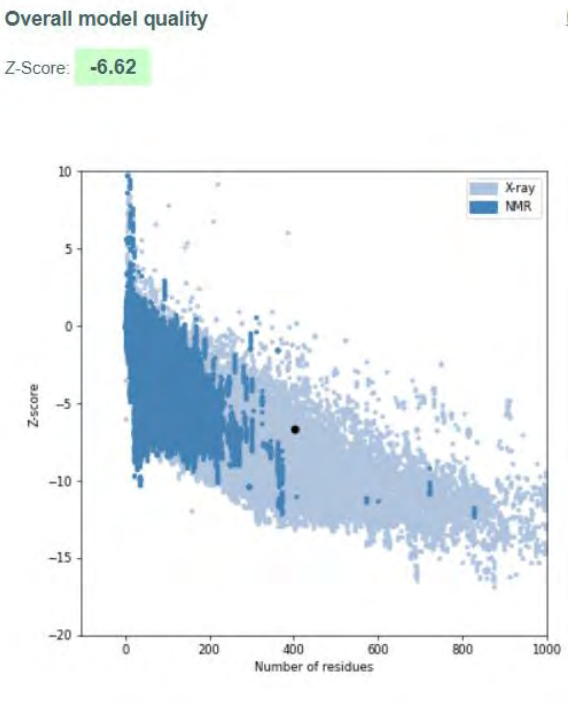


Figure 14: Overall model quality of the constructed vaccine (Wiederstein & Sippl, 2007)

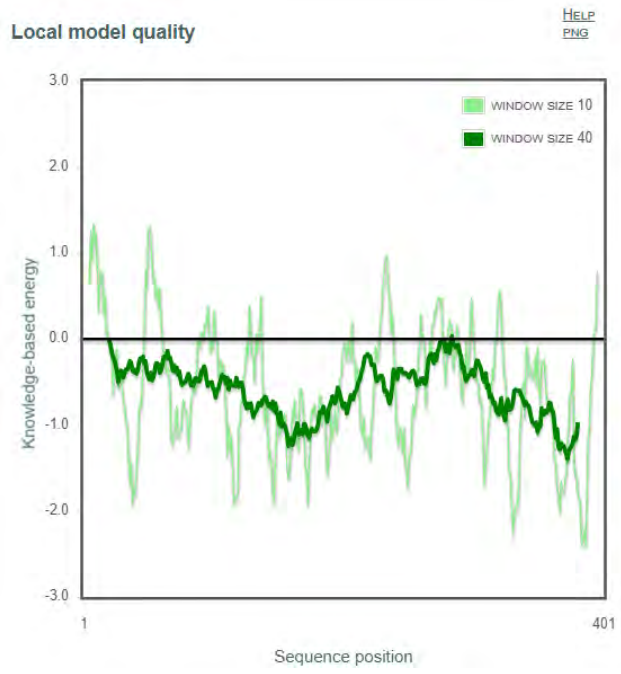


Figure 15: Local model quality represents knowledge based energy vs sequence graph (Wiederstein & Sippl, 2007)

3.10 Ramachandran Plotting

The Ramachandran plotting (Figure 16a) was checked by Swiss Model Expasy where it was found that the Ramachandran region was 95.99% and the Ramachandran outlier was 1.00% (Figure 16b). The acceptable range for Ramachandran favored region is greater than 90% and the outlier range is less than 2%. That means, the Ramachandran plotting falls within the desired range.

