In Silico Design of a Capsid Vertex Component 1 Protein Targeted Multi-Epitope Vaccine Against Cytomegalovirus

By

Md. Naimur Rahman Nabin
19346041

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
February 2024

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Md. Naimur Rahman Nabin
19346041

Approval

The thesis titled “In Silico Design of a Capsid Vertex Component 1 Protein Targeted Multi-Epitope Vaccine Against Cytomegalovirus” submitted by Md. Naimur Rahman Nabin (19346041) of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on February 2024.

Supervised by:

Supervisor:

Mohammad Kawsar Sharif Siam
Senior Lecturer, AMRSC (UK)
School of Pharmacy
BRAC University

Approved by:

Program Director:

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

Dean:

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
BRAC University

Ethics Statement

All ethical standards were maintained in the completion of this thesis. Conducting the research with integrity and following the ethical principles was given utmost importance.

Abstract

Cytomegalovirus is a virus causing irreversible neurological damage in newborns. In this study, in silico computational method was performed where, as primary protein, Capsid Vertex Component 1 Protein of cytomegalovirus from Uniprot server was selected. Using Vaxijen v2.0 it's antigenicity (0.5142) & Cytotoxic T lymphocyte,helper T lymphocyte,B cell epitopes was found via NetCTL-1.2, NetMHCIIpan 4.0, Bepipred servers. Helper T lymphocyte epitopes were Screened via IFN, IL-4Pred & IL-10Pred epitopes. For constructing the vaccine, linkers were used. A Biochemical analysis gave potential results. With ProtParam the instability index (34.10), grand average of hydropathy (-0.263) & molecular weight (56399.91 kDa) was predicted. Using AllergenOnline & T3DB, the vaccine was discerned as non-allergen & non-toxic. ProSAweb gave Z-score (-5.95) & SWISS-MODEL generated Ramachandran plots. By C-IMMSIM, desirable responses of immune cells & antibodies in accordance with three doses of the vaccine was analyzed. In vitro, in vivo methods can be done to validate the safety & efficacy of the vaccine.

Keywords: Cytomegalovirus; multi-epitope vaccine; in-silico; Capsid Vertex Component 1 protein; biochemical analysis.

Acknowledgement

I would like to express my deepest gratitude to my parents for their unwavering support and encouragement throughout my academic journey. Their efforts and belief in my abilities have been the cornerstone of my success, and I am forever grateful for their endless encouragement and understanding.

I am immensely thankful to my supervisor, Mohammad Kawsar Sharif Siam, Senior Lecturer, AMRSC (UK), School of Pharmacy, BRAC University for his guidance, expertise, and mentorship throughout the completion of this thesis. His insightful feedback, patience, and encouragement have been instrumental in shaping this work and pushing me to strive for excellence.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CMV	Cytomegalovirus
CVC1	Capsid Vertex Component 1

Chapter 1

Introduction

Cytomegalovirus (CMV) is a common virus which is a ubiquitous human herpes virus and with a seropositivity of > 60% in adults worldwide that is belonging to the family Herpesviridae. By the CMV there can happen infection and that is asymptomatic. In the developing fetus, CMV causes birth defects and by crossing the barrier there it could be infecting the placenta. There are tissues and organs for example endothelial cells, leukocytes, fibroblasts etc. which can be infected by CMV (Chauhan & Singh, 2020). The United States National Academy of Medicine has declared that for the public health there should be the development of the CMV vaccine as a high priority. Cytomegalovirus is happening in basically 3 forms which there are including no infection, primary which are seropositive and latent infections which there are evolving strains. There are some immunocompromised patients which they have uncontrolled high viral loads seen in the urine CMV infections are more severe (Akhtar et al., 2021). There are many developed countries in the world where CMV is causing the perinatal viral infection. In the United States, CMV is infecting 1% of the all new born which is 40000 infants per year. By CMV there is a causing of morbidity in the immunocompromised patients and there is including colitis, neuropathy, pneumonia etc. From the all-body fluids CMV can occur which is there including saliva, tears, urine etc. During the time of pregnancy CMV can be transmitted maternally. In the child care setting CMV can transmitted very easily. There is a high risk for bacterial, fungal and also viral superinfections which is increasing by the CMV disease. Right now, there is no vaccines for CMV. So, for controlling the large-scale outbreaks of CMV disease in future there is a need of effective vaccine against this disease (Plosa et al., 2012).

1.1 Structural and Genomic Features of Cytomegalovirus

Among the herpes viruses CMV has the large genome. Based on the biological properties these herpesviruses have been divided into three subfamilies. There is already identify of 100 herpesviruses. There are many proteins contain in CMV virions. There is a containing of simple set of proteins in the capsids of CMV (Spaete et al., 1994). In the herpesvirus virions there are a linear double-stranded DNA genome which is basically enclosed in an icosahedral which is T = 16 capsid. Herpes viruses have capsids which have a diameter of between 115 and 130 nm, and there while the whole virion has a diameter of between 150 and 200 nm. The Mammal-infecting herpesviruses are identified in every subfamily (Murphy & Shenk, n.d.). Right now there has also been genetic applications of chemical mutagenesis for CMV. There are Coupling groups were classified as a result of early CMV research (Ruzsics & Koszinowski, n.d.). The CMV genome is made up of a linear, double-stranded DNA molecule which is 50% bigger than the HSV-1 genome and that is also the biggest of all the human herpesviruses (236 kbp in the wild type virus) (Gibson et al., 1993). The viral proteins are basically synthesized into three categories which are immediate-early, early, and late. These categories are determined by a series of transcriptional events which is govern the expression of the CMV genome. By following a primary infection, reinfection, or reactivation there the clinical symptoms might be observed. Within six months of their mothers' transmission through the placenta, during childbirth, or through nursing, 10% of newborns had contracted the infection. There are 10% of babies who had acquired the virus within six months of the infection being passed on to them by their mothers through the placenta or during childbirth, or also through nursing. The CMV genome is boasting a high (G + C) concentration and for that reason it is the largest among all herpesviruses (Landolfo et al., 2003).

1.2 Cytomegalovirus Pathogenies and Viral Replication

AIDS patients the allograft recipients, and the fetuses' have been shown to be afflicted by cytomegalovirus (CMV). It has been identifying as a pathogen more recently for older people the general public, people in critical care units. There are all organs which might become infected with CMV through the bloodstream, and overt illness cannot occur unless the viral load reach to the extremely high levels. Moreover, a strong immune response can stop this from happen and there could be leaving the infected person largely asymptomatic. At the time of delivery there is about 2% of seronegative women and they have begun their pregnancies have seroconverted. Basically, for these mothers, small children, particularly toddlers they are the main source of CMV because of their urine and saliva has high CMV concentration. The pregnant women who have a primary infection there the 32 percent pass the virus across the placenta and after that there develop an intrauterine infection (Griffiths et al., 2015). It may be difficult to figure out cytomegaloviruses' coding potential because of their having large genomic size and that frequent usage of alternate transcription start sites. There is a number of the first studies on CMV transcriptomics which employed specialized DNA chips with probes targeted at the 150–200 open reading frames (ORFs) and that is identified by genome research. The theory about the primary source of infection in seropositive people is the virus's reactivation from the latency. But the majority of CMV infections when there is immunosuppressed individuals undergo immunosuppressive therapy after receive a solid organ, bone marrow transplant, or also when the patient has AIDS. Then the virus will be known as a significant pathogen, also when its symptoms are becoming silent (Landais & Nelson, 2013)

Chapter 2

Methodology

Figure 1 depicts a few approaches that have been built in response to the implementation of this in-silico CMV vaccine.

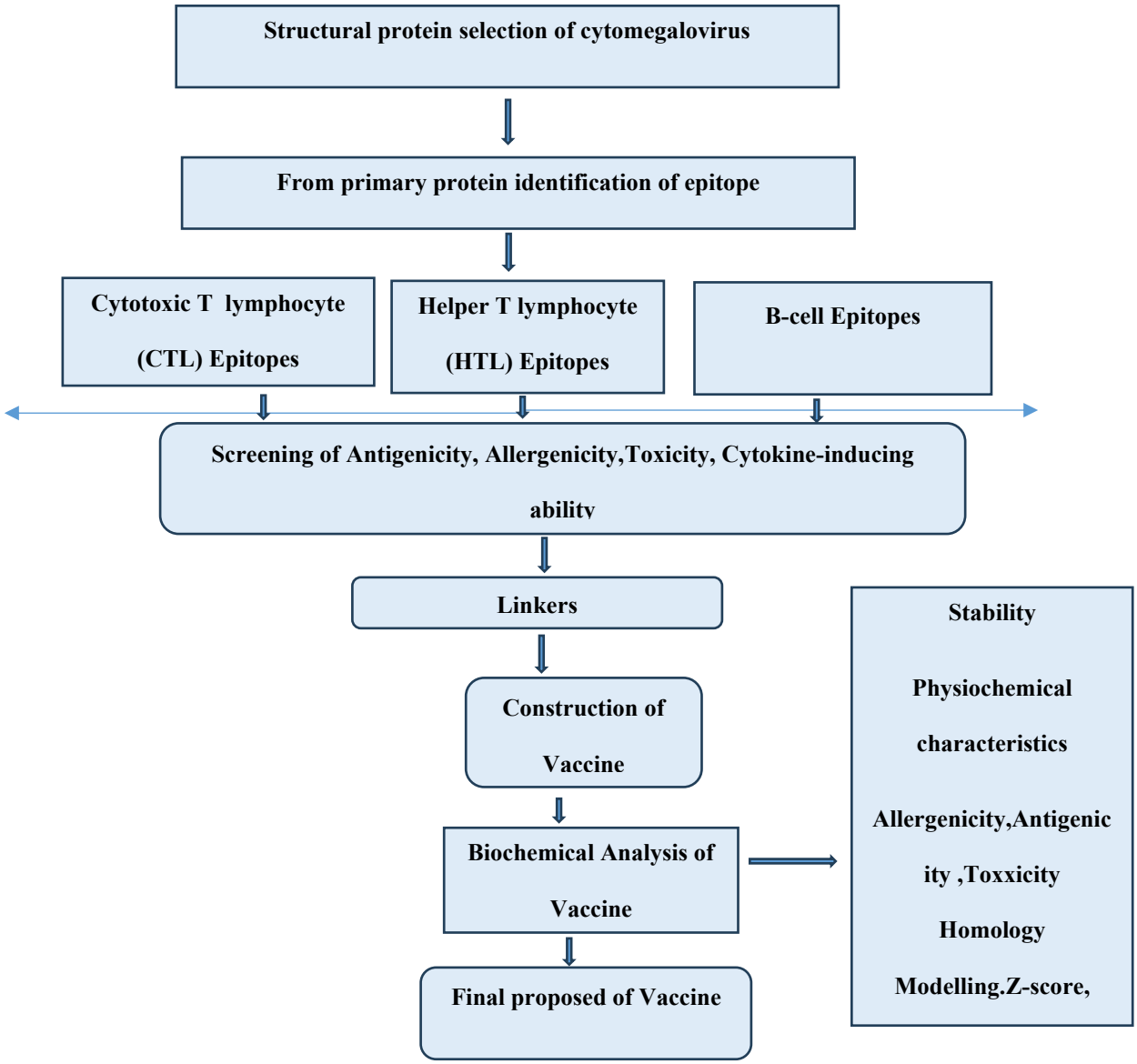


Figure 1 : Cytomegalovirus is serially cured employing in-silico vaccine design techniques

2.1 Acquiring and Gathering The Sequence of the Cytomegalovirus Envelope Protein

There is a use of Uniprot server (https://www.uniport.org) for collecting the different types of protein sequence of Cytomegalovirus strains. There are almost 19645 protein sequence have been finding out by searching results of Cytomegalovirus. The size of UniProt is basically an important collection which is containing protein sequences and their annotations that has been doubled to 80 million sequences in the last year. The UniProt number for accession space has been extended from six to ten characters for that reason as a result of this expansion in sequences (Bateman et al., 2015). In the Uniprot server there are listed lots of virus families with protein sequence. By applying the entire country column as a reference, the CMV strain was detected. Based on Fasta format the protein sequence is given and then basically compiled. There is a research paper about designing a targeted based on multi-epitope-based vaccine which is against the disease cytomegalovirus and there is following immunoinformatic method. There is using of the appropriate linkers for the epitopes that have been chosen and that had been combined for design a multi-epitope-based vaccination constructions (Bakkari, 2023). There is a use of Vaxijen 2.0 for predicting the antigenicity of the collecting protein sequences (Saha & L S Prasad, n.d.). That virus was fundamentally recognised as an attacking organism in the server. Also, they established 0.5 as their threshold for this. The projected date of complete antigenicity was noted. Following that, the protein containing the highest antigenicity was observed for matching.

