In-Silico Design of an E2 Envelope Glycoprotein Targeted Multi-Epitope Vaccine against Chikungunya Virus (CHIKV)

By

Intizar Ahmed
ID: 18146023

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
December 2022

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing a 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 of the main sources of help.

Student's Full Name & Signature:

INTIZAR AHMED

Student ID: 18146023

Approval

The thesis titled “In-Silico Design of an E2 Envelope Glycoprotein Targeted Multi-Epitope Vaccine against Chikungunya Virus (CHIKV)” submitted by Intizar Ahmed (ID: 18146023), of Summer, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

_____

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
Brac University

Approved By:

Program Director:

_____

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
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.

Acknowledgments

I am immensely grateful to the following people for guiding me through this project. I would like to thank wholeheartedly my respected project supervisor- Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University, who advised me throughout the project and provided me with valuable feedback. I am grateful to him for making me familiar with the new skills needed for conducting this research and for always guiding me. I am thankful towards Professor Dr. Eva Rahman Kabir, Dean, School of Pharmacy, Brac University.

Intizar Ahmed
ID: 18146023
School of Pharmacy
Brac University

Abstract

The mosquito-borne chikungunya virus has caused numerous outbreaks globally over the years and its transmission is expeditious. Due to the lack of an effective vaccine, chikungunya infection prevention mainly focuses on preventing mosquito bites. Chikungunya infection causes debilitating symptoms. An effective vaccine can provide stronger protection in future outbreaks. In this study, in-silico approach was taken to construct a multi-epitope vaccine against the chikungunya virus and in-silico biochemical analysis of the designed vaccine was performed. The E2 envelope glycoprotein of chikungunya virus, collected from Vipr database, was selected as primary protein. Its antigenicity (0.5679) was found using Vaxijen v2.0. Cytotoxic T lymphocyte epitopes, helper T lymphocyte epitopes, and B cell epitopes were identified using NetCTL-1.2, NetMHCIIpan 4.0, and Bepipred servers respectively. The helper T lymphocyte epitopes were screened using IFNepitope, IL-4Pred, and IL-10Pred. Linkers were used according to literature studies in order to construct the vaccine. Biochemical analysis of the final vaccine represented promising results. ProtParam was used to predict the instability index (35.68), grand average of hydropathy (-0.553), and molecular weight (63.72038 kDa) of the final vaccine. AllergenOnline and T3DB predicted the vaccine as non-allergen and non-toxic respectively. ProSAweb was applied to assess Z-score (-5.16) and SWISS-MODEL generated Ramachandran plots. C-IMMSIM predicted desirable responses of immune cells and antibodies in accordance with three doses of the vaccine. As this study was based on in-silico computational methods, future investigations incorporating in vitro and in vivo methods are needed to validate the safety and efficacy of the constructed chikungunya vaccine.

Keywords: Chikungunya virus, multi-epitope vaccine, in-silico, E2 envelope glycoprotein, biochemical analysis.

Table of Contents

List of Figures

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

(e) Number of CD4 T helper lymphocytes

(f) Entity state of CD4 Helper T cells

(g) Number of CD8 T cytotoxic (TC) cells

(h) Number of CD8 T cytotoxic (TC) cells per entity state

(i) Total number of dendritic cells (DC) and their states

(j) Levels of cytokines including IFN-γ, IL-4, and IL-10

List of Tables

Chapter 1

Introduction

Chikungunya is a debilitating viral disease caused by infection with the Chikungunya virus (CHIKV). Chikungunya virus is a mosquito-borne pathogen. Transmission of CHIKV occurs via *Aedes* spp. mosquitoes infected with the virus. Nonhuman primates act as the primary reservoir hosts for CHIKV. The origin of CHIKV is presumably Central and East Africa (McFee, 2018; Vairo et al., 2019) The virus has re-emerged in epidemic form in recent years and posed a significant threat to public health in particular regions of the world. In 2008, the first chikungunya outbreak in Bangladesh affecting 39 people was identified in Chapainawabganj and Rajshahi by the Institute of Epidemiology, Disease Control and Research (IEDCR) and International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) (ICDDR,B, 2009). After that, numerous chikungunya outbreaks have occurred in Bangladesh from 2012 to 2017 (Haque et al., 2019). However, the most threatening outbreak in Bangladesh occurred around April to September of 2017. It affected 23 districts and more than 13000 cases were reported from the capital Dhaka (Anwar et al., 2020). Although chikungunya is not a fatal disease, it can cause significant health problems especially for patients with comorbidities. Its symptoms including- combination of fever and serious arthralgia, weakness, swelling in joints, sleep disturbance, memory problems etc. can reduce quality of life drastically (Centers for Disease Control and Prevention [CDC], 2022). At present there are no vaccines for chikungunya fever. The preventative measures against this disease are based on prevention of mosquito bites. Thus the preventative actions are limited to using insect repellents, mosquito sprays, mosquito nets, wearing long-sleeved clothings, scrutinizing water-retaining containers etc (Anwar et al., 2020). However, controlling the vector population may not be sufficient to prevent disease transmission and protect public health if there is a massive outbreak of CHIKV again in the future. Thus, in order to prepare for managing large-scale outbreaks of chikungunya in future, effective vaccines against the virus are needed.

1.1 Structural and genomic features of Chikungunya Virus:

Chikungunya virus belongs to one of the genus of the Togaviridae family named Alphavirus. The size of the virus varies between 50-70 nm. Electron microscopy shows that it contains an

icosahedral nucleocapsid which is surrounded via the viral envelope. The nucleocapsid core has 240 copies of the capsid proteins and is surrounded with a lipid bilayer which is host-derived. The nucleocapsid constitutes a single-stranded RNA genome. The RNA is positive-strand, non-segmented, about 11.8 kb with a terminal methylguanylate cap in the 5' end and polyadenylation in the 3' end (Cavrini et al., 2009). It comprises two open reading frames located between the untranslated regions (NTR) 30 NTR and 50 NTR. The open reading frame at 50 end of the genome codes for the polyprotein precursor which creates the nonstructural proteins nsP1, nsP2, nsP3, nsP4. These proteins have proteolytic and/or replicative functions. The other open reading frame codes for the polyprotein precursor which creates the structural proteins E1, E2, C necessary for nucleocapsid and envelope formation. Molecular mass of Both E1 and E2 proteins are approximately 50 kD which create stable heterodimer. Viral surface spike is formed by the interaction of three E2-El heterodimers (Chevillon et al., 2008).

1.2 Viral replication and pathogenesis of Chikungunya Virus:

In mammalian cells, the E2 protein of CHIKV binds to the surface of the host cell using receptors and attachment factors. The entry of CHIKV inside the cell occurs via clathrin-mediated endocytosis. Then there is endosomal acidification that causes the fusion peptide of E1 protein of the virus to insert in the endosomal membrane. The nucleocapsid is released in the cytosol due to fusion between endosomal membrane and viral envelope. As the nucleocapsid disassembles, the positive-sense RNA of CHIKV becomes free (Silva & Dermody, 2017). Translation machinery of the host cell translates this viral genome and produces nonstructural polyprotein P1234. P1234 is broken down to precursor P123 and the nonstructural-protein nsP4 (Solignat et al., 2009). Replication complexes (RCs) are created by P123 and nsP4 (Van der Heijden & Bol, 2014). These replication complexes create negative-sense RNA as a template for synthesizing the positive-sense genomic RNA and sub-genomic RNA. This sub-genomic 26S RNA works as the mRNA that codes for viral structural proteins C-pE2-6K-E1. From this, the capsid protein C is liberated first due to its autocleavage function. This capsid protein then binds to newly formed viral RNA to start creating the nucleocapsid core. In the endoplasmic reticulum, the remainder precursor pE2-6K-E1 becomes processed to pE2, 6K, and E1. Heterodimer complexes are created by E1 and pE2, which use the Golgi secretory pathway to reach the cell membrane. Here host enzyme furin and furin-like proteases break down pE2 into

E2 and E3. E2 and E3 enclose the remainder of the virus genome. Lastly, the nucleocapsid core migrates to the host cell surface and combines with E1/E2 dimers. The new virus buds and exits from the cell and a new replication cycle is started (Abdelnabi et al., 2015; Battisti et al., 2021).

Viraemia occurs in the initial phase of infection which can be identified with real-time PCR. High levels of chemokines and cytokines including interferon-α, interferon-γ, interleukin-4, interleukin-6, interleukin-10 etc. are observed in patients. IgM antibodies are found within days after infection and anti-chikungunya IgG are found near the second week (Burt et al., 2017)**.** Different cells and tissues of the host that support the growth of CHIKV include- fibroblasts, brain cells, endothelial cells etc. Pathogenesis of CHIKV focuses on the ability of the virus to replicate in muscle tissues and joints. The virus affects mesenchymal stem cells of bone marrow, osteoblasts and chondrocytes in the joints. Joint pain and arthritis occurs due to cellular damage caused by CHIKV infection. Bone homeostasis is affected in patients since repair capacity and mineralization processes are hampered (Kril et al., 2021). Some patients experience chronic arthritis, the causes of which still remain elusive. It is assumed that persistence of virus or its nucleic acids and continued immune activation could cause such immunopathology even after infection has subsided (Labadie et al., 2010).