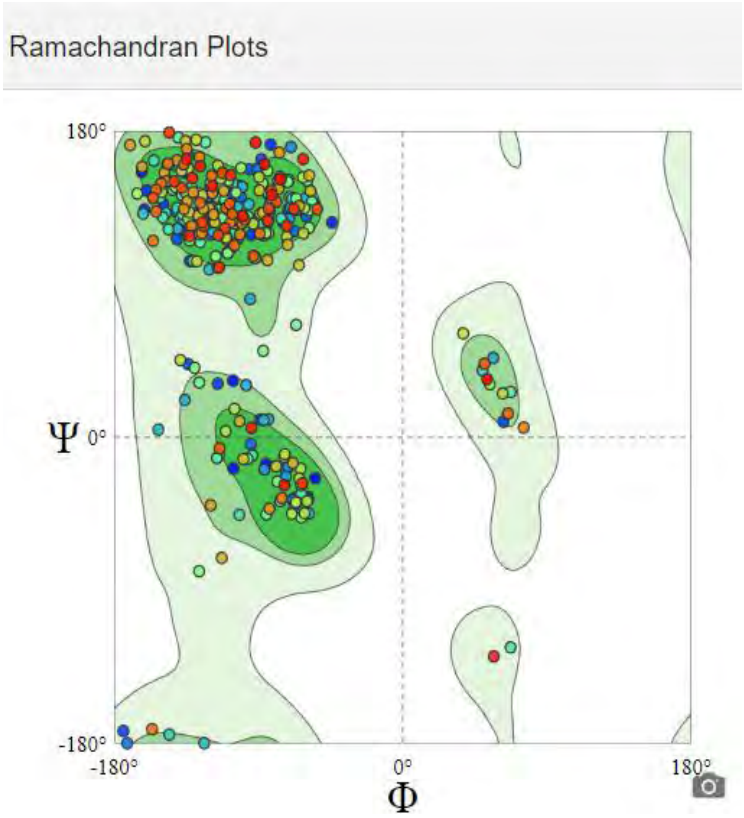


Figure 16(a): Ramchandran plots found from SWISS-MODEL tool (Schwede et al., 2003)

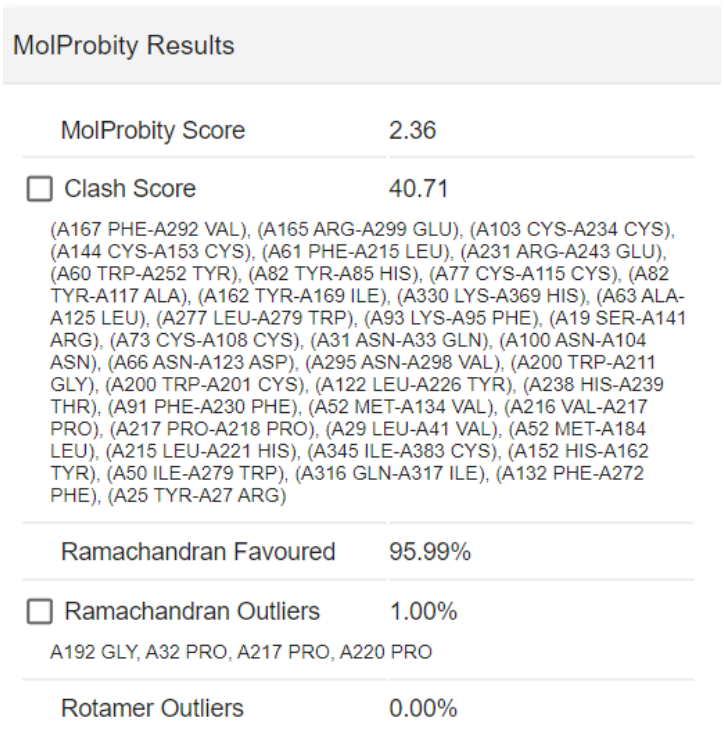


Figure 16(b): MolProbity results found from SWISS-MODEL tool (Schwede et al., 2003)

3.11 Molecular Docking with Relevant Human Receptor

Molecular docking was performed by using a server named ClusPro where human Toll Like Receptor 4 (TLR4) was chosen to bind with the ligand. The PDB file of TLR4 was inserted into the receptor folder and the PDB found from phyre2 uploaded in the ligand folder. The result gave 10 docked complexes where the first docked complex was the most stable docked complex (Figure 17). It scored the highest score which was -841.0 (Figure 18).