2.2 Discovering An Epitopes of Cytotoxic T Lymphocyte (CTL)

For picking primary protein sequences, the NetCTL-1.2 mechanism has been employed for recognizing CTL (cytotoxic T lymphocyte) epitopes. A prediction featuring the twelve major histocompatibility complex (MHC) class 1 supertypes is made in this tool. The important step

of creating the perfect peptide is determining of the cytotoxic T lymphocyte (CTL) epitope. For each of the selected protein which there are the CTL epitope is predict by using this server. By this server there can utilizing artificial neural networks so that there can predicting MHC class I (Bhattacharjee et al., 2023). In FASTA format the primary protein sequence which was being selected that is given. For epitope identification, a threshold of 0.75 was chosen and for default values basically there is yielding the best prediction performance there are the weights of C terminal cleavage which there is the TAP transport efficiency at their default value they were left. There is a combined score which was select and there the "sort by score" option was chosen, for that reason the CTL epitopes are now listed on the result page in descending order in a prediction scores. Determination of the T-cell epitopes and also the MHC-binding peptides is important because for developing vaccines (Lin et al., 2008).

2.3 Getting The MHC Class I Alleles

By using of the NetMHCpan-4.1 server, MHC I alleles is being unique to the CTL epitopes which were predicted. There is an use of Artificial neural networks (ANNs) which are used by this server for predict peptide interaction with MHC I molecules. 9-mer peptides were provided as the input peptide length. For that reason, Strong and weak binder thresholds were set at 0.5 and 2, respectively. All of the MHC molecules there was displayed were chosen from the "select allele(s)" list. The server was configured to including binding affinity (BA) and that is ranked by prediction scores. Based on the results page, there is the corresponding peptides and alleles were recorded. Only peptides containing an excellent binding level have been included in the screening procedure (Reynisson et al., 2021).

2.4 Establishing The Epitopes of Helper T Lymphocyte (HTL)

The NetMHCIIpan 4.0 server was employed to recognise the epitopes of helper T lymphocytes (HTLs). Through the programme of artificial neural networks, this website was capable to

figure out the sequence of peptide binding that has previously been viewed for the MHC II proteins. The preferred primary protein sequence that was offered in the FASTA format was provided based on the server. The threshold for a strong binder whoever matched the requirement was 1, though the threshold for a weak binder was 5. From the "select allele(s)" there is listing of, a maximum of 20 alleles which could be chosen for each submission. There is the peptides with the strongest binding (SB) levels which were the only ones gathered. Every allele was chosen and the procedure was repeated. There in a Microsoft Excel document which bis contained a compilation of all strong binding HTLs. The sheet was stripping of the SB HTL sequences that is being repeated (Reynisson et al., 2020).

2.5 Strong Binding HTL Epitopes' Capacity To Stimulate Cytokine Production

Helper T cells can produce a variety of cytokines which stimulate our body's immune cells, for example interleukin 4, interleukin 10, and interferon-gamma. So that after the inflammatory reactions, these cytokines endure and shield tissue. For that reason, it is therefore preferred that HTL epitopes induce this type of cytokine release while developing vaccines. For that reason, the first three criteria that were examined were all of the HTL epitopes that were founding in the preceding step. Whether or not the HTLs were interferon-gamma inducing was predicted by using the IFN server. There the IFN gamma vs. non-IFN gamma model was chosen in this server. A major obstacle in the in-silico creation of subunit vaccines has been locating MHC class II binders that can produce interferon-gamma and meaning activate T-helper cells. One Th1 cytokine that helps eliminate intracellular infections is IFN gamma. Therefore, it can also stimulate cell-mediated immunity (Dhanda et al., 2013).

Afterwards, it turned out that the HTL epitopes, about the application of the IL-4pred server, triggered interleukin-4 (IL-4). There the server is selected the hybrid (SVM+motif) model.

HTL epitopes' IL-4 inducibility is taken into because IL-4 is important for the synthesis of IgE, and then the switching of antibody isotypes, and the growth of antigen-presenting cells. For determine whether or not the HTL epitopes induced interleukin-10 (IL-10) production there the IL-10Pred server was used (Nagpal et al., 2017).

2.6 The Analysis of B cell epitopes:

An online computational tool titled the IEDB Analysis Resource is further utilized during B cell epitope identifications. Hence, the "bepipred linear epitope prediction 2.0" method was picked by the server. In this server there is creating a table of predicted peptides which is the primary protein sequence is submitted. Because of the matching graph and B

2.7 Vaccine Candidate Assembly

By the usage of proper linkers, a multi-epitope vaccine incorporating CTL, HTL & B cell epitopes was successfully created via interconnecting the primary protein sequence. For a consequence, in order to create the initial vaccine construct, every single one of the B cells, HTL, and CTL epitopes that had been validated during earlier stages had been combined. Afterwards, a number of combinations of CTL, HTL & B cell epitopes were essentially created and analyzed by biochemical methods to contrast and compare diverse vaccine design modifications. Therefore**,** there is effectively isolating epitopes and enhancing their presentation in host cells require linkers (Narula et al., 2018). There the vaccination design made by the use of one of four peptide linker types: KK, AAY, GPGPG, and EAAAK. Therefore, the initial CTL epitope and the main protein sequence were joined by the EAAAK linker. There is a closest CTL, HTL & also B cell epitopes which there had been distinguished by AAY, GPGPG, and KK linkers.

2.8 In silico Biochemical Evaluation of Vaccine Candidates

For every vaccination construct there is an in-silico biochemical study was done. Therefore, examining the vaccine construct's stability there is using the ProtParam tool which was the first stage in the biochemical analysis process. The ProtParam server after that shows the atomic composition, a total number of the amino acid. There is also showing of grand average of hydrophilicity (GRAVY), theoretical pI, molecular weight & there is other information for the vaccine construct sequence upon submission. For that reason, it is a great tool for comparing different aspects of the vaccination. The vaccines that have been generated and whose reliability was previously discovered as being determined by the instability index (Walker et al., n.d.).

2.9 Evaluation of Allergenicity and Toxicity

Employing the Toxin & Toxin Target Database, or T3DB for short enough that features both a database of toxin & a database of targets, the intended vaccinations have been assessed to discover what substances were confirmed to be toxic or non-toxic. The vaccination sequences were then provided by using the FASTA format, and the server's BLAST parameters were left at their default defaults (Lim et al., 2009). For estimating the allergenicity of the created vaccinations there is the AllergenOnline service (version 21) which was consulted. By the server there compares protein sequences which have been submitted and known allergens using bioinformatics techniques such as FASTA or BLASTP local alignment tools. Based on the server's search menu, "Show Z-score" was chosen (Goodman et al., 2016).

2.10 Modeling Homology of Three-Dimensional Model Builds

For showing the created vaccines there is three-dimensional (3D) structures, the Phyre2 server which was used to operate. By using an approach named remote homology identification, the Phyre 2 server has the capability to construct 3D models in besides pinpointing the precise

position of ligand binding sites. As a result of this, the secondary and tertiary structures of the model have been acknowledged and its quality—which incorporates coverage and confidence—is revealed (Kelley et al., 2015).

"Normal modelling" constitutes one of the options than were selected in Phyre 2. This server sent each vaccination with the PDB file on an authorised email. Then the PDB files were opened and the three-dimensional structures were interactively was seen by the using of the Discovery Studio Visualizer software.

2.11 Building Z-score Reviews of Vaccines and Ramachandran Plots

Ramachandran plots and related data were gathered for the built vaccines in the SWISS-MODEL server by employing one site, which is a structure assessment site (Guex et al., 2009). The PDB file that had been uploaded in the earlier phase was necessary. There is a protein model's stereochemical quality which can be evaluated by using the Ramachandran plot. The conformation angles, phi (φ) & psi (ψ), which are collectively referred to as Ramachandran angles and are depicted against one another, will be illustrated by the Ramachandran plot. By these angles there will affect a certain protein's polypeptide chain backbone (Gopalakrishnan et al., 2007). Next, the z-score for the created vaccinations was ascertained using the ProSA-web website.

The server received the PDB file. The z-score value showed up on the outcomes page. Also, an information-driven energy vs. sequence position graph and a z-score vs. number of residues graph were exhibited (Sippl, 1993).

2.12 Docking interaction Vaccines with the Correct Human Receptors Molecularly

By the PATCHDOCK molecular-docking algorithm that was used to guarantee binding affinity between the desired receptors and the produced vaccine (Schneidman-Duhovny et al., 2005).

5y3m was inserted to the server Patchdock as the receptor molecule, and the resulting vaccine was uploaded as the ligand. There is an acronym for the human Toll-like receptor 9 (TLR9) which is 5y3m. Then the PDB file for the top-scoring solution was retrieved from the results page.

2.13 The Simulations of Immune Response

The humoral & cell-mediated reactions of an immune system to antigens are replicated through the C-ImmSim model. By this server's immune simulation produces multiple graphs for antigen, immunoglobulin, lymphocyte, and other immune cell numbers which can be predicted in relation for period of time after immunization shots. There are the three doses were added, and the vaccination sequence was entered into the C-ImmSim server. Then there is the temporal steps corresponding for injection numbers 1, 2, and 3 were 1, 84, and 168. By this there is particular hours of time are represented by each time step. Then it is the recommended dosage for the immunization was three shots spaced 28 days apart.

2.14 Observations About the Methods

The methods which are involved in creating the cytomegalovirus vaccine and the biochemical analysis that followed were carried out by using an in-silico method in this study. For that reason, there is the instruments employed in the research which were internet-based tools that are frequently utilized in research on in-silico vaccine creation. In addition to these web resources, there is a thorough literature evaluation was conducted whenever necessary to substantiate the techniques applied in this investigation.

Chapter 3

Results

3.1 The Picked Protein's Antigenic Qualities:

During protein screening, which generally equivalent to primary protein screening, the Capsid vertex component 1 protein for the cytomegalovirus has been found to be the one with the highest antigenicity in the VaxiJen v2.0 server. The strain, P24441, received recognition from The GenBank Sequence Accession. The protein's sequence can be seen below:

METHLYYDTLYQYQGGVYPAHICLPTDVCLPMRVDCIESLYFRCVFFKSGMHYTEW
SKLKFTVISREIKFKDVLKDADSDEVFTGLVVMTIPIPIVDFHFDIDSVILKLVYPRLVH
REIVLRLYDLICVRPPSNRPSEASAKNIANDFYQLTSRENKQTPDEEKRCLFFQQGPLE
PPSTVRGLKAPGNEKPIQFPAHANEKMTESFLSDSWFGQKVRCKKILDFTQTYQVVV
CWYELSFSREMQIENNLLSASQLKRVNAADFWDRTNRYLRDIGSRVLTHIVKTLQIH
NRQFKQKFNCNFPDNFSFDRLLSFMQLGKDFWILNLTLDSCIIKAIICFLGFQNGGKSF
LAQDEVWGDLIDCSKGSVIYGEKIQWILDSTNNLYSTCREKQNKSWELYVDCCALY
VSEKLELDFVLPGGFAITGKFALTDGDIDFFNWRFGLS

In accordance with Vaxijenv2.0, this sequence's antigenicity was predicted as being 0.5142 (Probable ANTIGEN), which is demonstrated in Figure 2.

Figure 2 : Antigenicity of the indicated Capsid vertex component 1 protein of Cytomegalovirus

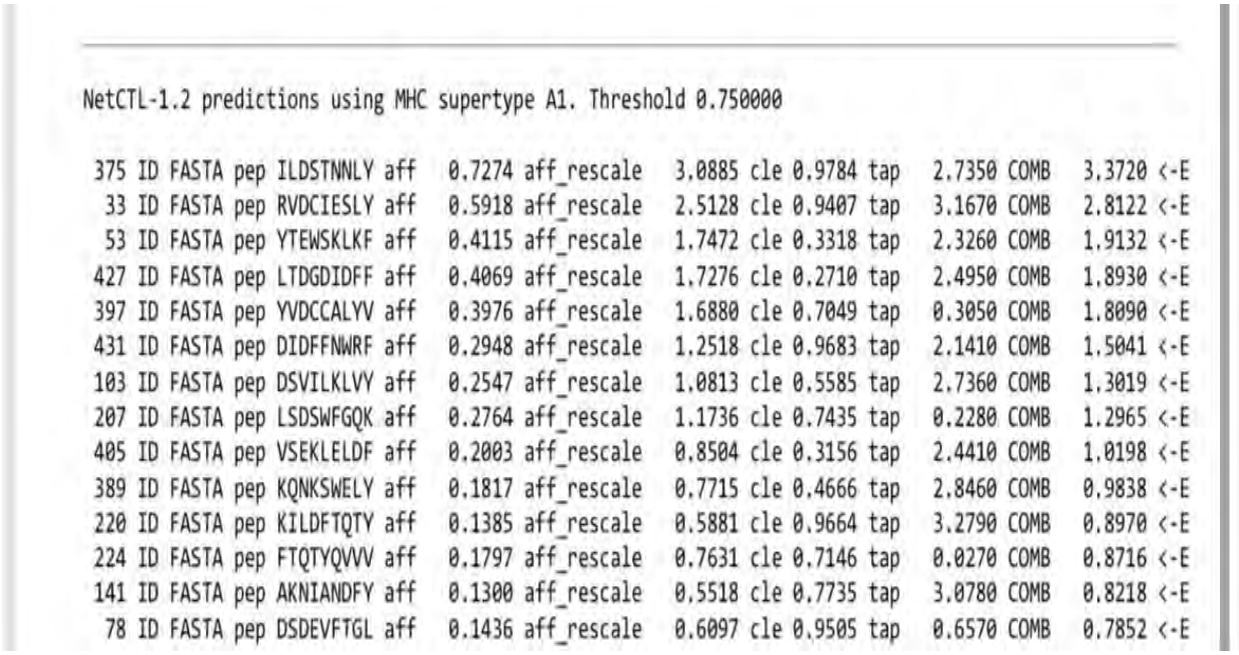
3.2 Finding the CTL Epitopes

Figure 3 demonstrates the identification of 14 epitopes from the server address NetCTL-1.2 discovered in the most significant protein sequence assuming MHC supertype A1 & the 0.75 threshold.