Vaccines play a significant role in cost-effective prevention of infectious disease. Similar to other vaccines, a potential vaccine candidate for CHIKV will need to have a substantial balance of safety and immunogenicity (Mahalingam et al., 2017). Different types of CHIKV vaccines have been experimented with virus-like particles (VLP), inactivated virions, subunit vaccines, live attenuated vaccines etc. (DeFilippis, 2022).

Chapter 2

Methodology

The step-by-step methods which were utilized in this in-silico vaccine design for the chikungunya virus are illustrated in Figure 1.

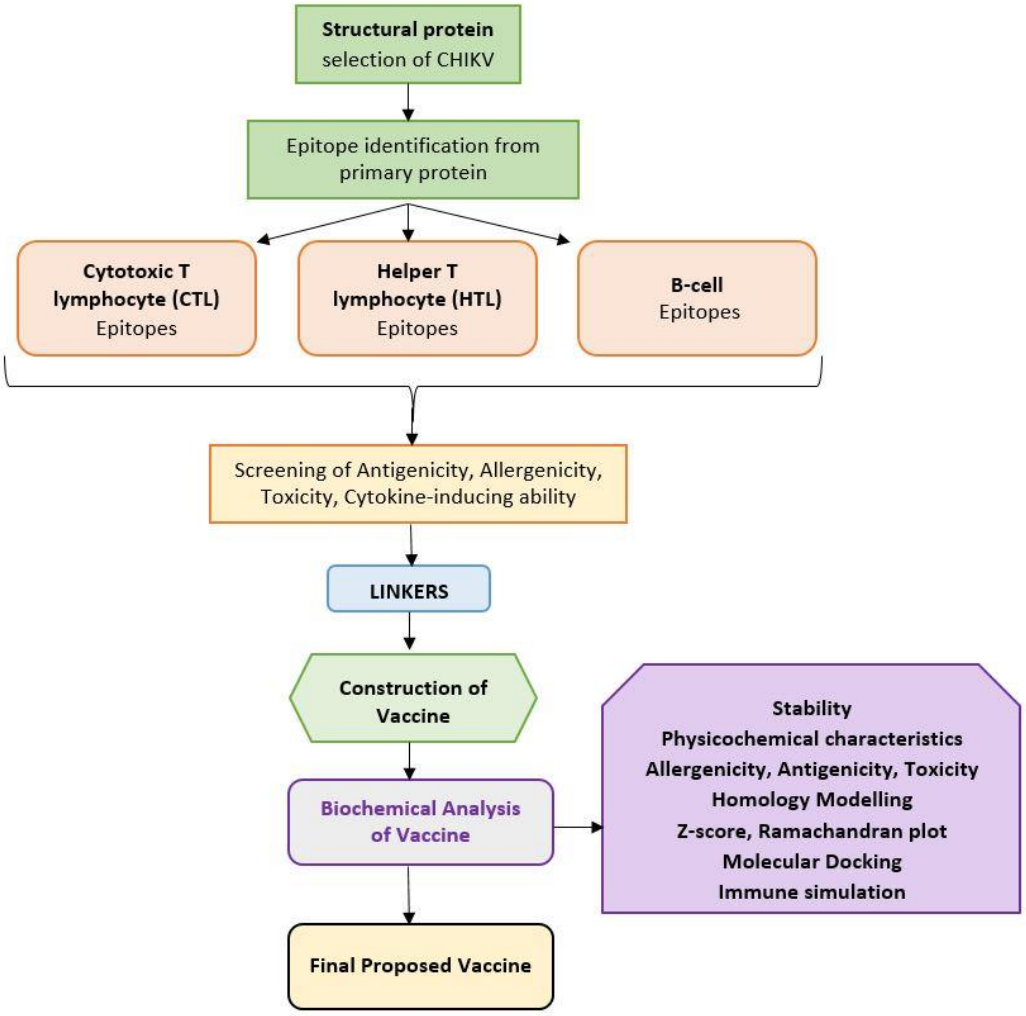


Figure 1: Step-by-step methods used in in-silico vaccine design for CHIKV.

2.1 Selection and Collection of Envelope Protein Sequence of Chikungunya Virus:

The Virus Pathogen Resource (ViPR) server (viprbrc.org) was used to collect different protein sequences of CHIKV strains. The ViPR server is a widely used computational tool used in structure, sequence and comparative genomic analysis of viruses (Pickett et al., 2012). In the ViPR server, there is a virus taxonomy browser which lists different virus families. Under the

Togaviridae family, the genus Alphavirus was selected. Then under the Alphavirus genus, the virus Chikungunya was selected. The server contained 5926 strains and 960 complete genomes for CHIKV. By observing the 'country' column, several strains for CHIKV were selected. Those included one strain from Kenya, one strain from India, and two strains from Bangladesh. Three proteins were selected from the Kenya strain named structural polyprotein, C and 6k protein. The E2 protein was selected from the strain '02 TANUVAS' from India. E1 protein was selected from 'CHIK-BD1052' Bangladesh strain. Six proteins were selected from Strain Bangladesh 2017 named- structural polyprotein, C, E1, E2, E3, 6K protein. The sequences of all these proteins were collected from the server in fasta format and compiled. ViPR has been used previously in research relating to CHIKV. For example, a group of researchers completed a study in 2020 involving the rise of ECSA Lineage of Chikungunya virus near the south border of the Amazon Forest. In their study, they collected all genomic data from the ViPR database (Da Silva Pessoa Vieira et al., 2020). Vaxijen 2.0 was used to predict the antigenicity of the collected protein sequences. In the Vaxijen server, prediction of protective antigens is independent of sequence alignment (Doytchinova & Flower, 2007). Virus was selected as the target organism in the server. The threshold was set to 0.5. The proteins were checked whether these were probable antigen or non-antigen. The overall predictions of antigenicity were noted and compared to screen the protein with the highest antigenicity.

2.2 Identifying Cytotoxic T Lymphocyte (CTL) Epitopes:

NetCTL-1.2 server was used to predict cytotoxic T lymphocyte (CTL) epitopes for the selected primary protein sequence. The CTL epitope predictions in this server are confined to the 12 major histocompatibility complex (MHC) class I supertype. This server utilizes artificial neural networks for predicting MHC class I binding and proteasomal cleavage (Larsen et al., 2007). The selected primary protein sequence was given in FASTA format. The threshold set for epitope identification was 0.75. Weight of C terminal cleavage and weight of TAP transport efficiency were set to default values since on average, default values provide optimum predictive performance. 'Combined score' was selected in the 'sort by score' section which lists CTL epitopes according to descending order of prediction scores on the result page. Determining MHC-binding peptides and T-cell epitopes play significant roles in the design of vaccines (Lin et al., 2008). The NetCTL-1.2 server has been used in other studies for the in-silico CTL epitope prediction of CHIKV strains (Kori et

al., 2015). The server ToxinPred was used to predict toxic or non-toxic peptides among the identified CTL epitopes (Gupta et al., 2013).

2.3 Determination of MHC class I Alleles:

NetMHCpan- 4.1 server was used for the prediction of MHC I alleles specific to the CTL epitopes. This server applies artificial neural networks (ANNs) to forecast peptide binding with MHC I molecules. The peptide length input given was 9 mer peptides. The threshold for strong binder and weak binder was 0.5 and 2 respectively. From the 'select allele(s)' list, all MHC molecules shown were selected. The options for including BA (binding Affinity) and sorting by prediction scores were selected in the server. The alleles and corresponding peptides were noted on the results page. Only the peptides with strong binding (bind level: SB) were selected (Reynisson, Alvarez, et al., 2020).

2.4 Identifying Helper T Lymphocyte (HTL) Epitopes:

The server NetMHCIIpan 4.0 was used to identify helper T lymphocyte (HTL) epitopes. This server utilizes artificial neural networks to forecast the peptide binding of a known sequence to the MHC II molecule (Reynisson, Barra, et al., 2020). The selected primary protein sequence was inputted in FASTA format on the server. The threshold for strong binder and threshold for weak binder were 1 and 5 respectively. Maximum of 20 alleles were selected from the 'select allele(s)' list in each submission. Only the peptides having strong binding (SB) bind level were collected. The process was repeated until all alleles were selected. All strong binding HTLs were compiled in a Microsoft Excel sheet. The repeating SB HTL sequences were omitted from the sheet.