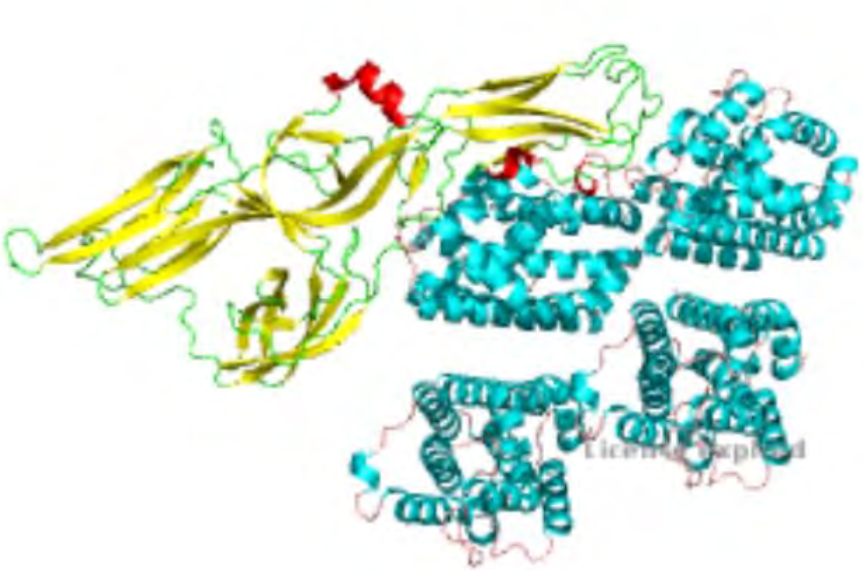


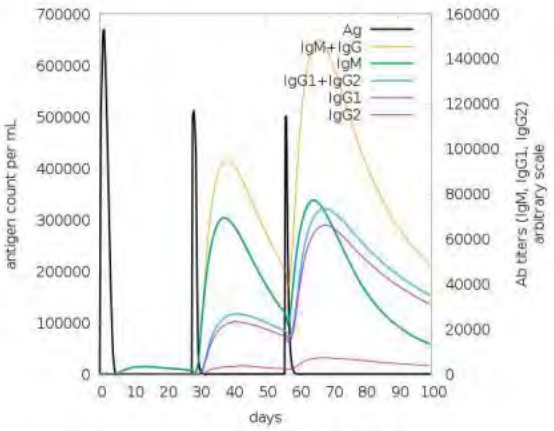
Figure 17: Molecular docking with suitable TLR-4 (Comeau et al., 2004)

Cluster	Members	Representative	Weighted Score
0	62	Center	-823.3
		Lowest Energy	-841.0
1	60	Center	-717.2
		Lowest Energy	-860.2
2	50	Center	-632.0
		Lowest Energy	-743.0
3	38	Center	-674.7
		Lowest Energy	-760.9
4	32	Center	-892.4
		Lowest Energy	-982.7

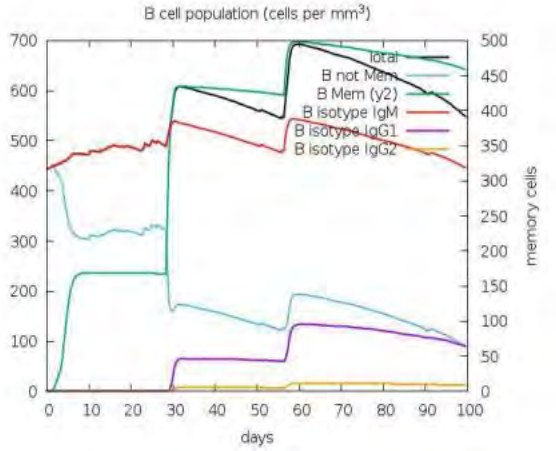
Figure 18: Highest scoring docked complex found from ClusPro (Comeau et al., 2004)

3.12 Immune Stimulation

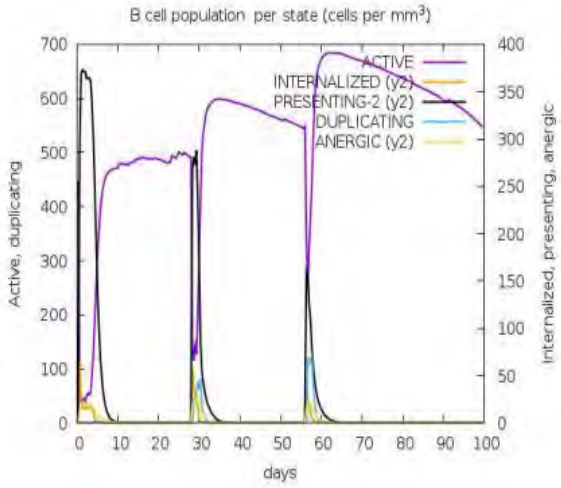
For immune stimulation the server C-IMMSIM was used where it showed normal termination and various graphical image were found (Figure 19 a-j).