Table 1 : CTL epitopes as well as the epitopes' total scores

CTL epitope	Combined Score
ILDSTNNLY	3.3720
RVDCIESLY	2.8122
YTEWSKLKF	1.9132
LTDGDIDFF	1.8930
YVDCCALYV	1.8090
DIDFFNWRF	1.5041
DSVILKLVY	1.3019
LSDSWFGQK	1.2965
VSEKLELDF	1.0198
KQNKSWELY	0.9838

KILDFTQTY	0.8970
FTQTYQVVV	0.8716
AKNIANDFY	0.8218
DSDEVFTGL	0.7852



NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

375	ID	FASTA	pep	ILDSTNNLY	aff	0.7274	aff_rescale	3.0885	cle 0.9784	tap	2.7350 COMB 3.3720 <-E
33	ID	FASTA	pep	RVDCIESLY	aff	0.5918	aff_rescale	2.5128	cle 0.9407	tap	3.1670 COMB 2.8122 <-E
53	ID	FASTA	pep	YTEWSKLKF	aff	0.4115	aff_rescale	1.7472	cle 0.3318	tap	2.3260 COMB 1.9132 <-E
427	ID	FASTA	pep	LTDGDIDFF	aff	0.4069	aff_rescale	1.7276	cle 0.2710	tap	2.4950 COMB 1.8930 <-E
397	ID	FASTA	pep	YVDCCALYV	aff	0.3976	aff_rescale	1.6880	cle 0.7049	tap	0.3050 COMB 1.8090 <-E
431	ID	FASTA	pep	DIDFFNWRF	aff	0.2948	aff_rescale	1.2518	cle 0.9683	tap	2.1410 COMB 1.5041 <-E
103	ID	FASTA	pep	DSVILKLVY	aff	0.2547	aff_rescale	1.0813	cle 0.5585	tap	2.7360 COMB 1.3019 <-E
207	ID	FASTA	pep	LSDSWFGQK	aff	0.2764	aff_rescale	1.1736	cle 0.7435	tap	0.2280 COMB 1.2965 <-E
405	ID	FASTA	pep	VSEKLELDF	aff	0.2003	aff_rescale	0.8504	cle 0.3156	tap	2.4410 COMB 1.0198 <-E
389	ID	FASTA	pep	KQNKSWELY	aff	0.1817	aff_rescale	0.7715	cle 0.4666	tap	2.8460 COMB 0.9838 <-E
220	ID	FASTA	pep	KILDFTQTY	aff	0.1385	aff_rescale	0.5881	cle 0.9664	tap	3.2790 COMB 0.8970 <-E
224	ID	FASTA	pep	FTQTYQVVV	aff	0.1797	aff_rescale	0.7631	cle 0.7146	tap	0.0270 COMB 0.8716 <-E
141	ID	FASTA	pep	AKNIANDFY	aff	0.1300	aff_rescale	0.5518	cle 0.7735	tap	3.0780 COMB 0.8218 <-E
78	ID	FASTA	pep	DSDEVFTGL	aff	0.1436	aff_rescale	0.6097	cle 0.9505	tap	0.6570 COMB 0.7852 <-E

Figure 3 : CTL epitope identification employing the NetCTL-1.2 tool

3.3 Establishing MHC Class I Alleles Distinct to CTL Epitopes

The MHC I alleles exclusive to the previously mentioned CTL epitopes have been found whereas NetMHCpan-4.1 was utilised (Reynisson et al., 2020b). For per epitope, there was peptide length of nine. The threshold for the classification of strong binders has been fixed at 0.5. The threshold for weak binders was set at 2.

Table 2 : The CTL epitope-specific, heavily binding MHC I alleles

Alleles	Peptides	Sequences no.	%Rank_EL	%Rank_BA	Aff(nM)
1 HLA A*01:01	ILDSTNNLY	1	0.007	0.014	14.40
1 HLA-A*01:01	RVDCIESLY	2	0.056	0.028	29.61
1 HLA-A*01:01	YTEWSKLKF	3	0.274	0.305	568.50
1 HLA-A*01:01	LTDGDIDFF	4	0.179	0.110	139.09
1 HLA-A*03:01	KILDFTQTY	5	0.238	0.946	359.80
1 HLA-A*26:01	DSVILKLVY	6	0.300	0.630	2008.66
1 HLA-A*26:01	KILDFTQTY	7	0.282	2.214	7687.25

1 HLA-B*58:01	KILDFTQTY	8	0.345	1.389	520.33
1 HLA-B*15:01	KQNKSWELY	9	0.058	0.312	56.45
1 HLA-B*15:01	KILDFTQTY	10	0.030	0.613	113.47

3.4 Antigenicity, Allergenicity, and Toxicity of CTL Epitopes Determined By In Silico

All of the CTL were highlighted as non-toxin in the toxinpred prediction system as a whole as can be observed in Figure 4.

Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq1	ILDSTNNLY	-1.03	Non-Toxin	-0.07	-0.13	-0.49	-1.00	1052.28
seq2	RVDCIESLY	-0.24	Non-Toxin	-0.15	0.14	0.09	-1.00	1260.56
seq3	YTEWSKLKF	-0.97	Non-Toxin	-0.19	-0.93	-0.12	1.00	1201.51
seq4	LTDGDIDFF	-1.00	Non-Toxin	0.03	0.26	0.00	-3.00	1042.23
seq5	KILDFTQTY	-1.55	Non-Toxin	-0.11	-0.28	-0.33	0.00	1128.42
seq6	DSVILKLVY	-1.15	Non-Toxin	0.09	1.22	-0.49	0.00	1049.41
seq7	KILDFTQTY	-1.55	Non-Toxin	-0.11	-0.28	-0.33	0.00	1128.42
seq8	KILDFTQTY	-1.55	Non-Toxin	-0.11	-0.28	-0.33	0.00	1128.42
seq9	KQNKSWELY	-0.59	Non-Toxin	-0.39	-1.94	0.24	1.00	1195.47
seq10	KILDFTQTY	-1.49	Non-Toxin	-0.10	-0.25	-0.30	0.00	1291.61
		-0.49	Non-Toxin					

1/1 50

Figure 4 : The picked CTL epitopes of the one eligible protein in the Toxinpred Server Results

Table 3 : A description of the indicated CTL epitopes' toxicity, allergenicity, and antigenicity

CTL epitopes	Antigenicity	Allergenicity	Toxicity
ILDSTNNLY	Non-antigen	Probable non-allergen	Non-toxin
RVDCIESLY	Antigen	Probable non-allergen	Non-toxin
YTEWSKLKF	Antigen	Probable non-allergen	Non-toxin
LTDGDIDFF	Antigen	Probable allergen	Non-toxin
KILDFTQTY	Antigen	Probable allergen	Non-toxin
DSVILKLVY	Non-antigen	Probable non-allergen	Non-toxin
KQNKSWELY	Antigen	Probable non-allergen	Non-toxin

3.5 Determination of the Binding HTL Epitopes:

The NetMHCIIpan 4.0 Server experienced an enormous amount of peptide sequences. It has a relationship to MHC II alleles (Reynisson et al., 2021b).

In results page for allele DRB1_0101 there is shown the percentile rank, score_EL, and the core of the peptides were also displayed. Figure 5 displays the first allele result page for the NetMHCIIpan 4.0 server. The result was in line with the score for prediction. This displayed the peptides with strong binding at the top of the listings. There is a total of strong binding (SB) HTL epitopes which had been known.

142	DRB1_0101	KNIANDFYQLTSREN	6	FYQLTSREN	0.980	FASTA	0.098949	9.93	NA	0.582612	91.47	10.86	
143	DRB1_0101	NIANDFYQLTSRENK	5	FYQLTSREN	1.000	FASTA	0.633212	1.70	NA	0.674318	33.91	4.07	<=WB
144	DRB1_0101	IANDFYQLTSRENKQ	4	FYQLTSREN	1.000	FASTA	0.828224	0.74	NA	0.690349	28.51	3.32	<=SB
145	DRB1_0101	ANDFYQLTSRENKQT	3	FYQLTSREN	1.000	FASTA	0.859977	0.59	NA	0.697134	26.49	3.03	<=SB
146	DRB1_0101	NDFYQLTSRENKQTP	2	FYQLTSREN	1.000	FASTA	0.609504	1.82	NA	0.673849	34.09	4.09	<=WB

Figure 5 : Both strong as well as weak binding peptides matching alleles in NetMHCIIpan 4.0 Server DRB1_0101 (Reynisson et al., 2021b).

3.6 Cytokine Stimulating Ability Taken through Defined HTL Epitopes

In order to screen for every one of the strong bindings HTL epitopes along with fulfil the three conditions of being cytokine inducers (IL-4, IL-10, and IFN-gamma), servers called IL4pred, IL10pred, and IFNepitope have been used. As the outcome, 13 HTL epitopes have been detected and are represented in yellow in the tables 4,5, 6, 7, and 8.

Table 4 : HTL epitopes which caused IFN-gamma, IL-4, and IL-10 (part 1)

	HTL peptide sequence	IFN	IL-10 Prediction	IL-4 Prediction
1	SREMQIENNLLSASQ	Negative	inducer	inducer
2	ETHLYYDTLYQYQGG	Positive	inducer	inducer
3	IVDFHFDIDSVILKL	Positive	inducer	inducer
4	RTNRYLRDIGSRVLT	Positive	inducer	inducer

	HTL peptide sequence	IFN	IL-10 Prediction	IL-4 Prediction
5	TNRYLRDIGSRVLTH	Negative	inducer	inducer
6	GEKIQWILDSTNNLY	Positive	non-inducer	inducer
7	EKIQWILDSTNNLYS	Negative	non-inducer	inducer
8	KIQWILDSTNNLYST	Negative	non-inducer	inducer
9	IQWILDSTNNLYSTC	Negative	non-inducer	inducer
10	FSREMQIENNLLSAS	Negative	inducer	inducer
11	REMQIENNLLSASQL	Negative	inducer	inducer
12	SREIKFKDVLKDADS	Positive	non -inducer	inducer

Table 5 : HTL epitopes which caused IFN-gamma, IL-4, and IL-10 (part 2)

13	IANDFYQLTSRENKQ	Negative	inducer	inducer
14	ANDFYQLTSRENKQT	Negative	inducer	inducer
15	NIANDFYQLTSRENK	Negative	inducer	inducer
16	NDFYQLTSRENKQTP	Negative	inducer	inducer
17	SFSREMQIENNLLSA	Negative	inducer	inducer
18	VFTGLVVMTIPIPIV	Positive	non-inducer	inducer

Table 6 : HTL epitopes which caused IFN-gamma, IL-4, and IL-10 (part 3)

19	FTGLVVMTIPIPIVD	Positive	non-inducer	inducer
20	TGLVVMTIPIPIVDF	Negative	non-inducer	inducer
21	CKKILDFTQTYQVVV	Positive	non-inducer	inducer
22	VISREIKFKDVLKDA	Positive	inducer	inducer
23	ISREIKFKDVLKDAD	Positive	non-inducer	inducer
24	REIKFKDVLKDADSD	Positive	non-inducer	inducer
25	EIKFKDVLKDADSDE	Negative	non-inducer	inducer
26	PGGFAITGKFALTDG	Positive	non-inducer	inducer
27	LPGGFAITGKFALTD	Positive	non-inducer	inducer
28	GGFAITGKFALTDGD	Negative	non-inducer	inducer
29	PSTVRGLKAPGNEKP	Positive	inducer	inducer

Table 7 : HTL epitopes which caused IFN-gamma, IL-4, and IL-10 (part 4)

30	SKLKFTVISREIKFK	Positive	inducer	inducer
31	KLKFTVISREIKFKD	Positive	inducer	inducer
32	SRVLTHIVKTLQIHN	Negative	inducer	inducer
33	GSRVLTHIVKTLQIH	Negative	inducer	inducer
34	ASQLKRVNAADFWDR	Positive	non-inducer	inducer
35	HIVKTLQIHNRQFKQ	Negative	inducer	inducer
36	IVKTLQIHNRQFKQK	Negative	non-inducer	inducer
37	VKTLQIHNRQFKQKF	Negative	inducer	inducer

Table 8 : HTL epitopes which caused IFN-gamma, IL-4, and IL-10 (part 5)

38	KTLQIHNRQF KQKFN	Negative	inducer	inducer
39	IKFKDVLKDA DSDEV	Negative	non-inducer	inducer
40	VLPGGFAITG KFALT	Positive	non-inducer	inducer
41	YDTLYQYQG GVYPAH	Positive	inducer	inducer
42	IPIVDFHFDIDS VIL	Positive	non-inducer	inducer
43	PIVDFHFDIDS VILK	Positive	inducer	inducer
44	VDFHFDIDSVI LKLV	Positive	inducer	inducer
45	WDRTNRYLR DIGSRV	Positive	inducer	inducer
46	DRTNRYLRDI GSRVL	Positive	inducer	inducer
47	NRYLRDIGSR VLTH	Positive	inducer	inducer
48	YGEKIQWILD STNNL	Negative	non-inducer	inducer

3.7 Evaluation of The Epitopes on B Cells

The main protein sequence that provided the results displayed herein as well as the has been saved in the IEDB Analysis Resource had been submitted to the Bepipred Linear Epitope Prediction 2.0 tools the strategy. A total of peptides have been proven to be B cell epitopes, and the results are listed in tables 9 and 10.