2.5 Cytokine Stimulating Ability of Strong Binding HTL Epitopes:

Different cytokines including- interferon-gamma, interleukin 4, and interleukin 10 are released by helper T lymphocytes which trigger immune cells in our body. These cytokines survive following inflammatory responses and prevent tissue damage. Thus, induction of such cytokine release by HTL epitopes is desirable for designing vaccines (Bhuiyan et al., 2020). Hence, all the identified HTL epitopes from the previous step were tested for three criteria initially. IFN server was utilized in order to predict whether the HTLs were interferon-gamma inducing or not. The model selected in this server was IFN gamma vs. non-IFN gamma. It is important to find out MHC class II binders

that are able to stimulate IFN gamma-inducing T-helper cells for the in-silico design of subunit vaccines. IFN gamma is a Th1 cytokine that promotes cell-mediated immunity for the elimination of intracellular pathogens (Dhanda, Vir, et al., 2013). After that, the IL-4pred server was used to predict whether the HTL epitopes were interleukin-4 (IL-4) inducing or not. The hybrid (SVM+motif) model was chosen in the server. IL-4 inducibility of HTL epitopes is considered since IL-4 plays important functions including IgE production, antibody isotype switching, and proliferation of antigen-presenting cells (Dhanda, Gupta, et al., 2013). IL-10Pred server was utilized to predict whether the HTL epitopes were interleukin-10 (IL-10) inducing or not (Nagpal et al., 2017).

2.6 Determination of B-cell Epitopes:

IEDB Analysis Resource was used as an online computational tool for B cell epitope predictions. The method chosen in the server was 'bepipred linear epitope prediction 2.0'. Upon submission of the primary protein sequence in this server, the results page generates a table of predicted peptides as B cell epitopes and its corresponding graph. The threshold value was 0.5 (Jespersen et al., 2017).

2.7 Assembly of Vaccine Candidates:

The primary protein sequence, CTL epitopes, HTL epitopes, and B cell epitopes were connected using appropriate linkers in order to construct a multi-epitope vaccine. The first vaccine construct was assembled by taking all CTL epitopes, HTL, and B cell epitopes screened from previous steps. Subsequently, several combinations of CTL, HTL, and B cell epitopes were assembled and tested through biochemical analysis to compare and contrast different variations of the vaccine construct. Linkers are needed for effectively separating epitopes and increasing the presentation of epitopes in host cells (Narula et al., 2018). Four types of peptide linkers- EAAAK, AAY, GPGPG, and KK were used in all vaccine constructs. The EAAAK linker was placed to connect the primary protein sequence and first CTL epitope. AAY, GPGPG, and KK linkers were placed between adjacent CTL, HTL, and B cell epitopes respectively.

2.8 In-silico Biochemical Analysis of Candidate Vaccines:

In-silico biochemical analysis was performed for all the vaccine constructs. The first step in biochemical analysis was to check the stability of the vaccine construct via the ProtParam tool.

Upon submission of the vaccine construct sequence, the ProtParam server shows the total amino acid number, amino acid composition, atomic composition, theoretical pI, molecular weight, grand average of hydropathicity (GRAVY) etc. for the submitted sequence. Thus, it is an excellent server to compare various attributes of the vaccine. The instability index indicates the stability of the constructed vaccine (Gasteiger et al., 2005; Tahir Ul Qamar et al., 2018).

2.9 Assessment of Toxicity and Allergenicity:

The Toxin and Toxin Target Database or T3DB is a comprehensive database of toxin and toxin targets that was applied to search whether the constructed vaccines were toxin or non-toxin. The BLAST parameters in the server were set to default settings and the vaccine sequences were submitted in FASTA format (Lim et al., 2010; Wishart et al., 2015).

The AllergenOnline server (version 21) was used to predict the allergenicity of the constructed vaccines. This server utilizes bioinformatics methods like FASTA or BLASTP local alignment tools in order to compare submitted protein sequences and known allergens. 'Show Z-score' was selected in the search option on the server (Goodman et al., 2016).

2.10 Three-Dimensional Model Creation via Homology Modeling:

Phyre2 server was operated to depict the three-dimensional or 3D structures of the constructed vaccines. The Phyre 2 server constructs 3D models and identifies ligand binding sites by following a remote homology detection approach. It helps to elucidate the secondary and tertiary structure of the model and shows the quality of the model including the coverage and confidence (Kelley et al., 2015). In Phyre 2, the 'Normal modeling' mode was chosen. This server provided a PDB file for each vaccine via the given email. The software Discovery Studio Visualizer was used for opening the PDB files and for interactively viewing the 3D structures.

2.11 Generating Ramachandran Plots and Z-score Evaluation of Vaccine:

In the SWISS-MODEL server, the structure assessment portal was used to generate Ramachandran plots and associated information for the constructed vaccines (Guex et al., 2009). The PDB file obtained in the previous step was submitted here. Ramachandran plot helps to assess the stereochemical quality of a protein model. Conformation angles- phi (φ) and psi (ψ), also named Ramachandran angles are plotted against each other in the Ramachandran plot. These angles

influence the polypeptide chain backbone of a particular protein (K. Gopalakrishnan et al., 2007). Then, the website ProSA-web was used for determining the z-score for the constructed vaccines. The PDB file was inputted to the server. The results page showed the z-score value, a z-score vs. number of residues graph, a knowledge-based energy vs. sequence position graph (Sippl, 1993; Wiederstein & Sippl, 2007).

2.12 Molecular Docking of Vaccine with Relevant Human Receptors:

In order to ensure binding affinity between the constructed vaccine and desired receptors, the molecular-docking algorithm PATCHDOCK was applied (Schneidman-Duhovny et al., 2005). The constructed vaccine was inputted as the ligand molecule, and 3W3G was inputted as the receptor molecule in patchdock. 3W3G symbolizes human Toll-like receptor 8 (TLR8). On the results page, the PDB file of the highest scoring solution was downloaded.

2.13 Immune Response Simulations:

The C-ImmSim model gives a representation of the humoral and cellular response of the immune system with respect to antigens. Immune simulation with this server generates several graphs to predict levels of antigen, immunoglobulins, lymphocytes, and other immune cells with respect to time following the vaccination doses (Castiglione et al., 2021; Rapin et al., 2010). In the C-ImmSim server, the vaccine sequence was inputted and three injections were added. The time steps of injection numbers 1, 2, and 3 were 1, 84, and 168 respectively. Each time step represents specific hours of time. The dosage regimen for the vaccination was assigned to be three injections at intervals of 28 days.

2.14 Remarks regarding Methodology:

The steps in chikungunya vaccine construction and subsequent biochemical analysis completed in this study were performed in an in-silico approach. Therefore, the tools used in the experimentation were online servers which are widely used in the field of in-silico vaccine design research. Besides these online servers, a comprehensive literature review was done wherever necessary to validate the methods used in this study.

Chapter 3

Results

3.1 Antigenicity of Selected Protein:

Among the screened primary proteins, the E2 protein of 'Chikungunya Strain Bangladesh 2017' showed the highest antigenicity in VaxiJen v2.0 server (Doytchinova & Flower, 2007). The GenBank Sequence Accession of this strain was MF773566. The sequence of the E2 protein is given below-

STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDD
NHDWTKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDS
RKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRT
LMSQQSGNVKITVNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQ
YNSPLVPRNAELGDRKGKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLS
YRNMGEEPNYQEEWVMHKKEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHP
HEIILYYYELYPTMTVVVVSVATFILLSMVGMAAGMCMCARRRCITPYELTPGATVPFL
LSLICCIRTAKA

The antigenicity of this sequence was predicted to be 0.5679 (Probable ANTIGEN) in VaxiJen v2.0 as shown in figure 2.

Figure 2: Antigenicity of the selected E2 protein of 'Chikungunya Strain Bangladesh 2017' (Doytchinova & Flower, 2007).

Besides high antigenicity, the E2 protein was selected as the primary protein sequence due to a number of reasons. The E2 protein has relations to viral replication, pathogenicity, monoclonal

antibody binding, and cell tropism. Each of its domains contributes to the life cycle of CHIKV (Weger-Lucarelli et al., 2016).

3.2 Identifying CTL Epitopes:

In the NetCTL-1.2 server, 11 epitopes were found for the primary protein sequence by using MHC supertype A1 and a threshold of 0.75 as shown in figure 3 (Larsen et al., 2007). The combined scores of these CTL epitopes are presented in table 1.