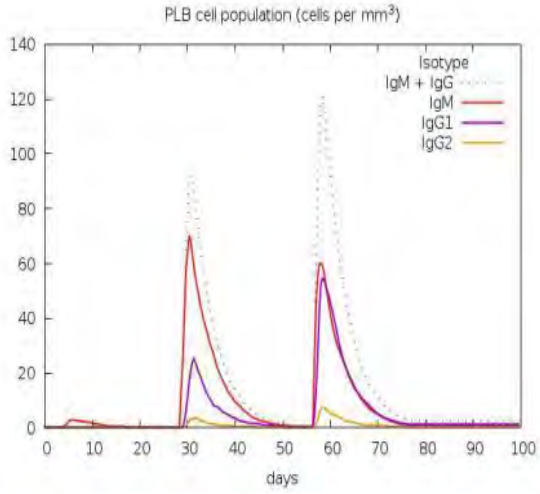
(a) Antigen count per ml & numerous antibody subtype (Rapin et al., 2010)



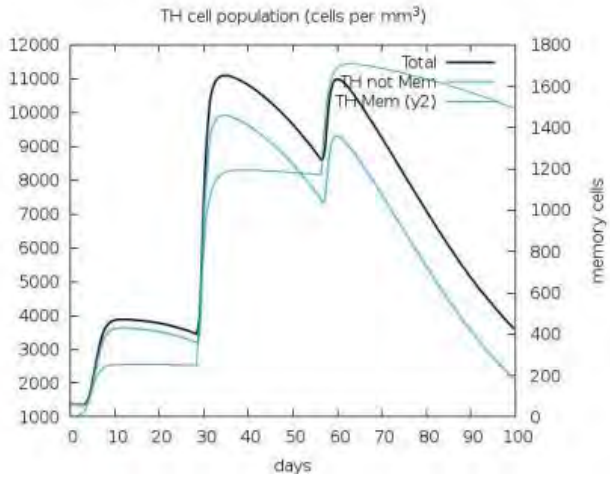
(b) Total count of B lymphocytes, memory cell and subdivided immunoglobulins (Rapin et al., 2010)



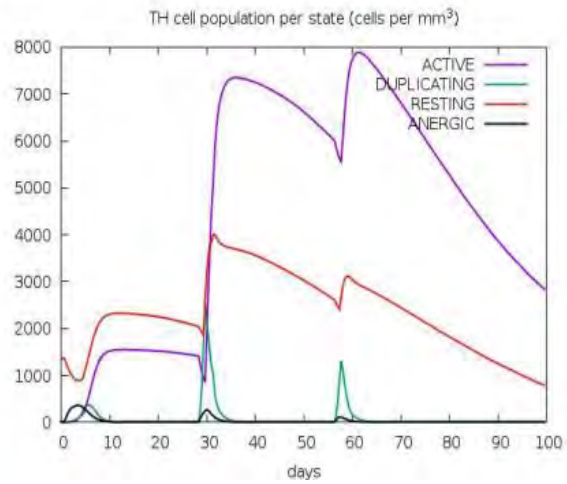
(c) Per entity state the population of B lymphocytes (Rapin et al., 2010)