Table 9 : B cell epitope estimation employing Bepipred Linear Epitope Prediction 2.0 (part 1)

No.	Start	End	Peptide	Length
1	12	19	QYQGGVYP	8
2	29	29	C	1
3	50	56	GMHYTEW	7
4	67	82	EIKFKDVLKDADSDEV	16
5	130	226	RPPSNRPSEASAKNIANDFYQLTSRENKQTPDEEKRCLFFQQGPL EPPSTVRGLKAPGNEKPIQFPAHANEKMTESFLSDSWFGQKVRC KKILDFTQ	97
6	238	266	SFSREMQIENNLLSASQLKRVNAADFWDR	29
7	288	295	IHNRQFKQ	8
8	340	353	GFQNGGKSFLAQDE	14
9	366	375	VIYGEKIQWI	10

Table 10 : B cell epitope estimation employing Bepipred Linear Epitope Prediction 2.0 (part 2)

10	387	396	REKQNKSWEL	10
11	430	438	GDIDFFNWR	9

Figure 6 : The graph of B cell epitopes' score against position is presented by the IEDB service (Jespersen et al., 2017b). B cell epitopes exhibited an average score of 0.491, a minimum score of 0.229, and a maximum score of 0.670, as can be seen in figure 6 (Jespersen et al., 2017b).

3.8 Combination of Promising Vaccines

The different types of epitopes were connected by using the linkers which were discussed in the techniques section. For ensuring for the essential functional domain segregation, the EAAAK linker minimizes interference between the epitopes. Ala-Ala-Tyr, or AAY, is a linker that functions inside mammalian cells. There was location of proteasome cleavage. For that

reason, inside of cells, CTL epitopes were linked by AAY will effectively separate. Then there were by decreasing junctional immunogenicity and increasing the immunogenicity of individual epitopes, the GPGPG linker increases T helper cell responses. Cathepsin B, which is a lysosomal protease, that is targeting the di-lysine junction (KK). Thereby increasing the vaccine's immunogenicity, the KK linker reduces junctional immunogenicity. For constructing the suggested potential vaccines and for connecting all of the chosen epitopes, these linkers were very helpful. That linkers will improve antigen presentation and processing and reduce the likelihood of junctional antigen formation in these multi-epitope candidate vaccines. Then the built-in vaccination will also be more stable, stiff, and flexible structurally thanks to these linkers (Ayyagari et al., 2022).

A meticulous procedure is being subsequently followed, so as a result, seven CTL epitopes, thirteen HTL epitopes, while eleven B cell epitopes have been discovered established. thereby, the first vaccination is produced integrating two antigenic, non-toxin, non-allergen CTL, HTL, and B-cell epitopes. Consequently, by eliminating specific CTL, HTL, and B cell epitopes from every single vaccine—which are collectively referred to as i–viii here—six distinct vaccinations were successfully generated. The main objective of creating this vaccination with changes in combination impacted the vaccines' biochemical analysis. The list of all discovered vaccinations appears below.

Vaccine (i): Vaccine is Constructed incorporating 2 CTL, 2 HTL, 2 B-Cell epitopes which are all antigenic, Non toxin, Non allergen.

Vaccine (ii): Vaccine Constructed incorporating 3 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic, Non toxin, Non allergen.

Vaccine (iii): Vaccine Constructed incorporating 2 CTL, 1 HTL, 2 B-Cell epitopes which are all antigenic, Non toxin, Non allergen.

Vaccine (iv): Vaccine Constructed incorporating 2 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic, non-toxin, non-allergen.

Vaccine (v): Vaccine construction by incorporating 1 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic , Non toxin , Non allergen.

Vaccine (vi): Vaccine Constructed incorporating epitopes which are all antigenic, Non toxin , Non allergen.

Table 11 : AllerTOP, ToxinPred, and VaxiJen v2.0 CTL epitope antigenicity, allergenicity, & toxicity predictions (part 1)

CTLs	**Antigenicity (%)**	**Allergenicity**	**Toxicity**
ILDSTNNLY	0.4041 (Probable Non-antigen)	Non-allergen	Non-toxin
RVDCIESLY	0.9272 (Probable Antigen)	Probable non-allergen	Non-toxin
YTEWSKLKF	0.6968 (Probable Antigen)	Probable non-allergen	Non toxin
LTDGDIDFF	1.0123 (Probable Antigen).	Probable Allergen	Non-toxin
KILDFTQTY	1.2739 (Probable Antigen)	Probable Allergen	Non-toxin

Table 12 : AllerTOP, ToxinPred, and VaxiJen v2.0 CTL epitope antigenicity, allergenicity, & toxicity predictions (part2)

DSVILKLVY	0.2607 (Probable Non-antigen)	PROBABLE non-allergen	Non-toxin
KQNKSWELY	0.9114 (Probable Antigen)	Probable non-allergen	Non Toxin

Table 13 : AllerTOP, ToxinPred, VaxiJen v2.0, and allergenicity and toxicity estimates regarding HTL epitopes (part 1)

HTLs	**Antigenicity (%)**	**Allergenicity**	**Toxicity**
ETHLYYDTL YQYQGG	0.0420 (Probable Non-antigen)	Non-allergen	Non toxin
IVDFHFDIDS VILKL	0.8284 (Probable antigen)	Allergen	Non toxin

RTNRYLRDIG SRVLT	-0.3926(Probable Non-antigen)	Allergen	Non toxin
VISREIKFKD VLKDA	0.7512 (Probable Antigen)	Allergen	Non toxin
PSTVRGLKAP GNEKP	-0.213(Probable Non-antigen)	Non-allergen	Non toxin
SKLKFTVISR EIKFK	1.4981(Probable Antigen)	Non-allergen	Non toxin
KLKFTVISREI KFKD	1.8350 (Probable Antigen)	Non-allergen	Non toxin

Table 14 : AllerTOP, ToxinPred, VaxiJen v2.0, and allergenicity and toxicity estimates regarding HTL epitopes (part 2)

YDTLYQYQGGVYPAH	0.2708 (Probable Non-antigen).	Allergen	Non toxin
PIVDFHFDIDSVILK	0.7103 (Probable Antigen).	Allergen	Non toxin
VDFHFDIDSVILKLV	0.8831 (Probable Antigen).	Allergen	Non toxin
WDRTNRYLRDIGSRV	0.3102 (Probable Non-antigen).	Non-allergen	Non toxin
DRTNRYLRDIGSRVL	-0.2986 (Probable Non-antigen).	Non-allergen	Non toxin
NRYLRDIGSRVLTH	-0.6715 (Probable Non-antigen).	Non-allergen	Non toxin

Table 15 : All B cell epitope allergenicity, toxicity, and antigenicity projections in AllerTOP, ToxinPred, & VaxiJen v2.0

B cell epitopes	Antigenicity (%)	Allergenicity	Toxicity
QYQGGVYP	0.4069 (Probable Non-antigen).	Non-allergen	Non-toxin
GMHYTEW	1.2303 (Probable Antigen)	Non-allergen	Non-toxin
EIKFKDVLKDADSDEV	0.8947 (Probable Antigen).	Non-allergen	Non-toxin
SFSREMQIENNLLSASQLKRVNAADFWDR	0.5734 (Probable Antigen).	Non-allergen	Non-toxin
IHNRQFKQ	-0.1543 (Probable Non-antigen).	Non-allergen	Non-toxin
GFQNGGKSFLAQDE	0.0475 (Probable Non-antigen).	Non-allergen	Non-toxin

VIYGEKIQWI	1.1003 (Probable Antigen).	Allergen	Non-toxin
REKQNKSWEL	0.8581 (Probable Antigen).	Allergen	Non-toxin
GDIDFFNWR	1.6291 (Probable Antigen).	Allergen	Non-toxin

Table 16 : A candidate vaccine is made up of constructs that make use of linkers to link together several CTL, HTL, and B cell epitope combinations

SL	**Criteria**	**Sequence of the vaccine construction**
1	Vaccine i Constructed incorporating 2 CTL, 2 HTL, 2 B-Cell epitopes which are all antigenic, non-toxin, non-allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLPMRVDCIESLYFRCVFFKSGMHY TEWSKLKFTVISREIKFKDVLKDADSDEVFTGLVVMTIPIPIVDFHFDIDSVILKLV YPRLVHREIVLRLYDLICVRPPSNRPSEASAKNIANDFYQLTSRENKQTPDEEKR CLFFQQGPLEPPSTVRGLKAPGNEKPIQFPAHANEKMTESFLSDSWFGQKVRCK KILDFTQTYQVVVCWYELSFSREMQIENNLLSASQLKRVNAADFWDRTNRYLR DIGSRVLTHIVKTLQIHNRQFKQKFNCNFPDNFSFDRLLSFMQLGKDFWILNLTL DSCIIKAIICFLGFQNGGKSFLAQDEVWGDLIDCSKGSVIYGEKIQWILDSTNNLY STCREKQNKSWELYVDCCALYVSEKLELDFVLPGGFAITGKFALTDGDIDFFNW RFGLSEAAAKRVDCIESLYAAYKQNKSWELYGPGPGSKLKFTVISREIKFKGPGP GKLKFTVISREIKFKDKKGMHYTEWKKEIKFKDVLKDADSDEV

2	Vaccine ii (Constructed incorporating 3 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic, non-toxin, non-allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLPMRVDCIESLYFRCVFFKSGMHY TEWSKLKFTVISREIKFKDVLKDADSDEVFTGLVVMTIPIPIVDFHFDIDSVILKLV YPRLVHREIVLRLYDLICVRPPSNRPSEASAKNIANDFYQLTSRENKQTPDEEKR CLFFQQGPLEPPSTVRGLKAPGNEKPIQFPAHANEKMTESFLSDSWFGQKVRCK KILDFTQTYQVVVCWYELSFSREMQIENNLLSASQLKRVNAADFWDRTNRYLR DIGSRVLTHIVKTLQIHNRQFKQKFNCNFPDNFSFDRLLSFMQLGKDFWILNLTL DSCIIKAIICFLGFQNGGKSFLAQDEVWGDLIDCSKGSVIYGEKIQWILDSTNNLY STCREKQNKSWELYVDCCALYVSEKLELDFVLPGGFAITGKFALTDGDIDFFNW RFGLSEAAAKRVDCIESLYAAYYTEWSKLKFAAYKQNKSWELYGPGPGSKLKF TVISREIKFKGPGPGKLKFTVISREIKFKDKKGMHYTEW
3	Vaccine iii (Constructed incorporating 2 CTL, 1 HTL, 2 B-Cell epitopes which are all antigenic, non-toxin, non-allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLPMRVDCIESLYFRCVFFKSGMHY TEWSKLKFTVISREIKFKDVLKDADSDEVFTGLVVMTIPIPIVDFHFDIDSVILKLV YPRLVHREIVLRLYDLICVRPPSNRPSEASAKNIANDFYQLTSRENKQTPDEEKR CLFFQQGPLEPPSTVRGLKAPGNEKPIQFPAHANEKMTESFLSDSWFGQKVRCK KILDFTQTYQVVVCWYELSFSREMQIENNLLSASQLKRVNAADFWDRTNRYLR DIGSRVLTHIVKTLQIHNRQFKQKFNCNFPDNFSFDRLLSFMQLGKDFWILNLTL DSCIIKAIICFLGFQNGGKSFLAQDEVWGDLIDCSKGSVIYGEKIQWILDSTNNLY STCREKQNKSWELYVDCCALYVSEKLELDFVLPGGFAITGKFALTDGDIDFFNW RFGLSEAAAKYTEWSKLKFAAYKQNKSWELYGPGPGSKLKFTVISREIKFKKKG MHYTEWKKSFSREMQIENNLLSASQLKRVNAADFWDR