Table 1: CTL epitopes and their respective combined scores

CTL epitopes	Combined Score
VTNHKKWQY	2.7636
STKDNFNVY	2.1435
FTDSRKISH	1.4179
TVNGQTVRY	1.3134
VTWGNNEPY	1.2063
IILYYYELY	1.0370
NHDWTKLRY	0.9362
STAATTEEI	0.8655
KNQVIMLLY	0.8541
TTDKVINNC	0.8219
HPHEIILYY	0.7537

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

229 ID FASTA pep VTNHKKWQY aff   0.5847 aff_rescale   2.4827 cle 0.8643 tap   3.0250 COMB   2.7636 <-E
  1 ID FASTA pep STKDNFNVY aff   0.4356 aff_rescale   1.8493 cle 0.9772 tap   2.9520 COMB   2.1435 <-E
115 ID FASTA pep FTDSRKISH aff   0.3322 aff_rescale   1.4104 cle 0.3355 tap  -0.8580 COMB   1.4179 <-E
191 ID FASTA pep TVNGQTVRY aff   0.2399 aff_rescale   1.0188 cle 0.9784 tap   2.9570 COMB   1.3134 <-E
328 ID FASTA pep VTWGNNEPY aff   0.2147 aff_rescale   0.9117 cle 0.9741 tap   2.9700 COMB   1.2063 <-E
355 ID FASTA pep IILYYYELY aff   0.2035 aff_rescale   0.8639 cle 0.1786 tap   2.9270 COMB   1.0370 <-E
 61 ID FASTA pep NHDWTKLRY aff   0.1557 aff_rescale   0.6610 cle 0.9503 tap   2.6540 COMB   0.9362 <-E
159 ID FASTA pep STAATTEEI aff   0.1666 aff_rescale   0.7072 cle 0.8775 tap   0.5340 COMB   0.8655 <-E
280 ID FASTA pep KNQVIMLLY aff   0.1497 aff_rescale   0.6354 cle 0.5186 tap   2.8170 COMB   0.8541 <-E
212 ID FASTA pep TTDKVINNC aff   0.1915 aff_rescale   0.8131 cle 0.1236 tap  -0.1940 COMB   0.8219 <-E
351 ID FASTA pep HPHEIILYY aff   0.1132 aff_rescale   0.4807 cle 0.9768 tap   2.5290 COMB   0.7537 <-E
 57 ID FASTA pep KTDDNHDWT aff   0.1804 aff_rescale   0.7660 cle 0.0407 tap  -0.7870 COMB   0.7327
330 ID FASTA pep WGNNEPYKY aff   0.1035 aff_rescale   0.4393 cle 0.7313 tap   2.5470 COMB   0.6764
```

Figure 3: CTL epitopes found in NetCTL-1.2 server (Larsen et al., 2007).

3.3 Determination of MHC Class I Alleles Specific to CTL Epitopes:

NetMHCpan-4.1 was used for the identification of MHC I alleles specific to the previously found CTL epitopes (Reynisson et al., 2020). The peptide length was 9 for all epitopes. The threshold for

strong binders was 0.5 whereas the threshold for weak binders was 2. Alleles were found for 6 CTL epitopes among the 11 CTL epitopes. The strong binding alleles are tabulated in table 2.

Table 2: Strong binding MHC I alleles specific to CTL epitopes

Allele	Peptide	Sequence no.	%Rank_EL	%Rank_BA	Aff(nM)
1 HLA-A*01:01	VTNHKKWQY	1	0.154	0.487	1074.18
1 HLA-A*01:01	STKDNFNVY	2	0.232	0.705	1831.77
1 HLA-A*01:01	FTDSRKISH	3	0.349	0.61	1498.6
1 HLA-A*01:01	TVNGQTVRY	4	0.153	0.515	1165.32
1 HLA-A*01:01	VTWGNNEPY	5	0.415	0.353	682.8
1 HLA-A*26:01	STKDNFNVY	2	0.022	0.054	102.82
1 HLA-A*26:01	TVNGQTVRY	4	0.013	0.113	235.08
1 HLA-A*26:01	HPHEIILYY	11	0.088	0.507	1568.24
1 HLA-B*15:01	STKDNFNVY	2	0.062	0.321	57.8
1 HLA-B*15:01	TVNGQTVRY	4	0.165	1.945	622.57

3.4 In-silico Prediction of Antigenicity, Allergenicity, and Toxicity of CTL Epitopes:

In the ToxinPred predictions, it was seen that all the CTL epitopes were non-toxin as shown in figure 4 (Gupta et al., 2013).

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq1	VTNHKKWQY	-1.05	Non-Toxin	-0.35	-1.86	-0.19	2.50	1203.50
seq2	STKDNFNVY	-1.51	Non-Toxin	-0.26	-1.13	0.00	0.00	1087.27
seq3	FTDSRKISH	-1.03	Non-Toxin	-0.37	-1.12	0.49	1.50	1090.32
seq4	TVNGQTVRY	-0.85	Non-Toxin	-0.24	-0.69	-0.30	1.00	1037.27
seq5	VTWGNNEPY	-1.16	Non-Toxin	-0.12	-1.24	-0.47	-1.00	1079.26
seq6	IILYYYELY	-0.61	Non-Toxin	0.22	0.88	-1.49	-1.00	1252.61
seq7	NHDWTKLRY	-0.94	Non-Toxin	-0.43	-1.97	0.09	1.50	1232.49
seq8	STAATTEEI	-1.02	Non-Toxin	-0.09	-0.20	0.26	-2.00	922.07
seq9	KNQVIMLLY	-1.01	Non-Toxin	0.02	0.67	-0.79	1.00	1121.55
seq10	TTDKVINNC	-0.43	Non-Toxin	-0.24	-0.51	0.14	0.00	1007.25
seq11	HPHEIILYY	-0.22	Non-Toxin	0.06	-0.14	-0.89	0.00	1184.50

Figure 4: ToxinPred results for CTL epitopes (Gupta et al., 2013).

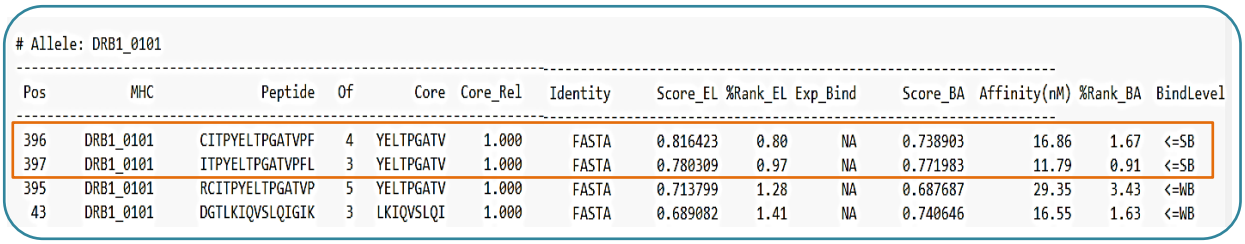
VaxiJen v2.0 server predicted two CTL epitopes as non-antigen whereas the other four epitopes were predicted to be probable antigens. AllerTOP v. 2.0 predicted three CTL epitopes to be probable allergens out of the six CTL epitopes. These attributes of selected CTL epitopes are summarized in table 3.

Table 3: Selected CTL epitopes: Summary of antigenicity, allergenicity, and toxicity

CTL epitopes	Antigenicity	Allergenicity	Toxicity
VTNHKKWQY	Antigen	Probable allergen	Non-toxin
STKDNFNVY	Antigen	Probable non-allergen	Non-toxin
FTDSRKISH	Antigen	Probable non-allergen	Non-toxin
TVNGQTVRY	Non-antigen	Probable allergen	Non-toxin
VTWGNNEPY	Antigen	Probable allergen	Non-toxin
HPHEIILYY	Non-antigen	Probable non-allergen	Non-toxin

3.5 Identification of Strong Binding HTL Epitopes:

In the NetMHCIIpan 4.0 Server, numerous peptide sequences were obtained along with the corresponding MHC II alleles (Reynisson et al., 2020). The core of the peptides, score_EL, and percentile rank were also shown on the results page. The first result page for allele DRB1_0101 in NetMHCIIpan 4.0 server is shown in figure 5. The output was according to the prediction score, which showed the strong binding peptides on top of the lists. A total of 62 strong binding (SB) HTL epitopes were found.