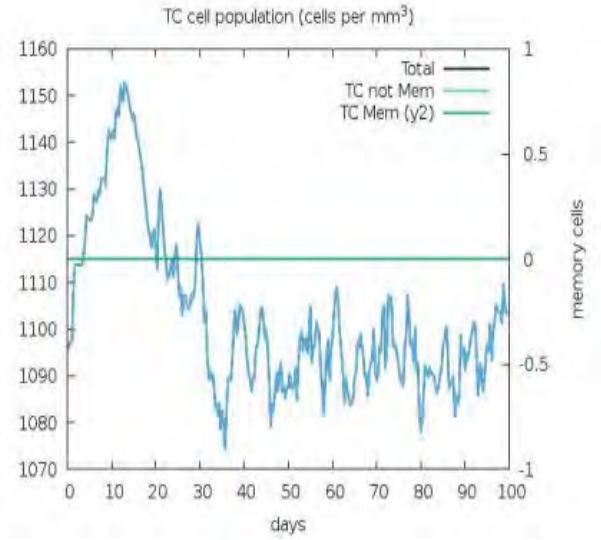
(d) Total count of plasma B lymphocytes that subdivided per isotype (Rapin et al., 2010)



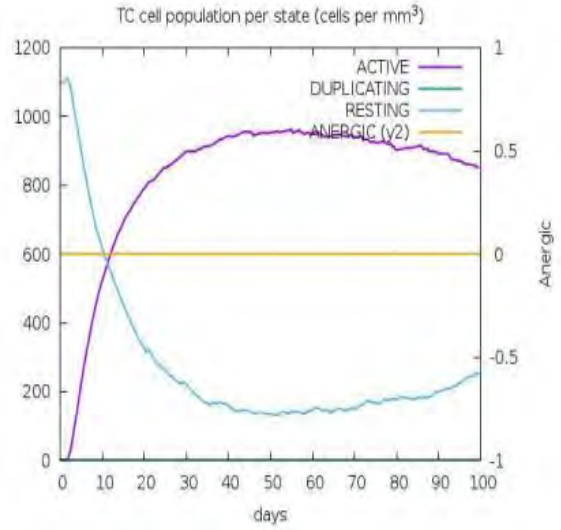
(e) Total count of CD4 T-helper lymphocytes (Rapin et al., 2010)



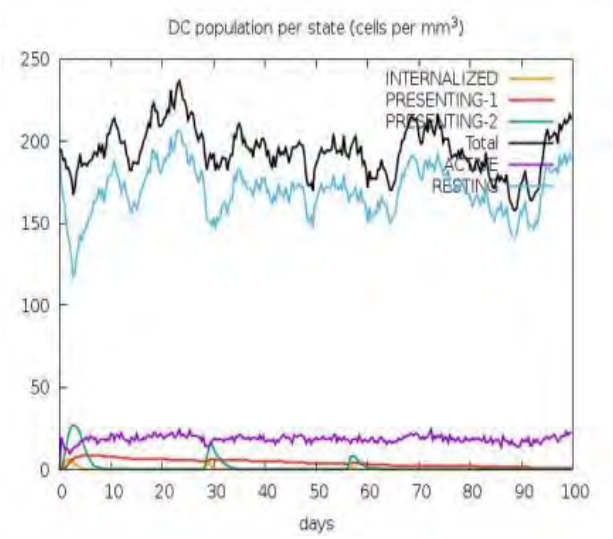
(f) population of CD4 T-helper lymphocytes per entity state (Rapin et al., 2010)



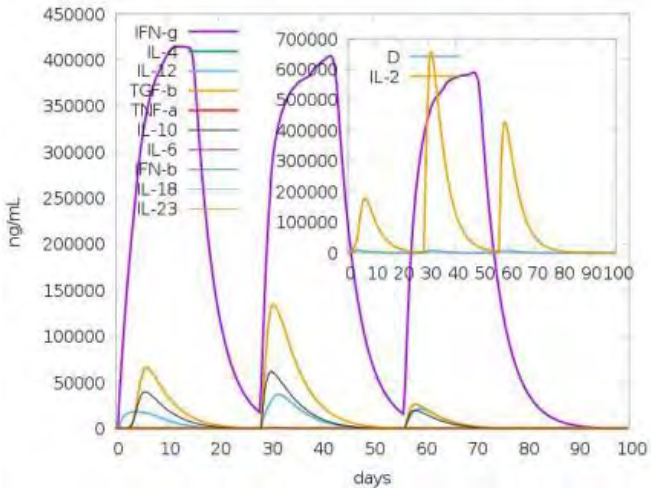
(g) Count of CD8 T-cytotoxic lymphocytes (Rapin et al., 2010)