4	Vaccine iv Constructed incorporating 2 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic, non-toxin, non allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLP MRVDCIESLYFRCVFFKSGMHYTEWSKLKFT VISREIKFKDVLKDADSDEVFTGLVVMTIPI PIVDFHFDIDSVILKLVYPRLVHREIVLRLY DLICVRPPSNRPSEASAKNIANDFYQLTSRE NKQTPDEEKRCLFFQQGPLEPPSTVRGLKAP GNEKPIQFPAHANEKMTESFLSDSWFGQKVR CKKILDFTQTYQVVVCWYELSFSREMQIENN LLSASQLKRVNAADFWDRTNRYLRDIGSRVL THIVKTLQIHNRQFKQKFNCNFPDNFSFDRL LSFMQLGKDFWILNLTLDSCIIKAIICFLGF QNGGKSFLAQDEVWGDLIDCSKGSVIYGEKI QWILDSTNNLYSTCREKQNKSWELYVDCCAL YVSEKLELDFVLPGGFAITGKFALTDGDIDF FNWRFGLSEAAAKYTEWSKLKFAAYKQNKSW ELYGPGPGSKLKFTVISREIKFKGPGPGKLK FTVISREIKFKDKKGMHYTEW

5	Vaccine v Constructed incorporating 1 CTL, 2 HTL, 1 B-Cell epitopes which are all antigenic, non-toxin, non-allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLP MRVDCIESLYFRCVFFKSGMHYTEWSKLKFT VISREIKFKDVLKDADSDEVFTGLVVMTIPI PIVDFHFDIDSVILKLVYPRLVHREIVLRLY DLICVRPPSNRPSEASAKNIANDFYQLTSRE NKQTPDEEKRCLFFQQGPLEPPSTVRGLKAP GNEKPIQFPAHANEKMTESFLSDSWFGQKVR CKKILDFTQTYQVVVCWYELSFSREMQIENN LLSASQLKRVNAADFWDRTNRYLRDIGSRVL THIVKTLQIHNRQFKQKFNCNFPDNFSFDRL LSFMQLGKDFWILNLTLDSCIIKAIICFLGF QNGGKSFLAQDEVWGDLIDCSKGSVIYGEKI QWILDSTNNLYSTCREKQNKSWELYVDCCAL YVSEKLELDFVLPGGFAITGKFALTDGDIDF FNWRFGLSEAAAKKQNKSWELYGPGPGSKLK FTVISREIKFKGPGPGKLKFTVISREIKFKD KKGMHYTEW

6	Final vaccine vi Constructed incorporating epitopes which are all antigenic, non-toxin, non-allergen	METHLYYDTLYQYQGGVYPAHICLPTDVCLPMRVDCIESLYFRCVFF KSGMHYTEWSKLKFTVISREIKFKDVLKDADSDEVFTGLVVMTIPIPIV DFHFDIDSVILKLVYPRLVHREIVLRLYDLICVRPPSNRPSEASAKNIAN DFYQLTSRENKQTPDEEKRCLFFQQGPLEPPSTVRGLKAPGNEKPIQFP AHANEKMTESFLSDSWFGQKVRCKKILDFTQTYQVVVCWYELSFSRE MQIENNLLSASQLKRVNAADFWDRTNRYLRDIGSRVLTHIVKTLQIH NRQFKQKFNCNFPDNFSFDRLLSFMQLGKDFWILNLTLDSCIIKAIICF LGFQNGGKSFLAQDEVWGDLIDCSKGSVIYGEKIQWILDSTNNLYSTC REKQNKSWELYVDCCALYVSEKLELDFVLPGGFAITGKFALTDGDID FFNWRFGLSEAAAKKQNKSWELYGPGPGKLKFTVISREIKFKDKKGM HYTEW

3.9 In-silico Biochemical Analysis of Candidate Vaccines:

Some key distinct features were seen in the designed vaccines(i-vi) while doing the biochemical analysis. A short review contrasting the vaccines along with biochemical analysis is shown in the tables 17.

Table 17 : Biochemical analysis differentiation of 6 vaccines samples

	Vaccine i	**Vaccine ii**	**Vaccine iii**
PROTPARAM VACCINE STABILITY [primary protein instability index= 32.60]	The instability index (II) is computed to be 33.57 protein as stable. **Aliphatic index:** 83.96 **Grand average of hydropathicity (GRAVY):** -0.283 **The quantity of amino acids**: 535 **Molecular weight:** 62045.47	The instability index (II) is computed to be 33.70 protein is stable. **Aliphatic index:** 83.27 **Grand average of hydropathicity (GRAVY):** -0.260 **The quantity of amino acids:** 529 **Molecular weight:** 61444.80	The instability index (II) is computed to be 33.48 protein as stable. **Aliphatic index:** 82.33 **Grand average of hydropathicity (GRAVY):** -0.301 **The quantity of amino acids:** 528 **Molecular weight:** 61525.75
VAXIJEN-2.0 ANTIGENICITY [primary protein antigenicity=0.5142]	**0.5654** (Probable **ANTIGEN**).	**0.5671** (Probable **ANTIGEN**).	**0.5391** (Probable **ANTIGEN**).
ALLERGEN - ONLINE ALLERGENICITY	No allergenicity	No allergenicity	No allergenicity
Z score	-6.2	-6.2	-6.35
RAMACHANDRAN FAVORED REGION	86.21%	86.21%	86.06%
Phyre 2	**Confidence 100 % Coverage 70%**	**Confidence 100% Coverage 71 %**	**Confidence 100% Coverage 70%**

	Vaccine (iv)	Vaccine (v)	Vaccine (vi) [FINAL VACCINE]
PROTPARAM VACCINE STABILITY [primary protein instability index= 32.60]	The instability index (II) is computed to be 31.99 protein as stable. **Aliphatic index:** 82.75 **Grand average of hydropathicity (GRAVY):** -0.273 **The quantity of amino acids:** 517 **Molecular weight:** 60060.23	The instability index (II) is computed to be 32.94 protein is stable. **Aliphatic index:** 83.54 **Grand average of hydropathicity (GRAVY):** -0.268 **The quantity of amino acids:** 505 **Molecular weight:** 58571.53	The instability index (II) is computed to be 34.10 protein is stable. **Aliphatic index:** 83.98 **Grand average of hydropathicity (GRAVY):** -0.263 **The quantity of amino acids:** 485 **Molecular weight:** 56399.91
VAXIJEN-2.0 ANTIGENICITY [primary protein antigenicity=0.5142	0.5699 (Probable ANTIGEN).	0.5703 (Probable ANTIGEN).	0.5566 (Probable ANTIGEN
ALLERGEN -ONLINE ALLERGENICITY	No allergenicity	No allergenicity	No allergenicity

Z score	-6.39	-6.07	-5.95
RAMACHANDRAN FAVORED REGION	86.06%	86.74%	87.53%
Phyre 2	**Confidence 100%** Coverage 72%	**Confidence 100 %** Coverage 74%	**Confidence 100%** Coverage 77%

Finally, vaccine(vi) was selected, all of the vaccines biochemical results.

Both figures depict the manner in which the ProtParam server was implemented for evaluating the final vaccine's consistency. The vaccine was having 485 amino acids, as well as the molecular weight of it has been determined to be 56399.91 Da founded on the ProtParam.The probable pI for that vaccine has been estimated through ProtParam and found to be 7.81. Around the pH of the isoelectric point (pI), the net charge associated with a specific protein is zero. Each amino acids a percentage in the vaccination was highlighted. It is clear from the amino acid structure that the amino acid threonine has the largest sum. In all, there were 61 positively charged residues as well as 59 negatively charged residues in altogether. An anticipated stable protein in ProtParam is denoted having an instability index that's below 40. This vaccine's destabilization index was 34.10, demonstrating that it was consistent. The sum of the of the hydropathy of every one of them reduced by the number of residues generated the vaccine's average level of hydropathy (GRAVY), which found determined to be -0.263. (Walker et al., n.d.-b).

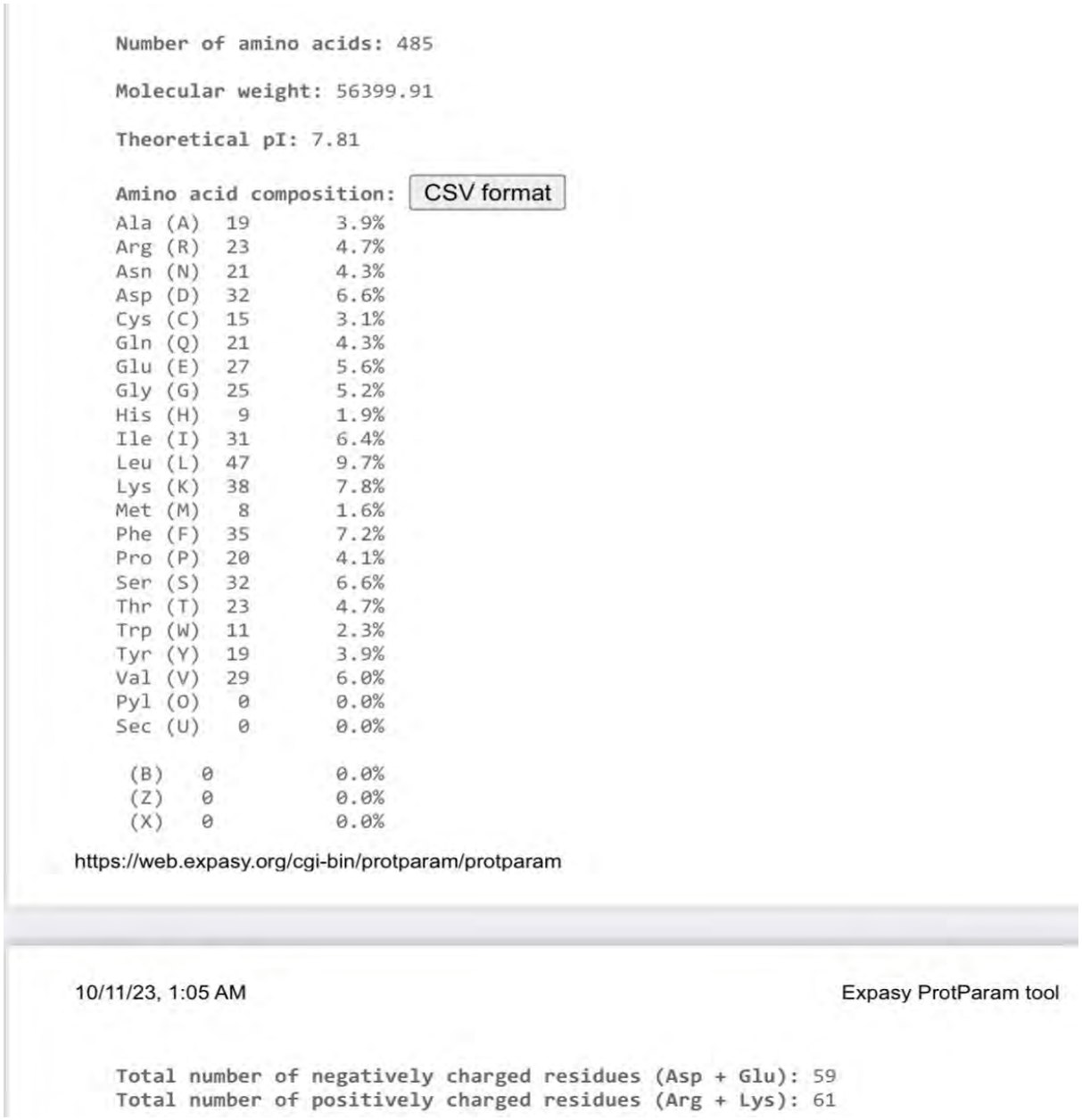


Figure 7 : The ProtParam server's amino acid amount and composition, theoretical pI, molecular weight, and charge residues (Walker et al., n.d.-b).

Atomic composition:

Carbon	C	2572
Hydrogen	H	3933
Nitrogen	N	663
Oxygen	O	720
Sulfur	S	23

Formula: $C_{2572}H_{3933}N_{663}O_{720}S_{23}$
Total number of atoms: 7911

Extinction coefficients:

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

Ext. coefficient 89685
Abs 0.1% (=1 g/l) 1.590, assuming all pairs of Cys residues form cystines

Ext. coefficient 88810
Abs 0.1% (=1 g/l) 1.575, 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 34.10
This classifies the protein as stable.