Allele: DRB1_0101

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
396	DRB1_0101	CITPYELTPGATVPF	4	YELTPGATV	1.000	FASTA	0.816423	0.80	NA	0.738903	16.86	1.67	<=SB
397	DRB1_0101	ITPYELTPGATVPFL	3	YELTPGATV	1.000	FASTA	0.780309	0.97	NA	0.771983	11.79	0.91	<=SB
395	DRB1_0101	RCITPYELTPGATVP	5	YELTPGATV	1.000	FASTA	0.713799	1.28	NA	0.687687	29.35	3.43	<=WB
43	DRB1_0101	DGTLKIQVSLQIGIK	3	LKIQVSLQI	1.000	FASTA	0.689082	1.41	NA	0.740646	16.55	1.63	<=WB

Figure 5: Strong and weak binding peptides for allele DRB1_0101 in NetMHCIIpan 4.0 Server (Reynisson, Alvarez, et al., 2020)

3.6 Cytokine Stimulating Ability of Obtained HTL Epitopes:

IL4pred, IL10pred and IFNepitope servers were used respectively to screen all strong binding HTL epitopes whether these fulfilled all three criteria of being inducers of cytokines IL-4, IL-10, and IFN-gamma. 7 such HTL epitopes were identified which are highlighted in yellow color as shown in table 4 and 5.

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

1	HTL peptide sequence	IFN	IL-4 prediction	IL-10 prediction
2	CITPYELTPGATVPF	NEGATIVE	Inducer	Non-inducer
3	ITPYELTPGATVPFL	POSITIVE	Non-inducer	Non-inducer
4	KIQVSLQIGIKTDDN	NEGATIVE	Inducer	Non-inducer
5	DGTLKIQVSLQIGIK	NEGATIVE	Inducer	Non-inducer
6	LKIQVSLQIGIKTDD	NEGATIVE	Inducer	Non-inducer
7	FNVYKATRPYLAHCP	NEGATIVE	Non-inducer	Non-inducer
8	DNFNVYKATRPYLAH	NEGATIVE	Non-inducer	Non-inducer
9	NSPLVPRNAELGDRK	NEGATIVE	Non-inducer	Non-inducer
10	RCITPYELTPGATVP	NEGATIVE	Inducer	Non-inducer
11	THPFHHDPPVIGREK	POSITIVE	Non-inducer	Non-inducer
12	CTHPFHHDPPVIGRE	POSITIVE	Non-inducer	Non-inducer
13	HPFHHDPPVIGREKF	POSITIVE	Non-inducer	Non-inducer
14	SCTHPFHHDPPVIGR	POSITIVE	Inducer	Non-inducer
15	PCSTYVQSTAATTEE	NEGATIVE	Non-inducer	Non-inducer
16	GNVKITVNGQTVRYK	NEGATIVE	Non-inducer	Non-inducer
17	LQIGIKTDDNHDWTK	NEGATIVE	Inducer	Non-inducer
18	NVKITVNGQTVRYKC	POSITIVE	Inducer	Non-inducer
19	SGNVKITVNGQTVRY	POSITIVE	Non-inducer	Non-inducer
20	IMLLYPDHPTLLSYR	POSITIVE	Inducer	Inducer
21	VIMLLYPDHPTLLSY	POSITIVE	Inducer	Inducer
22	PYKYWPQLSTNGTAH	NEGATIVE	Non-inducer	Non-inducer
23	WTKLRYMDNHMPADA	NEGATIVE	Inducer	Non-inducer
24	HPTLLSYRNMGEEPN	NEGATIVE	Inducer	Inducer
25	DWTKLRYMDNHMPAD	NEGATIVE	Inducer	Inducer
26	DHPTLLSYRNMGEEP	NEGATIVE	Inducer	Non-inducer
27	TKLRYMDNHMPADAE	NEGATIVE	Inducer	Non-inducer
28	PVALERIRNEATDGT	NEGATIVE	Inducer	Non-inducer
29	PTLLSYRNMGEEPNY	NEGATIVE	Inducer	Non-inducer
30	ALERIRNEATDGTLK	NEGATIVE	Inducer	Non-inducer
31	LERIRNEATDGTLKI	NEGATIVE	Inducer	Non-inducer
32	ERIRNEATDGTLKIQ	NEGATIVE	Inducer	Non-inducer

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

33	EPYKYWPQLSTNGTA	NEGATIVE	Inducer	Non-inducer
34	TLTVGFTDSRKISHS	POSITIVE	Inducer	Inducer
35	ETLTVGFTDSRKISH	NEGATIVE	Inducer	Inducer
36	TPYELTPGATVPFLL	POSITIVE	Non-inducer	Non-inducer
37	GETLTVGFTDSRKIS	POSITIVE	Inducer	Inducer
38	RRCITPYELTPGATV	NEGATIVE	Inducer	Non-inducer
39	NFNVYKATRPYLAHC	POSITIVE	Non-inducer	Non-inducer
40	NVYKATRPYLAHCPD	POSITIVE	Inducer	Inducer
41	LTVGFTDSRKISHSC	NEGATIVE	Inducer	Inducer
42	KDNFNVYKATRPYLA	NEGATIVE	Non-inducer	Non-inducer
43	DPPVIGREKFHSRPQ	POSITIVE	Inducer	Inducer
44	HDPPVIGREKFHSRP	NEGATIVE	Inducer	Inducer
45	STYVQSTAATTEEIE	NEGATIVE	Inducer	Non-inducer
46	CSTYVQSTAATTEEI	POSITIVE	Non-inducer	Non-inducer
47	LPCSTYVQSTAATTE	NEGATIVE	Non-inducer	Non-inducer
48	HAAVTNHKKWQYNSP	NEGATIVE	Inducer	Non-inducer
49	CHAAVTNHKKWQYNS	NEGATIVE	Inducer	Non-inducer
50	AAVTNHKKWQYNSPL	NEGATIVE	Inducer	Non-inducer
51	QCHAAVTNHKKWQYN	NEGATIVE	Inducer	Non-inducer
52	HHDPPVIGREKFHSR	NEGATIVE	Inducer	Inducer
53	PPVIGREKFHSRPQH	NEGATIVE	Inducer	Inducer
54	YQEEWVMHKKEVVLT	POSITIVE	Non-inducer	Non-inducer
55	QEEWVMHKKEVVLTV	NEGATIVE	Non-inducer	Non-inducer
56	NEPYKYWPQLSTNGT	NEGATIVE	Inducer	Non-inducer
57	YNSPLVPRNAELGDR	NEGATIVE	Non-inducer	Non-inducer
58	PDHPTLLSYRNMGEE	NEGATIVE	Inducer	Inducer
59	QVIMLLYPDHPTLLS	POSITIVE	Inducer	Inducer
60	NQVIMLLYPDHPTLL	POSITIVE	Inducer	Inducer
61	KNQVIMLLYPDHPTL	NEGATIVE	Inducer	Inducer
62	TKDNFNVYKATRPYL	NEGATIVE	Non-inducer	Non-inducer
63	STKDNFNVYKATRPY	NEGATIVE	Non-inducer	Non-inducer

3.7 Determination of B-cell Epitopes:

In IEDB Analysis Resource, the Bepipred Linear Epitope Prediction 2.0 method applied for the primary protein sequence gave the following result. A total of 14 peptides were found as B cell epitopes which are tabulated in table 6.

Table 6: B cell epitopes predicted via Bepipred Linear Epitope Prediction 2.0

No.	Start	End	Peptide	Length
1	6	17	FNVYKATRPYLA	12
2	20	29	PDCGEGHSCH	10
3	36	42	RIRNEAT	7
4	56	80	IKTDDNHDWTKLRYMDNHMPADAER	25
5	118	121	SRKI	4
6	131	165	HDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTE	35
7	172	180	PPDTPDRTL	9
8	208	212	EGLTT	5
9	214	215	DK	2
10	217	219	INN	3
11	234	252	KWQYNSPLVPRNAELGDRK	19
12	273	278	NPTVTY	6
13	299	314	NMGEEPNYQEEWVMHK	16
14	331	350	GNNEPYKYWPQLSTNGTAHG	20

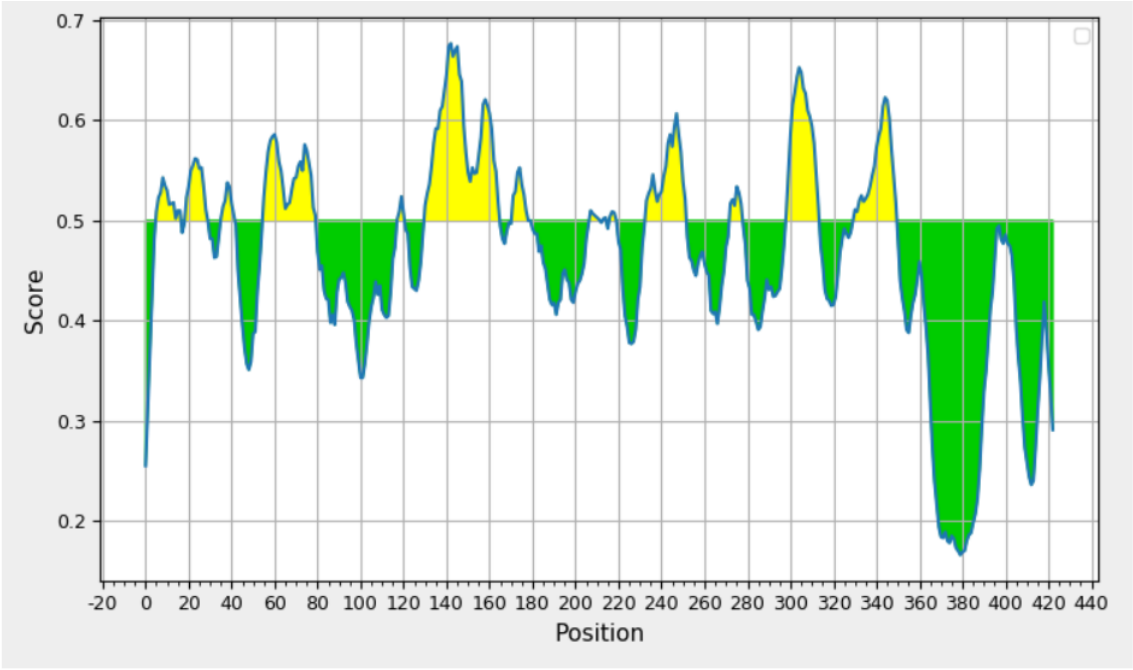