(h) Population of B lymphocytes per entity state (Rapin et al., 2010)



(i) Number and various states of dendritic cell (Rapin et al., 2010)



(j) Level of different interleukins and cytokines (Rapin et al., 2010)

Figure 19 (a-j): Immune stimulation via C-ImmSim (Rapin et al., 2010)

In the graph (a) it was seen that the antigen level increases after the immunization. However, it also showed that after 28 days the peak raised for IgG and IgM and after 60 days the antibody level significantly increased which indicates the vaccine was able to produce the immune response. Memory cells are produced from B lymphocytes along with antibodies. These

memory cells store the memory and when the pathogen of hantavirus again attacks the immune system in the body that time it produces faster response. Here in the graph (b) total count of B lymphocytes and memory cells and their different isotypes was shown. In graph (c) it was seen that it represented the population of B lymphocytes per entity state. At the same time also represented the count of active B lymphocytes, how many B cell present on class II, which B cells were internalized the Ag and also showed how many B cells pass through duplication and anergy. The graph (d) represented the total count of plasma B lymphocytes that were subdivided per isotype. The isotypes were IgM, IgG+IgM, IgG1, IgG2.After the vaccination dose the number of CD4 T lymphocytes were represented in graph (e) where it was shown that the helper T cells were elevated on days 10, 30 and 60. And the memory cells were higher in number compared to the not-memory cells. Here in graph f) it was observed that the peaks for resting cells were higher than duplicating cells. And peaks for active cells were higher than resting cells. This graph basically represented the number of active, resting, anergic and duplicating cells of CD4 T-helper lymphocytes that were subdivided per entity state. In the graph g) the number of memory cells were at a low level whereas the number of not memory cells were observed in a good number. Active, duplicating, resting and anergic state of TC cells were observed in graph (h). Active, resting, internalized states of dendritic cells were present in the graph (i) with various numbers. Where presenting 1 target the DC that was present in MHC class-I and for presenting 2 DC were present in MHC class II. In the graph (j) it was observed that the concentration level for interleukins increased where in the insert box level for interleukin-2 was increased on day 10, between 30-50 and between 60-80 and D indicated the danger signal.

Chapter 4

Discussion

The diseases caused by hantavirus are increasing gradually in China and Europe. The targeted organ for hantavirus is lung and kidney. It is assumed that the rate may double very soon. This virus is a threat for future pandemics and now it is a global concern. That's why this is the high time to prepare a vaccine that is preventative for hantavirus. The aim of this research was to generate an in-silico multiple vaccine based on a bioinformatic approach.

For this research, M polyprotein (Glycoprotein) of hantavirus was chosen as the primary protein for constructing the vaccine. Based on antigenicity this protein was selected by using Vaxijen server. Targeted the primary sequence one CTL epitope, one HTL epitope and four B cell epitopes were found by using different servers and these epitopes were screened out based on various parameters. By using different combinations of epitopes, six candidate vaccines were constructed. The final vaccine that maintained all the criteria was prepared using one CTL, one HTL & one B cell epitope. The antigenicity of the constructed vaccine was also checked and it showed the antigenicity was better compared to the antigenicity of primary protein sequence. Protparam tool was used to detect the prepared vaccine was stable or not, where it showed it was stable. The number of amino acids was desired and the GRAVY value was negative which means the absorption will be better. For the 3D model of the structure, phyre2 was used where it gave with 100% confidence the coverage of the model was 78% which was quite satisfactory. Then the allergenicity and toxicity of the prepared vaccine was tested using Allergen Online and T3DB tools respectively. Non-allergen and non-toxic are the two major criteria of a vaccine that is called an ideal vaccine because the vaccine ideal doesn't cause any reaction or toxicity when it is injected. Then the 3D model quality of the prepared vaccine was tested using the ProsaWeb tool where it was found that the value was acceptable. The Z-score of any model indicates the model quality and energy distribution throughout the

model. Then Swiss Model Expasy was used to check the Ramachandran plotting of the model and found that they were of desirable range. After that for molecular docking human toll-like receptor 4 was chosen which is the suitable TLR for this protein of hantavirus and got 10 docked complex but choose one the docked complex based on energy because to create a chemical bond, the more the energy released, the more stable the structure and here the energy was -841.0 for stable docked complex. Finally, the stimulation result of the immune response was stimulated by C-IMMSIM and it showed positive results. And considering all the results it was analyzed that this will be a novel vaccine for hantavirus.