Aliphatic index: 83.98

Grand average of hydropathicity (GRAVY): -0.263

Figure 8 : Formula, half-life, instability index, atomic composition, & GRAVY within the ProtParam server (Walker et al., n.d.-b).

3.10 Toxicity and Allergenicity Prediction for Made Vaccine:

While the vaccination code had been transmitted to the Allergen Online server, the vaccine had not been anticipated from the program to be allergic. With the goal to ascertain whether or not the searched sequence could possess allergenic characteristics the Allergen Online server then examines the inputted amino acid sequence with allergen sequences within its database of entries (Goodman et al., 2016). The vaccination had been expected to be non-allergic

considering the web server couldn't find any correspondence against its recognized allergic sequences, the server determined that there had been no match among the vaccine's known allergen combinations and as a result indicated the vaccine would prove non-allergenic. The vaccine sequence was created and presented to the T3DB server; no results showed up, demonstrating that the sequence was non-toxic (Lim et al., 2009). The server was going to get a toxicity match result if the sequence was comparable a dangerous component in the server library. Nevertheless, considering the website didn't yield any conclusions, it ended up being decided that the vaccine had been anticipated to be secure.

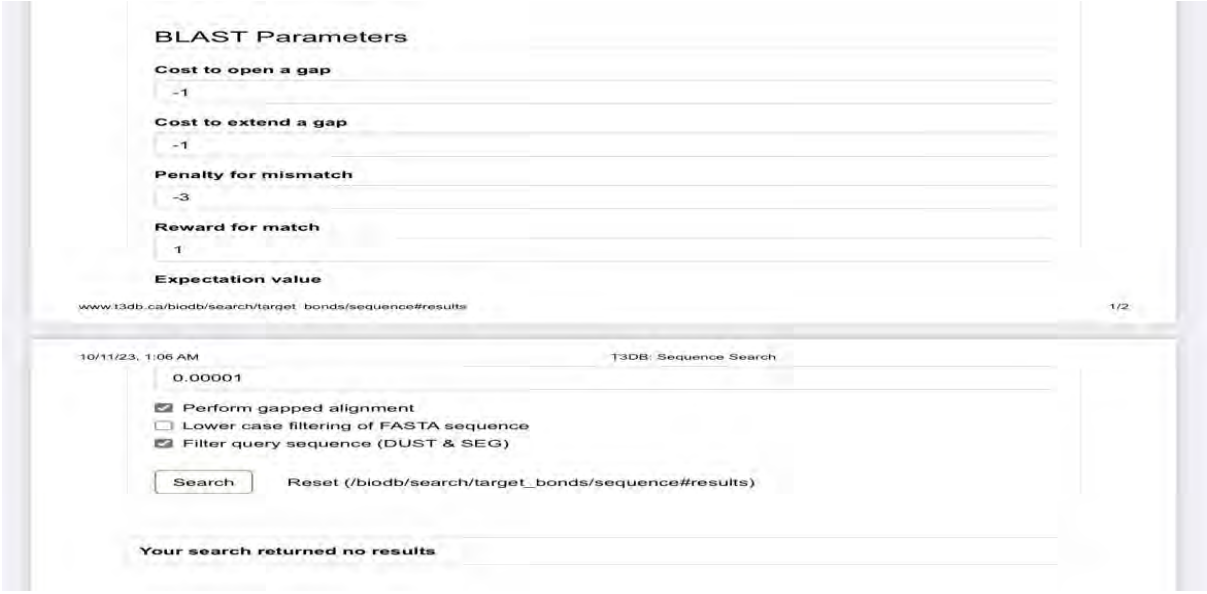


Figure 9 : Finding the toxicity output via the T3DB server (Wishart et al., 2015).

3.11 Models homology

Thus, the Phyre 2 server was implemented to acquire the vaccine's homology modeling (figure 10). Following this, the server produces a 77% coverage, 100% confidence 3-dimensional model of the vaccination. In the present scenario, 375 residues, or 77% of the sequence, were 100% confidently modeled. For the experiment to produce a precise conclusion, a high coverage percentage and confidence are essential. Consequently, in alongside easing the visual

examination of the vaccine's structure, the 3D model of the vaccine contributed in complete the upcoming biochemical examination of the vaccine. (Kelley et al., 2015).

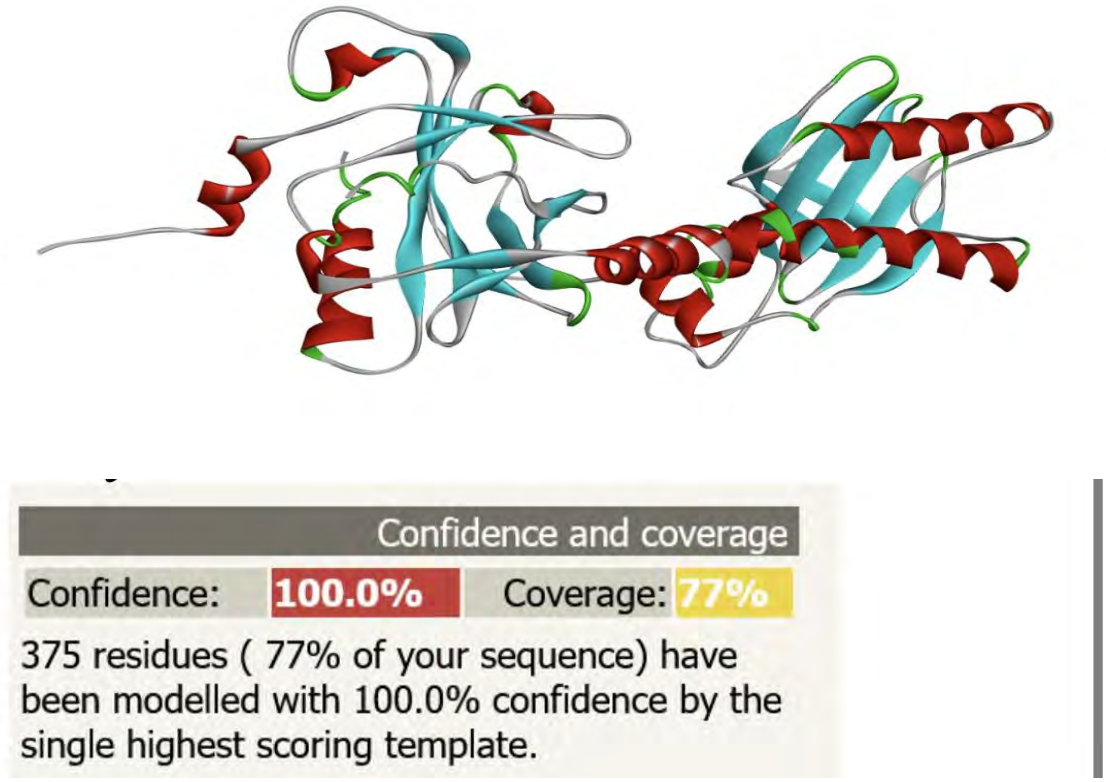


Figure 10 : Phyre 2 server homology modelling of vaccination (Kelley et al., 2015).

3.12 Evaluation using Z-Score with Ramachandran Plots

The last version of the vaccine's pdb file has been generated using the phyre 2 server which could be examined with the Discovery Studio program. In the near future, biochemical analysis was performed as well employing the pdb file. Z-score relative number of residues graph was compiled on the overall model quality thanks to vaccination pdb file submission in ProSAweb. The Z-score [figure: 11(a)] was -5.95. A second graph that highlighted the correlation across knowledge-based energy as well as sequence position was acquired by the local model quality website (Wiederstein & Sippl, 2007).

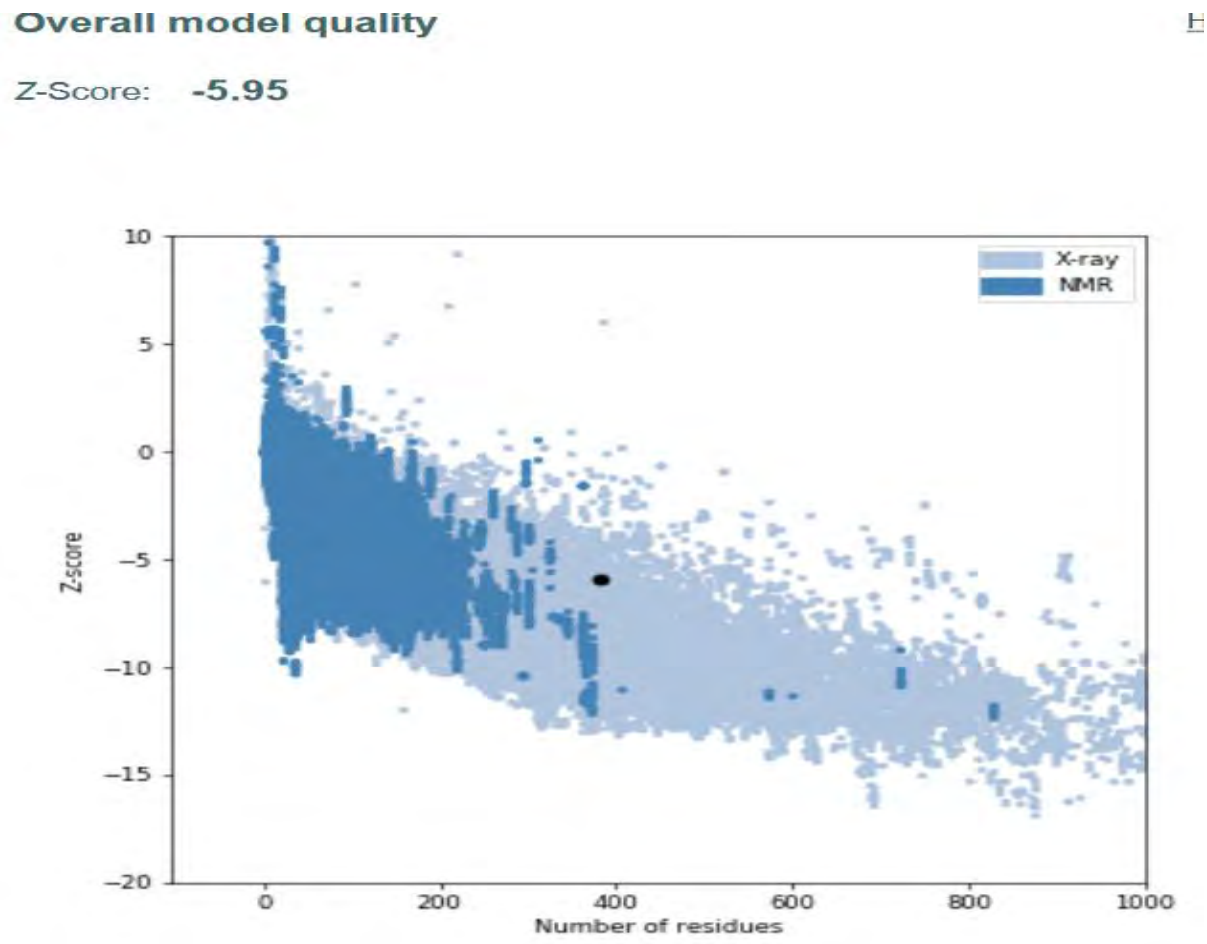
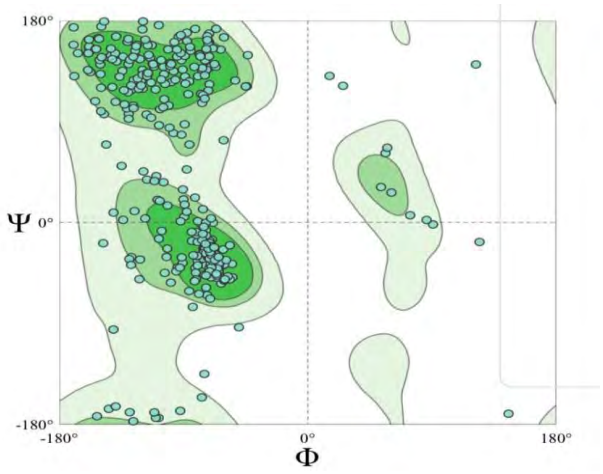


Figure 11 : (a) Z-score within the ProSAweb server and the Z-score vs. number of residues graph (Wiederstein & Sippl, 2007).

Figure 12 : (b) Local model quality connecting sequence position and knowledge-based energy (Wiederstein & Sippl, 2007).

The Ramachandran plots of the vaccine, in conjunction with the accuracy of estimation, residue quality, as well as MolProbity data, had been generated from the SWISS-MODEL server's structure review (Gopalakrishnan et al., 2007b). As per to the MolProbity data, the vaccine's Ramachandran favored region was 87.53%, and 3.18% Ramachandran outliers with 0.0% rotamer outliers.