Figure 6: Score vs Position graph of B cell epitopes from IEDB server (Jespersen et al., 2017).

The average score of B cell epitopes was 0.466, the minimum score was 0.166 and the maximum score was 0.677 as shown in figure 6 (Jespersen et al., 2017).

3.8 Assembly of Vaccine Candidates:

The linkers mentioned in the methods section were used to connect the different types of epitopes. EAAAK linker ensures minimum interference between the epitopes and provides adequate segregation of functional domains. Inside mammalian cells, AAY or Ala-Ala-Tyr linker acts as a cleavage site for proteasomes. Hence, CTL epitopes connected by AAY will be efficiently detached inside cells. GPGPG linker stimulates T helper lymphocyte responses, lowers junctional immunogenicity, and thereby enhances the immunogenicity of each epitope. The di-lysine linker (KK) is targeted by the lysosomal protease cathepsin B. KK linker lowers junctional immunogenicity and enhances the immunogenicity of the vaccine. These linkers helped to connect all the selected epitopes and to design the proposed candidate vaccines. Incorporation of such linkers in these multi-epitope candidate vaccines will facilitate antigen processing and presentation, decrease the chances of forming junctional antigens. These linkers will also improve the structural flexibility, rigidity, and stability of the constructed vaccine (Ayyagari et al., 2020).

In this systematic process, 6 CTL epitopes, 7 HTL epitopes, and 14 B cell epitopes were found. Preliminary, the first vaccine was constructed incorporating all of these epitopes, excluding none. Subsequently, eight different vaccines were constructed by excluding specific CTL, HTL, and B cell epitopes in each vaccine based on different criteria which are named i-viii here. The goal of assembling the vaccines with variations in combinations was to assess how the different combinations influenced the biochemical analysis of the vaccines. All of the vaccine constructs are shown in table 6.

Vaccine (i): Vaccine A constructed incorporating all 6 CTL epitopes, 7 HTL epitopes, and 14 B cell epitopes.

Vaccine (ii): Vaccine constructed excluding toxin HTL epitopes (table 7).

Vaccine (iii): Vaccine construction taking HTL epitopes that are antigenic (table 7).

Table 7: HTL epitope antigenicity and toxicity predictions in VaxiJen v2.0 and ToxinPred respectively

HTLs	Antigenicity	Antigenicity %	Toxicity
VIMLLYPDHPTLLSY	NON-ANTIGEN		Non-toxin
TLTVGFTDSRKISHS	**ANTIGEN**	**1.1916**	Non-toxin
GETLTVGFTDSRKIS	**ANTIGEN**	**1.0040**	Non-toxin

HTLs	Antigenicity	Antigenicity %	Toxicity
NVYKATRPYLAHCPD	NON-ANTIGEN		Non-toxin
DPPVIGREKFHSRPQ	**ANTIGEN**	**1.2728**	Toxin
QVIMLLYPDHPTLLS	NON-ANTIGEN		Non-toxin
NQVIMLLYPDHPTLL	NON-ANTIGEN		Non-toxin

Vaccine (iv): Vaccine construction excluding overlapping HTL epitopes:

The following were the overlapping HTL epitopes-

VIMLLYPDHPTLLSY

QVIMLLYPDHPTLLS

NQVIMLLYPDHPTLL

Vaccine (v): Vaccine construction taking antigenic CTLs and antigenic HTLs (table 8).

Table 8: Antigenicity scores of CTL epitopes in VaxiJen v2.0

CTLs	Antigenicity	Antigenicity %
VTNHKKWQY	**ANTIGEN**	**1.0047**
STKDNFNVY	**ANTIGEN**	**0.6446**
FTDSRKISH	**ANTIGEN**	**0.6346**
TVNGQTVRY	NON-ANTIGEN	
VTWGNNEPY	**ANTIGEN**	**1.2738**
HPHEIILYY	NON-ANTIGEN	

Vaccine (vi): Vaccine construction taking antigenic CTL, antigenic HTL epitopes and excluding toxin HTL epitope.

Vaccine (vii): Taking Antigenic CTLs, Antigenic HTLs, excluding the non-toxin HTL epitope, and taking B cell epitopes length higher than 10, B cell epitopes length higher than 10 amino acids (table 9).

Table 9: B cell epitopes having length higher than 10 amino acids.

No.	Start	End	Peptide	Length	Antigen-icity %	Antigenicity
1	6	17	FNVYKATRPYLA	12	-0.1710	NON-ANTIGEN
2	20	29	PDCGEGHSCH	10	-0.3375	NON-ANTIGEN
4	56	80	IKTDDNHDWTKLRYMDNHMP ADAER	25	0.4977	NON-ANTIGEN
6	131	165	HDPPVIGREKFHSRPQHGKELP CSTYVQSTAATTE	35	0.6639	**Probable ANTIGEN**
11	234	252	KWQYNSPLVPRNAELGDRK	19	1.3857	**Probable ANTIGEN**
13	299	314	NMGEEPNYQEEWVMHK	16	0.4771	NON-ANTIGEN
14	331	350	GNNEPYKYWPQLSTNGTAHG	20	0.0261	NON-ANTIGEN

Vaccine (viii): Taking antigenicity score wise antigenic CTL, antigenic HTL, excluding non-toxin HTL epitope, taking B cell epitope length higher than 10, and antigenic B cell epitopes. According to the above-mentioned criteria, the sequences of different vaccine constructs are shown (table 10).

Table 10: 8 candidate vaccine constructs with varying combinations of CTL, HTL, and B cell epitopes with linkers.

Sl	Criteria	Sequence of Vaccine Construct
1	Vaccine (i): constructed incorporating all epitopes	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYTVNGQTVRYAAYVTWGNNEPYAAYHPHEIILYYGPGPGVI MLLYPDHPTLLSYGPGPGTLTVGFTDSRKISHSGPGPGGETLTVGFTDSRKISGPGPGNVYKAT RPYLAHCPDGPGPGDPPVIGREKFHSRPQGPGPGQVIMLLYPDHPTLLSGPGPGNQVIMLLYP DHPTLLKKFNVYKATRPYLAKKPDCGEGHSCHKKRIRNEATKKIKTDDNHDWTKLRYMDNHM PADAERKKSRKIKKHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEKKPPDTPDRTLKKEGL TTKKDKKKINNKKKWQYNSPLVPRNAELGDRKKKNPTVTYKKNMGEEPNYQEEWVMHKKKG NNEPYKYWPQLSTNGTAHG