Chapter 5

Conclusion

To sum up, there are many cases of hantavirus all over the world but fortunately there is no case of hantavirus in the perspective of Bangladesh still now but no guarantee that this virus will never attack in Bangladesh. Still now no proper vaccine is available for this virus and many studies & researches were conducted to propose a suitable vaccine candidate. In this research different computational tools were used for the purpose of screening and checking different parameters to construct a novel vaccine against hantavirus. Immune stimulation was also checked. All the outcomes were quite promising. The purpose of this research was to design a multi epitope vaccine based on in-silico approach. One limitation of this research is that this research does not assess the safety and efficacy of the vaccine, it just constructs a vaccine model based on different parameters and criteria. To market the vaccine it is very essential to check its safety and efficacy parameters which means in-vivo and in-vitro tests are very important in that case.

References

Abdulla, F., Nain, Z., Hossain, M. M., Syed, S. Bin, Ahmed Khan, M. S., & Adhikari, U. K. (2021). A comprehensive screening of the whole proteome of hantavirus and designing a multi-epitope subunit vaccine for cross-protection against hantavirus: Structural vaccinology and immunoinformatics study. *Microbial Pathogenesis*, *150*.

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*.

Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Research*, *32*.

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*.

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2--a server for in silico prediction of allergens. *Journal of Molecular Modeling*, *20*.

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

Garg, V. K., Avashthi, H., Tiwari, A., Jain, P. A., Ramkete, P. W. R., Kayastha, A. M., & Singh, V. K. (2016). MFPPI – Multi FASTA ProtParam Interface. *Bioinformation*, *12*.

Ghafoor, D., Kousar, A., Ahmed, W., Khan, S., Ullah, Z., Ullah, N., Khan, S., Ahmed, S., Khan, Z., & Riaz, R. (2021). Computational vaccinology guided design of multi-epitopes subunit vaccine designing against Hantaan virus and its validation through immune simulations. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *93*.

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & 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*.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLOS ONE*, *8*.

Joshi, A., Ray, N. M., Singh, J., Upadhyay, A. K., & Kaushik, V. (2022). T-cell epitope-based vaccine designing against Orthohantavirus: a causative agent of deadly cardio-pulmonary disease. *Network Modeling and Analysis in Health Informatics and Bioinformatics*, *11*.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nature Protocols*, *10*.

Kupani, M., Pandey, R. K., Vashisht, S., Singh, S., Prajapati, V. K., & Mehrotra, S. (2023). Prediction of an immunogenic peptide ensemble and multi-subunit vaccine for Visceral leishmaniasis using bioinformatics approaches. *Heliyon*, *9*.

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*.

Muyangwa, M., Martynova, E. V., Khaiboullina, S. F., Morzunov, S. P., & Rizvanov, A. A. (2015). Hantaviral Proteins: Structure, Functions, and Role in Hantavirus Infection. *Frontiers in Microbiology*, *6*.

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*.

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. *PLOS ONE*, *5*.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*.

Singh, O., Hsu, W. L., & Su, E. C. Y. (2022). Ileukin10pred: A computational approach for predicting il-10-inducing immunosuppressive peptides using combinations of amino acid global features. *Biology*, *11*.

Sun, P., Ju, H., Liu, Z., Ning, Q., Zhang, J., Zhao, X., Huang, Y., Ma, Z., & Li, Y. (2013). Bioinformatics Resources and Tools for Conformational B-Cell Epitope Prediction. *Computational and Mathematical Methods in Medicine*, *2013*, 11.

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*.

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2015). T3DB: the toxic exposome database. *Nucleic Acids Research*, *43*.