(a)

MolProbity Results	
MolProbity Score	3.04

	Ramachandran Favoured	87.53%
☐	Ramachandran Outliers	3.18%
	Rotamer Outliers	0.00%
	C-Beta Deviations	0

(b)

Figure 13 : (a) The vaccine's Ramachandran plots created by SWISS-MODEL; (b) MolProbity conclusions in SWISS-MODEL (Gopalakrishnan et al., 2007b).

3.13 The Last Vaccine's Molecular Docking with the Appropriate Human Receptor

As the receptor molecule exhibiting the human toll-like receptor 9 (TLR9) crystal structure, 5y3m was registered into the PATCHDOCK server. In nature, the ligand molecule became available as the final vaccine. Twenty possibilities for the receptor-ligand complex are displayed with the server there. With a score of 21044, an area of 3265.90, and an atomic contact energy (ACE) of 333.11, the highest scoring complex (solution 1) was determined. -0.54 0.29 2.40 390.74 13.86 -176.60 being the advancement of this complex (figure 13 and 14). The molecule surface model has been employed via the Patchdock server to create flat, concave, as well as convex patches. The server subsequently connects complementary patches for producing alternative modifications. Later that, a rating was then carried out using the root mean square deviation (RMSD), which is the last stage in eradicating any redundant alternatives (Schneidman-Duhovny et al., 2005b).

The Toll-like receptors (TLRs) are inherent immune receptors that recognise pathogens by detecting their linked molecular patterns. The combination of they have a significant effect on both the acute and long-lasting inflammatory processes, as well as several medical conditions. TLR9 has been linked to a number of illnesses, namely psoriasis and defense against viral infections, making it a viable target for medicinal properties. discovery. TLR9 is found in the endosome region, and dsDNA, a protein with CpG motives seen in microbial DNA, was engaging the area in question (Zatsepin et al., 2016).

Molecular Docking Algorithm Based on Shape Complementarity Principles

[About PatchDock] [Web Server] [Download] [Help] [FAQ] [References]

Receptor	Ligand	Complex Type	Clustering RMSD	User e-mail
5y3m	10 2.pdb	Default	4.0	md.naimur.rahman.nabin@g.bracu.ac.bd

Solution No	Score	Area	ACE	Transformation
1	21044	3265.90	333.11	-0.54 0.29 2.40 390.74 13.86 -176.60

Figure 14 : Highest-scoring PATCHDOCK solution (Schneidman-Duhovny et al., 2005b).

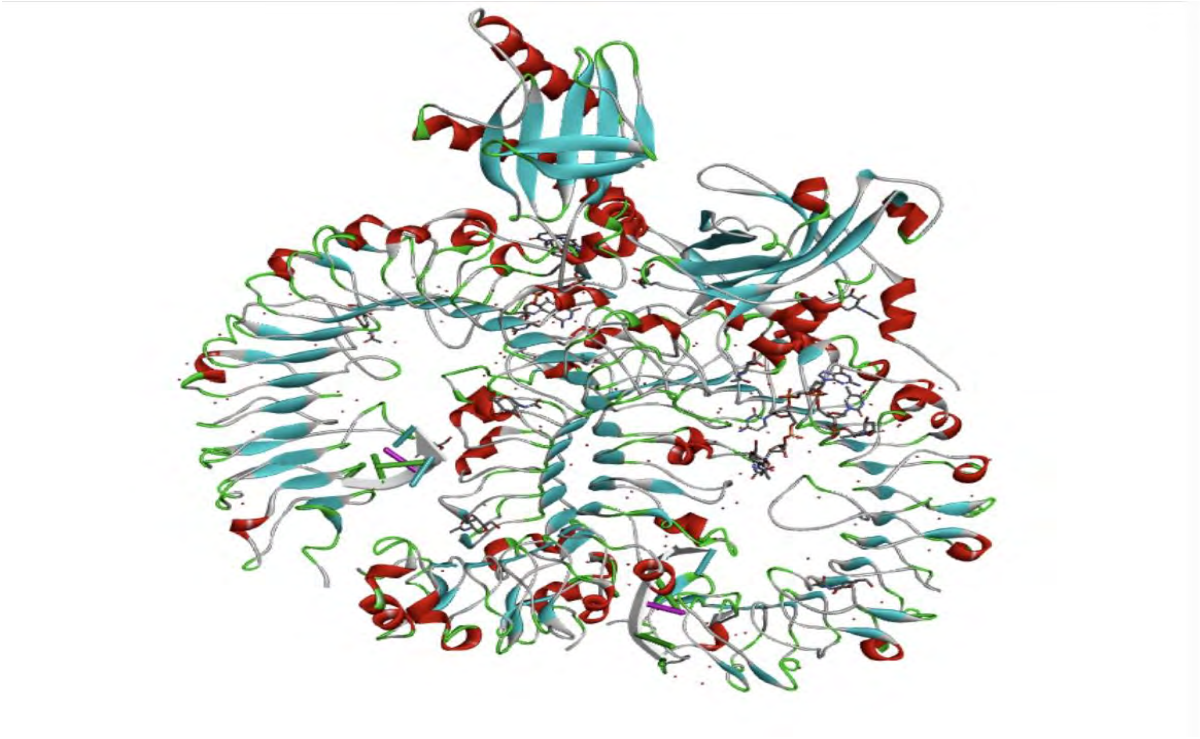
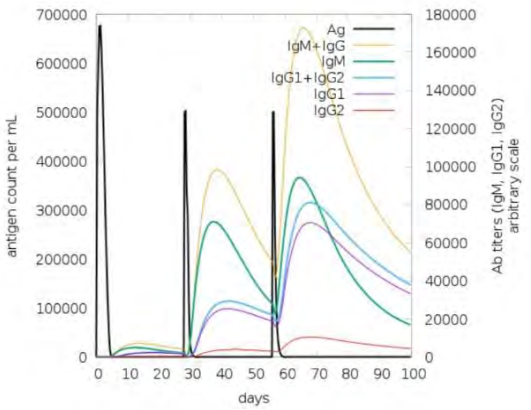


Figure 15 : Vaccine and receptor molecular docking in PATCHDOCK (Schneidman-Duhovny et al., 2005b).

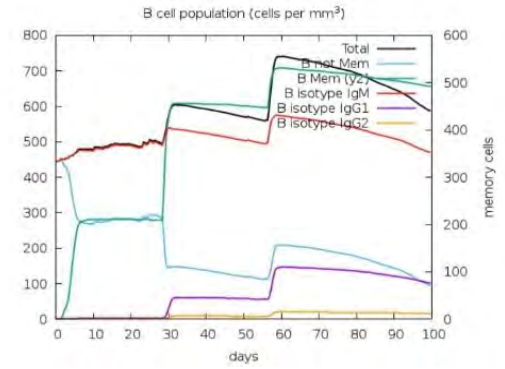
3.14 Immune Simulations:

To maintain a certain point of vaccine in the body there was enough for producing the intended prolonged immunological response, many doses which there including booster doses that are typically required. Thereby monitoring the levels of particular antibodies or immunoglobulin there might be in relation to the vaccination schedule and duration is important for evaluating

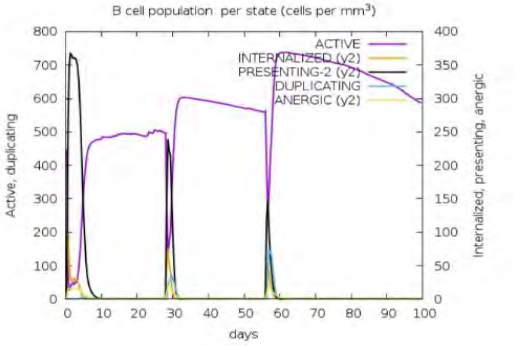
the immunological response that the body is being mounted. Then there was evaluation is required of the rise and fall in various immune cell populations which in response to vaccination doses. Based on the figure, the C-IMMSIM server anticipated and graphically depicted these characteristics in response for the vaccination (Rapin et al., 2010).



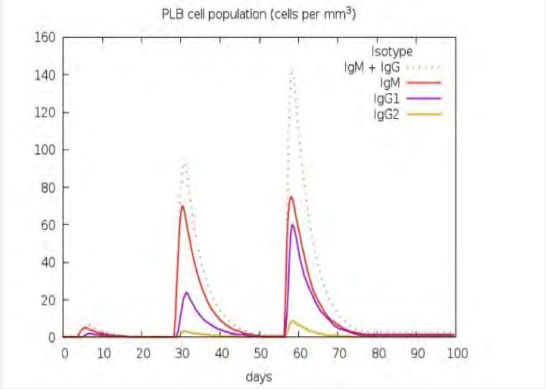
(a) The count of antigen per mL and antibody titers (Rapin et al., 2010).



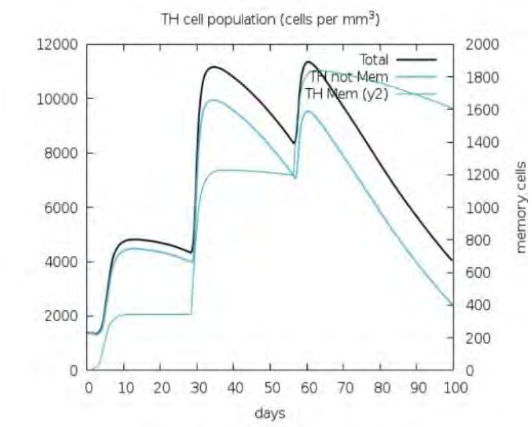
(b) B lymphocyte and memory number of cells together (Rapin et al., 2010).



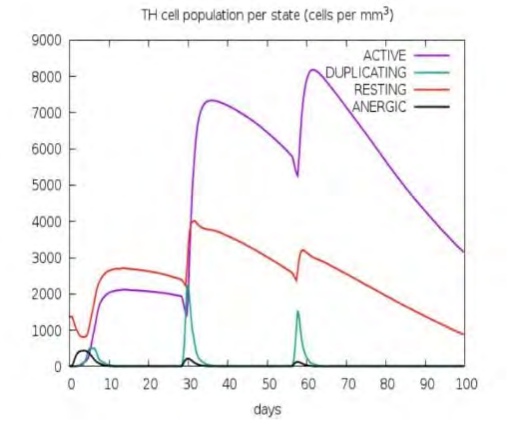
(c) B cell population's entity-state (Rapin et al., 2010).



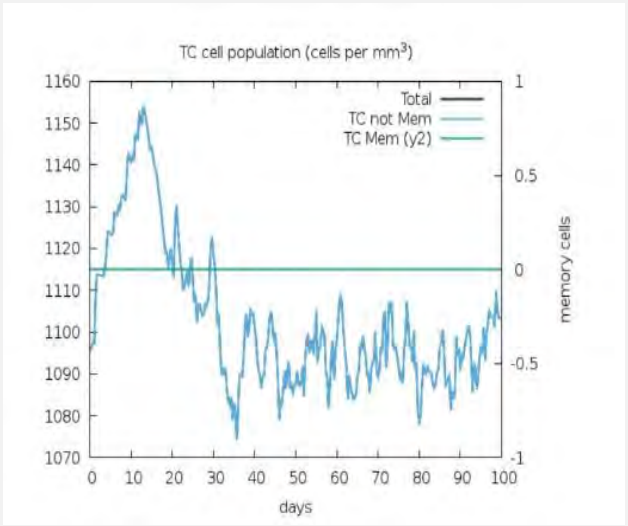
(d) Volume of plasma B lymphocytes defined by the isotypes (Rapin et al., 2010).



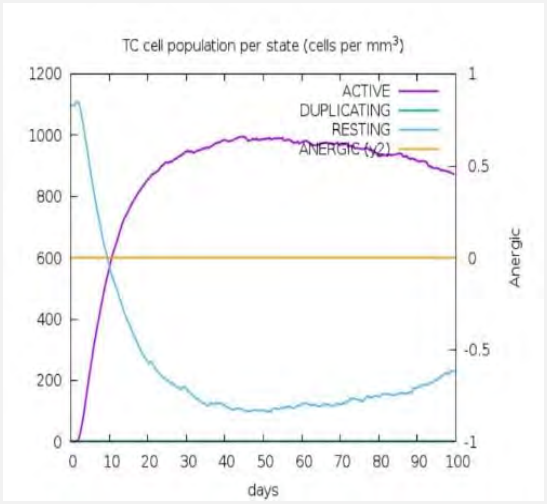
(e) Frequency of T helper lymphocytes (CD4) (Rapin et al., 2010).