Sl	Criteria	Sequence of Vaccine Construct
2	Vaccine (ii): constructed excluding the toxin HTL epitope	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYTVNGQTVRYAAYVTWGNNEPYAAYHPHEIILYYGPGPGVI MLLYPDHPTLLSYGPGPGTLTVGFTDSRKISHSGPGPGGETLTVGFTDSRKISGPGPGNVYKAT RPYLAHCPDGPGPGQVIMLLYPDHPTLLSGPGPGNQVIMLLYPDHPTLLKKFNVYKATRPYLAK KPDCGEGHSCHKKRIRNEATKKIKTDDNHDWTKLRYMDNHMPADAERKKSRKIKKHDPPVIG REKFHSRPQHGKELPCSTYVQSTAATTEKKPPDTPDRTLKKEGLTTKKDKKINNKKKWQYNSP LVPRNAELGDRKKKNPTVTYKKNMGEEPNYQEEWVMHKKKGNNEPYKYWPQLSTNGTAHG
3	Vaccine (iii): Taking HTL epitopes that are antigenic	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYTVNGQTVRYAAYVTWGNNEPYAAYHPHEIILYYGPGPGTL TVGFTDSRKISHSGPGPGGETLTVGFTDSRKISGPGPGDPPVIGREKFHSRPQKKFNVYKATRPY LAKKPDCGEGHSCHKKRIRNEATKKIKTDDNHDWTKLRYMDNHMPADAERKKSRKIKKHDPP VIGREKFHSRPQHGKELPCSTYVQSTAATTEKKPPDTPDRTLKKEGLTTKKDKKINNKKKWQY NSPLVPRNAELGDRKKKNPTVTYKKNMGEEPNYQEEWVMHKKKGNNEPYKYWPQLSTNGT AHG
4	Vaccine (iv): Excluding overlapping HTL epitopes	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYTVNGQTVRYAAYVTWGNNEPYAAYHPHEIILYYGPGPGTL TVGFTDSRKISHSGPGPGGETLTVGFTDSRKISGPGPGNVYKATRPYLAHCPDGPGPGDPPVIG REKFHSRPQKKFNVYKATRPYLAKKPDCGEGHSCHKKRIRNEATKKIKTDDNHDWTKLRYMD NHMPADAERKKSRKIKKHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEKKPPDTPDRTLK KEGLTTKKDKKINNKKKWQYNSPLVPRNAELGDRKKKNPTVTYKKNMGEEPNYQEEWVMH KKKGNNEPYKYWPQLSTNGTAHG
5	Vaccine (v): Taking Antigenic CTLs and Antigenic HTLs	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYVTWGNNEPYGPGPGTLTVGFTDSRKISHSGPGPGGETLTV GFTDSRKISGPGPGDPPVIGREKFHSRPQKKFNVYKATRPYLAKKPDCGEGHSCHKKRIRNEAT KKIKTDDNHDWTKLRYMDNHMPADAERKKSRKIKKHDPPVIGREKFHSRPQHGKELPCSTYV

		QSTAATTEKKPPDTPDRTLKKEGLTTKKDKKINNKKKWQYNSPLVPRNAELGDRKKNPTVTY KKNMGEEPNYQEEWVMHKKKGNNEPYKYWPQLSTNGTAHG
Sl	**Criteria**	**Sequence of Vaccine Construct**
6	Vaccine (vi): Taking Antigenic CTLs, Antigenic HTLs, Excluding toxin HTL	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYVTWGNNEPYGPGPGTLTVGFTDSRKISHSGPGPGGETLTV GFTDSRKISKKFNVYKATRPYLAKKPDCGEGHSCHKKRIRNEATKKIKTDDNHDWTKLRYMDN HMPADAERKKSRKIKKHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEKKPPDTPDRTLKK EGLTTKKDKKINNKKKWQYNSPLVPRNAELGDRKKNPTVTYKKNMGEEPNYQEEWVMHK KKGNNEPYKYWPQLSTNGTAHG
7	Vaccine (vii): Taking Antigenic CTLs, Antigenic HTLs, excluding the non-toxin HTL, B cell epitopes length> 10	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTNHKKWQYAAYS TKDNFNVYAAYFTDSRKISHAAYVTWGNNEPYGPGPGTLTVGFTDSRKISHSGPGPGGETLTV GFTDSRKISKKFNVYKATRPYLAKKPDCGEGHSCHKKIKTDDNHDWTKLRYMDNHMPADAER KKHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEKKKWQYNSPLVPRNAELGDRKKNM GEEPNYQEEWVMHKKKGNNEPYKYWPQLSTNGTAHG
8	Vaccine (viii): Taking score wise antigenic CTLs, antigenic HTLs, excluding non-toxin HTL, taking B cell epitope length> 10, antigenic B cell epitopes	STKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALERIRNEATDGTLKIQVSLQIGIKTDDNHDW TKLRYMDNHMPADAERAGLFVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHSCTHP FHHDPPVIGREKFHSRPQHGKELPCSTYVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKIT VNGQTVRYKCNCGSSNEGLTTTDKVINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK GKIHIPFPLANATCRVPKARNPTVTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVMHK KEVVLTVPTEGLEVTWGNNEPYKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVATFIL LSMVGMAAGMCMCARRRCITPYELTPGATVPFLLSLICCIRTAKAEAAAKVTWGNNEPYAAY VTNHKKWQYAAYSTKDNFNVYAAYFTDSRKISHGPGPGTLTVGFTDSRKISHSGPGPGGETLT VGFTDSRKISKKKWQYNSPLVPRNAELGDRKKKHDPPVIGREKFHSRPQHGKELPCSTYVQSTA ATTE

3.9 In-silico Biochemical Analysis of Candidate Vaccines:

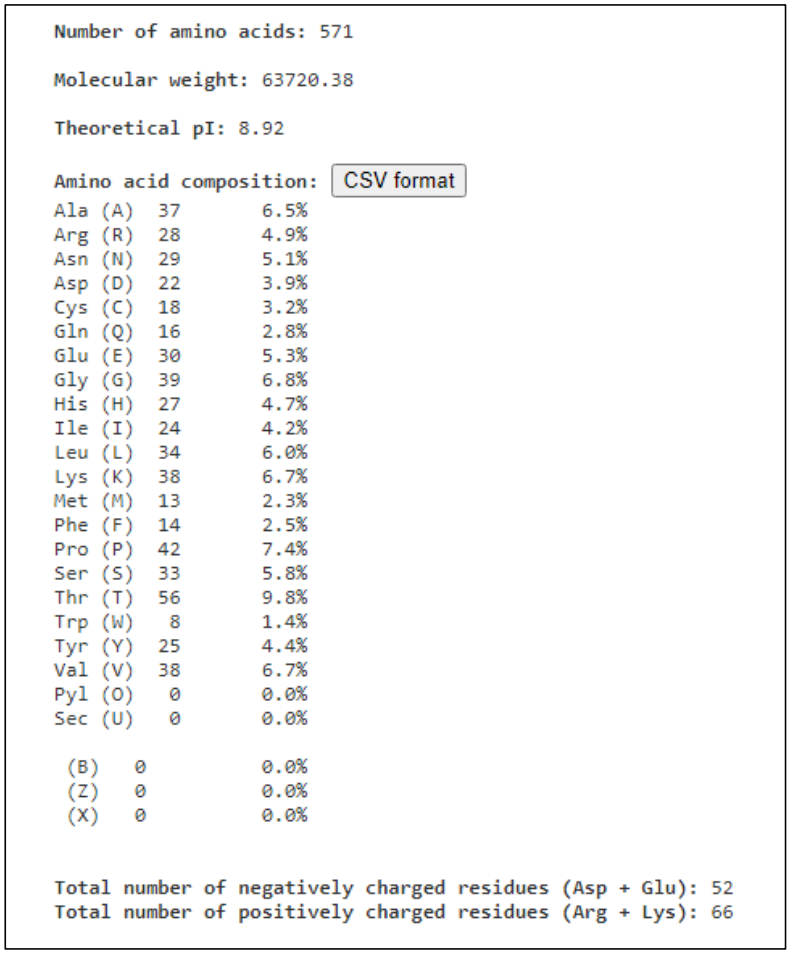
The biochemical analysis of the constructed vaccines (i-viii) showed differences in characteristics of the vaccine constructs. A summary and comparison of the biochemical analysis of each vaccine construct is shown in table 11.

Table 11: Comparison of biochemical analysis of all 8 vaccine candidates

	Vaccine (i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)
PROTPARAM- VACCINE STABILITY [primary protein instability index = 38.20]	Instability index 32.02 (stable), GRAVY: -0.709, M.W: 93926.03	Instability index 31.38 (stable), GRAVY: -0.695, M.W: 91815.67	Instability index 34.82 (stable), GRAVY: -0.783 M.W: 85506.21	Instability index 34.26 (stable), GRAVY: -0.783, M.W: 87601.57
VAXIJEN 2.0- ANTIGENICITY [primary protein antigenicity = 0.5679]	0.5153	0.5122	0.5748	0.5489
ALLERGEN ONLINE- ALLERGENICITY	No allergenicity	No allergenicity	No allergenicity	No allergenicity
Z score	-5.16	-5.16	-5.16	-5.16
Ramachandran favored region	86.33%	86.33%	86.33%	86.33%
Phyre 2	Confidence: 100.0%, Coverage: 50%	Confidence: 100.0%, Coverage: 51%	Confidence: 100.0%, Coverage: 55%	Confidence: 100.0%, Coverage: 54%
	Vaccine (v)	**Vaccine (vi)**	**Vaccine (vii)**	**Vaccine (viii) [FINAL VACCINE]**
PROTPARAM- VACCINE STABILITY [primary protein instability index = 38.20]	Instability index: 34.58 (stable), GRAVY: -0.805, M.W: 82710.07	Instability index: 33.92 (stable), GRAVY: -0.791, M.W: 80599.71	Instability index: 35.60 (stable), GRAVY: -0.703, M.W: 74724.74	Instability index: 35.68 (stable), GRAVY: -0.553, M.W: 63720.38