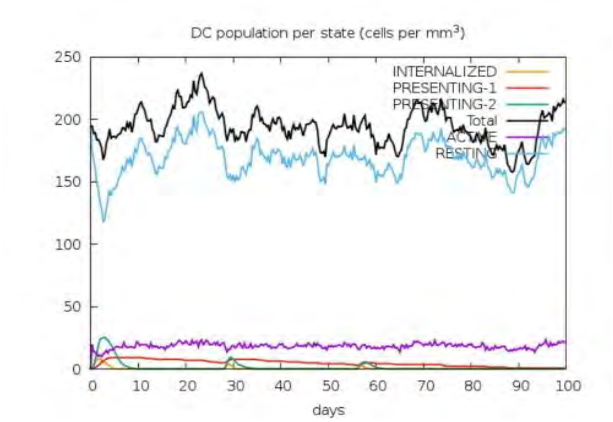
(f) A CD4 Helper T cell's entity designation (Rapin et al., 2010).



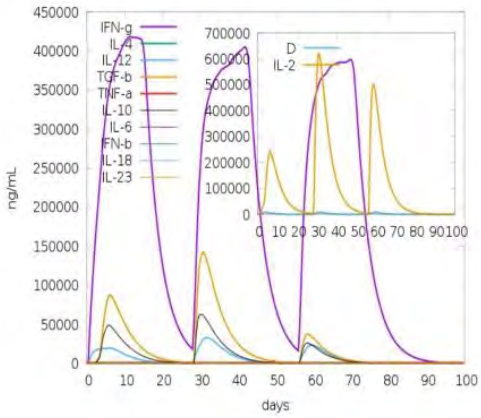
(g) Ratio of cytotoxic (TC) CD8 T cells (Rapin et al., 2010).



(h) Percentage of cytotoxic (TC) CD8 T lymphocytes per entity (Rapin et al., 2010).



(i) *Dendritic cells in general quantity and state (Rapin et al., 2010).*



(j) Levels of cytokines, among which are IL-4, IL-10, and IFN-γ (Rapin et al., 2010).

Figure 16 : Graphs showing the C-immsim immune simulation (a–j)

There is an essential component of the acquired immunity is antibodies which there the increase in antigen count per milliliter then after the initial vaccination is seen in Figure (a). IgM and IgG antibodies grow somewhat over the first 28 days of the period and after that they climb dramatically and that is because of the frequent delivery and completion of the vaccination doses, roughly 60 days.

There was an antibody production which supported the vaccine's predicted for the induction of the desirable immunological response. Then there were the immune cells which is called B lymphocytes that are responsible for both the production of antibodies and memory cells in the body. In the body there can more quickly produce protective immune function and recognize the same disease in the future thanks to these memory cells made from B cells. A graph representing the total number of memory cells and B lymphocytes upon immunization based on several isotypes is shown in Figure (b). In the Figure (c)'s graph there distinguished between and showed the B cells for each entity state. Thereby there was providing an information on the quantity of B cells that were internalizing the antigen and the quantity of B cells which

were active and after that the quantity of B cells exhibiting on Class II, and the quantity of B cells going through duplication and anergy. The quantity of plasma B lymphocytes (PLB) and their isotypes—IgG1, IgG2, IgM, and IgM + IgG—were demonstrated in Graph (d). In the Figure (e) there was projected the number of CD4 T helper cells after each dosage of immunization. There, the amount of Helper T cells that were predicted to be active and hence completing duplication, in a resting state which might be in an anergy state that was simulated using the C-IMMSIM (f). There the total and memory numbers of CD8 T cytotoxic (TC) cells as well as their active, duplicative. In the Figures (g) and (h) there is depicted the simulated resting and anergic states. Thereby in the figure (i) the graph displays the differentiation of dendritic cells into active, resting, antigen-presenting, and internalized states. PRESENTING-1 signified the DC cells reporting on MHC class I molecules in the present case, but PRESENTING-2 marked the DC cells reporting on MHC class II molecules. Following that, the interleukin-2 level and warning signal were highlighted in the graph's inset figure, as illustrated in figure (j).

Chapter 4

Discussion

These days, cytomegalovirus infections and the diseases their generate are important health issues. The study states that CMV infection has been associated to T-cell and lymphocyte limitation, mostly MHC class I and II. A specific antigen that is safe, provides a wide spectrum of immunogenicity, and is capable of maintaining an immune response is what's needed to deal with CMV strains (Sabbaghian et al., 2014).

The primary protein that was intended to be created into a vaccine was the Cytomegalovirus Capsid Vertex Component 1 protein, whose has a relationship with the virus's life cycle and contributes for creating the structure of a viral capsid, encapsulating and safeguarding the genetic material of the virus. Using the primary protein, the CTL, HTL, and B-cell epitopes were investigated. In this experiment, the NetCTL-1.2 set with a 0.75 threshold has been employed to detect 14 CTL epitopes for the protein. After that, epitopes were removed based on binding alleles, allergenicity (AllerTOP v. 2.0), toxicities (ToxinPred), and antigenicity (VaxiJen v2.0). In the end, seven CTL epitopes were chosen from the group. NetMHCIIpan 4.0 Server was applied in the HTL epitopes selection areas for finding HTL epitopes with strong binding, of which 48 were found for the primary protein. Following that, the CTL epitopes screening for allergenicity and toxicity were examined. The subsequent steps were assessing their capacities to see whether they might induce IL-4 (IL4pred), IFN-γ (IFNepitope), IL-10 (IL10pred), and 13. In addition, the IEDB tool's Bepipred Linear Epitope Prediction 2.0 approach was applied in selecting the primary protein epitopes for B cells. Nine B cell epitopes were selected using the CTL epitope screening technique. All among the main protein's chosen epitopes—seven CTL epitopes, thirteen HTL epitopes, and nine B cell epitopes—were utilised to construct vaccines, and linkers were added to improve communication between the protein

and epitopes. CTL, HTL, and B cell epitopes have been combined in order to generate six vaccines (i to vi) (table 16). Ultimately, vaccine (vi) with prime scores across all the created vaccines was ultimately selected as the ultimate vaccine based on the findings of the biochemical study. The vaccination indicated the stability score range when the Protparam tool was used for confirming stability. In addition to stability, the server provided an aliphatic index score, hydrophilic characteristic, and negative GRAVY score. The vaccine's antigenicity was elevated from the primary protein, which was suitable given the escalating immune response. The AllergenOnline tool classified the vaccine as non-allergen and states that it won't cause any allergic reactions. With the use of the T3DB server, no harmful entity was discovered in the vaccination. Having 100% confidence, 77% coverage was found employing the Phry2 tool for homology modelling. Using human TLR9 and a produced vaccine, the Patchdook tool is being used to determine binding affinity. Z-score and Ramachandran plot analysis is done using ProSAweb and SWISS-MODEL. The bottom set, the C-IMMSIM tool, was used to assess the immunological responses to the vaccine. It showed multiple rising peaks of different antibodies, such as IgG and IgM, as well as B-cells and CTL cells.

Chapter 5

Conclusion

In the end, there is currently no vaccine to aid in preventing the spread of the CMV disease. Everyone can be affected by that infection. In particular, this virus can stay in the body forever without causing any disease. CMV could end up in visual issues, hearing loss, etc. For these motives, should a virus threat arise in the future, there may be a vaccination that is effective and has a better prognosis. In this research, the Capsid Vertex Component 1 Protein of the cytomegalovirus was selected using an in-silico technique to create a multi-epitope vaccination. The vaccination originated through the use of a methodical, step-by-step approach to gather essential epitopes that relied on in-silico technologies. The biochemical study showed encouraging results; this was a critical criterion. Also, as the vaccine created using an in-silico approach has specific limitations, further research in vitro and in vivo is required. It's possible that the advised vaccination will stimulate the immune system to fight the dangerous CMV infection.

Chapter 6

References

Akhtar, N., Joshi, A., Singh, J., & Kaushik, V. (2021). Design of a novel and potent multivalent epitope based human cytomegalovirus peptide vaccine: An immunoinformatics approach. *Journal of Molecular Liquids*, *335*

Ayyagari, V. S., Venkateswarulu, T. C., Abraham Peele, K., & Srirama, K. (2022). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, *40*

Bakkari, M. A. (2023). Targeted Protein-Specific Multi-Epitope-Based Vaccine Designing against Human Cytomegalovirus by Using Immunoinformatics Approaches. *Vaccines*, *11*

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., … Zhang, J. (2015). UniProt: A hub for protein information. *Nucleic Acids Research*, *43*

Bhattacharjee, M., Banerjee, M., & Mukherjee, A. (2023). In silico designing of a novel polyvalent multi-subunit peptide vaccine leveraging cross-immunity against human visceral and cutaneous leishmaniasis: an immunoinformatics-based approach. *Journal of Molecular Modeling*, *29*

Chauhan, V., & Singh, M. P. (2020). Immuno-informatics approach to design a multi-epitope vaccine to combat cytomegalovirus infection. *European Journal of Pharmaceutical Sciences*, *147*.

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*, *2013*, 263952.

Gibson, W., Murphy, E., Shenk, T., Ruzsics, Z., Koszinowski, U., Kalejta, R., & Pari, G. S. (1993). *Structure and Formation of the Cytomegalovirus Virion*.

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 and Food Research*, *60*

Gopalakrishnan, K., Sowmiya, G., Sheik, S. S., & Sekar, K. (2007a). Ramachandran Plot on The Web (2.0). In *Protein & Peptide Letters* (Vol. 14).

Gopalakrishnan, K., Sowmiya, G., Sheik, S. S., & Sekar, K. (2007b). Ramachandran Plot on The Web (2.0). In *Protein & Peptide Letters* (Vol. 14).

Griffiths, P., Baraniak, I., & Reeves, M. (2015). The pathogenesis of human cytomegalovirus. *Journal of Pathology*, *235*

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, *30*

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*

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. *Nature Protocols*, *10*

Landais, I., & Nelson, J. A. (2013). Functional genomics approaches to understand cytomegalovirus replication, latency and pathogenesis. In *Current Opinion in Virology* (Vol. 3, Issue 4, pp. 408–415).

Landolfo, S., Gariglio, M., Gribaudo, G., & Lembo, D. (2003). The human cytomegalovirus. In *Pharmacology and Therapeutics* (Vol. 98).

Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A. C., Neveu, V., & Wishart, D. S. (2009). T3DB: A comprehensively annotated database of common toxins and their targets. *Nucleic Acids Research*, *38*

Lin, H. H., Ray, S., Tongchusak, S., Reinherz, E. L., & Brusic, V. (2008). Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunology*, *9*.

Lundegaard, C., Lund, O., Buus, S., & Nielsen, M. (2010). Major histocompatibility complex class i binding predictions as a tool in epitope discovery. In *Immunology* (Vol. 130, Issue 3, pp. 309–318).

Murphy, E., & Shenk, T. (n.d.). *Human Cytomegalovirus Genome*.

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*.

Narula, A., Pandey, R. K., Khatoon, N., Mishra, A., & Prajapati, V. K. (2018). Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infection, Genetics and Evolution*, *61*

Plosa, E. J., Esbenshade, J. C., Fuller, M. P., & Weitkamp, J. H. (2012). Cytomegalovirus infection. In *Pediatrics in Review* (Vol. 33, Issue 4, pp. 156–163).

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*

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2021a). 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*

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2021b). 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*

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., Peters, B., Nielsen, M., & Nielsen, M. (2020a). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *Journal of Proteome Research*, *19*

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., Peters, B., Nielsen, M., & Nielsen, M. (2020b). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *Journal of Proteome Research*, *19*

Ruzsics, Z., & Koszinowski, U. H. (n.d.). *Mutagenesis of the Cytomegalovirus Genome*.

Sabbaghian, E., Roodbari, F., Rafiei, A., & Amani, J. (2014). In Silico Design of a Multimeric Polytope as a Highly Immunogenic DNA Vaccine Against Human Cytomegalovirus. In *Journal of Applied Biotechnology Reports Original Article Journal of Applied Biotechnology Reports* (Vol. 1, Issue 4).

Saha, R., & L S Prasad, B. V. (n.d.). *In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus 1 (SARS-COV-2)*.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005a). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005b). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*

Sippl, M. J. (1993). Recognition of Errors in Three-Dimensional Structures of Proteins. In *Structure, Function, and Genetics* (Vol. 17).

Spaete, R. R., Gehrz, R. C., & Landini, M. P. (1994). Human cytomegalovirus structural proteins. In *Journal of General Virology* (Vol. 75).

Subramanian Iyer, S., & Cheng, G. (2012). Role of Interleukin 10 Transcriptional Regulation in Inflammation and Autoimmune Disease. In *Critical Reviews ™ in Immunology* (Vol. 32, Issue 1).

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (n.d.-a). *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Edited Protein Identification and Analysis Tools on the ExPASy Server*.

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (n.d.-b). *Protein Analysis Tools on the ExPASy Server 571 571 From:*

The Proteomics Protocols Handbook Edited Protein Identification and Analysis Tools on the ExPASy Server.

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*

Zatsepin, M., Mattes, A., Rupp, S., Finkelmeier, D., Basu, A., Burger-Kentischer, A., & Goldblum, A. (2016). Computational Discovery and Experimental Confirmation of TLR9 Receptor Antagonist Leads. *Journal of Chemical Information and Modeling*, *56*