VAXIJEN 2.0- ANTIGENICITY [primary protein antigenicity = 0.5679]	0.5796	0.5807	0.6061	0.6394
ALLERGEN ONLINE- ALLERGENICITY	No allergenicity	No allergenicity	No allergenicity	No allergenicity
Z score	-5.16	-5.16	-5.16	-5.16
Ramachandran favored region	86.33%	86.33%	86.33%	86.33%
Ramachandran Outliers	0.48%	0.48%	0.48%	0.48%
Phyre 2	Confidence: 100.0%, Coverage: 57%	Confidence: 100.0%, Coverage: 59%	Confidence: 100.0%, Coverage: 63%	Confidence: 100.0%, Coverage: 73%

Abbreviations: M.W.: Molecular weight (unit: Dalton), *GRAVY: grand average of hydropathicity*. After comparing the results of biochemical analysis of all vaccine constructs, vaccine (viii) was chosen as the final vaccine.

The stability of the final vaccine was checked using the ProtParam server as shown in figure 7 and 8. According to ProtParam, the number of amino acids in the vaccine was 571 and the molecular weight determined was 63720.38 Da. ProtParam computed the theoretical pI of the vaccine to be 8.92. The net charge of a particular protein remains zero at the pH of the isoelectric point (pI). The percentage of each amino acid in the vaccine was represented. From the amino acid composition, it can be seen that threonine amino acid has the highest percentage. The total number of positively and negatively charged residues were 66 and 52 respectively. In ProtParam, an instability index lower than 40 indicates a stable protein is predicted. The instability index of this vaccine was 35.68 which indicated that the vaccine was stable. The grand average of hydropathy (GRAVY) of the vaccine was -0.553, which was determined by the summation of hydropathy of every amino acid divided by the number of residues (Gasteiger et al., 2005).



Number of amino acids: 571

Molecular weight: 63720.38

Theoretical pI: 8.92

Amino acid composition: CSV format

Ala (A)	37	6.5%
Arg (R)	28	4.9%
Asn (N)	29	5.1%
Asp (D)	22	3.9%
Cys (C)	18	3.2%
Gln (Q)	16	2.8%
Glu (E)	30	5.3%
Gly (G)	39	6.8%
His (H)	27	4.7%
Ile (I)	24	4.2%
Leu (L)	34	6.0%
Lys (K)	38	6.7%
Met (M)	13	2.3%
Phe (F)	14	2.5%
Pro (P)	42	7.4%
Ser (S)	33	5.8%
Thr (T)	56	9.8%
Trp (W)	8	1.4%
Tyr (Y)	25	4.4%
Val (V)	38	6.7%
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): 52
Total number of positively charged residues (Arg + Lys): 66

Figure 7: Amino acid number and composition, theoretical pI, molecular weight, charged residues in ProtParam server (Gasteiger et al., 2005).

Atomic composition:

Carbon	C	2810
Hydrogen	H	4376
Nitrogen	N	800
Oxygen	O	835
Sulfur	S	31

Formula: $C_{2810}H_{4376}N_{800}O_{835}S_{31}$
Total number of atoms: 8852

Extinction coefficients:

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

Ext. coefficient 82375
Abs 0.1% (=1 g/l) 1.293, assuming all pairs of Cys residues form cystines

Ext. coefficient 81250
Abs 0.1% (=1 g/l) 1.275, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is S (Ser).

The estimated half-life is: 1.9 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 35.68
This classifies the protein as stable.

Aliphatic index: 65.39

Grand average of hydropathicity (GRAVY): -0.553

Figure 8: Atomic composition, formula, half-life, instability index, GRAVY in ProtParam server (Gasteiger et al., 2005).

3.10 Allergenicity and Toxicity Prediction of Constructed Vaccine:

As the vaccine sequence was submitted to the AllergenOnline server, the server did not predict the vaccine as allergenic. The AllergenOnline server matches inputted protein sequence with sequences of allergens from its database to check whether the searched sequence may have allergenic properties (Goodman et al., 2016). Since the server did not find any match with known allergen sequences, thus the vaccine was predicted as non-allergenic.

As the vaccine sequence was inputted into the T3DB server, the server returned no results. This indicated that the vaccine was predicted as non-toxic (Wishart et al., 2015). If the sequence had matched a toxic entity from the server database, then the server would have returned a toxicity match result. Since the server returned no results, it was concluded that the vaccine was predicted to be non-toxic (figure 9).

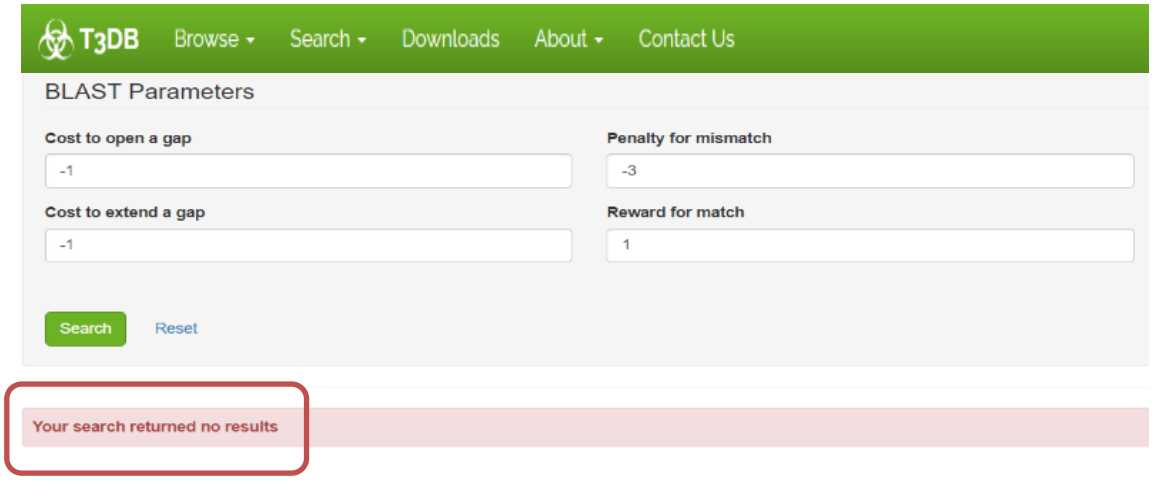


Figure 9: Result of toxicity prediction of the vaccine in the T3DB server (Wishart et al., 2015).

3.11 Homology modeling:

Homology modeling of the vaccine was obtained via the Phyre 2 server (figure 10). The server generated a 3-dimensional model of the vaccine with 100% confidence and 73% coverage. Here 419 residues (73% of the sequence) was modelled with 100% confidence. High confidence and coverage percentage are desired to obtain accurate results. This 3D model of the vaccine helped to complete the subsequent biochemical analysis of the vaccine along with aiding the visual observation of the vaccine structure (Kelley et al., 2015).

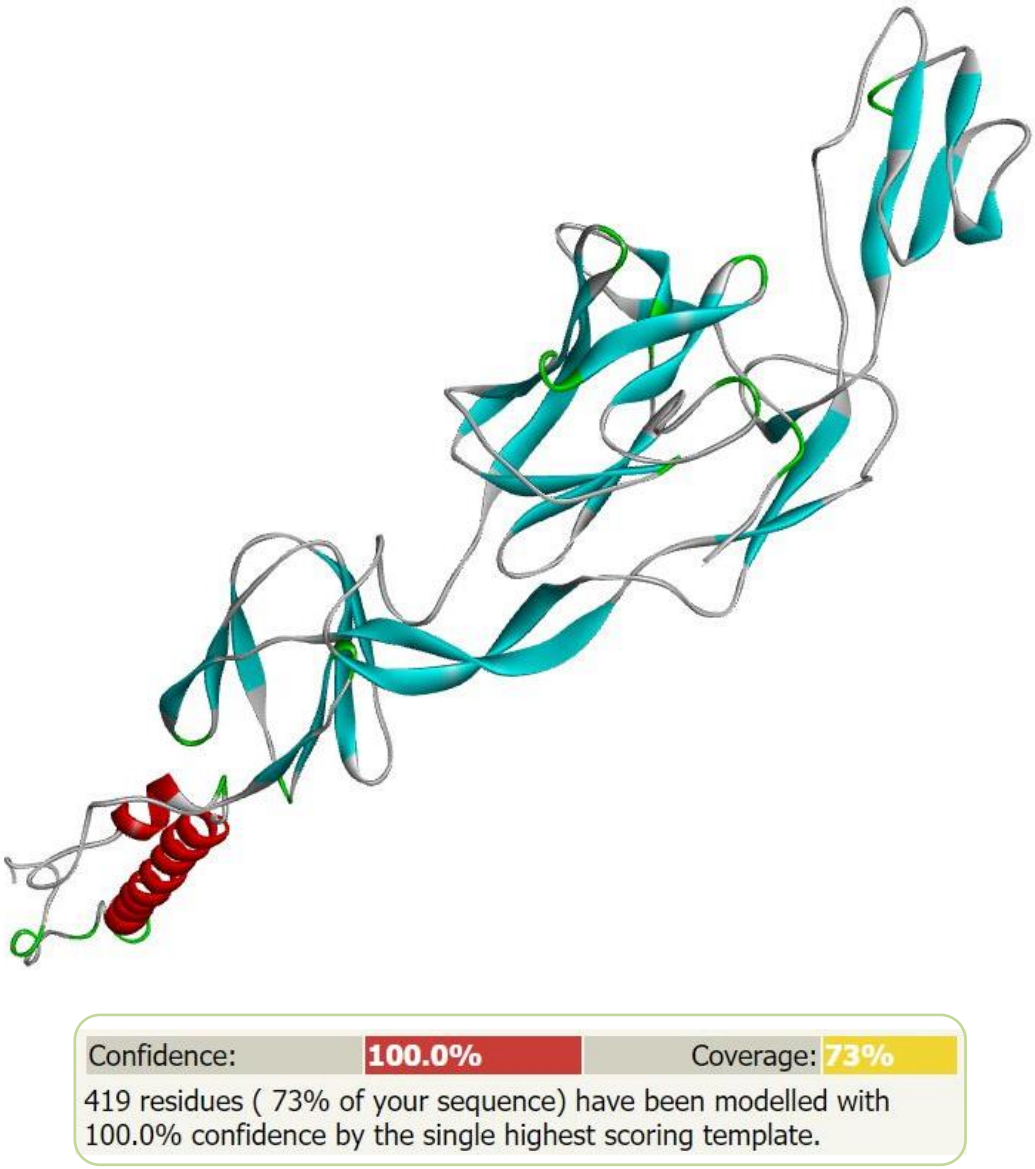
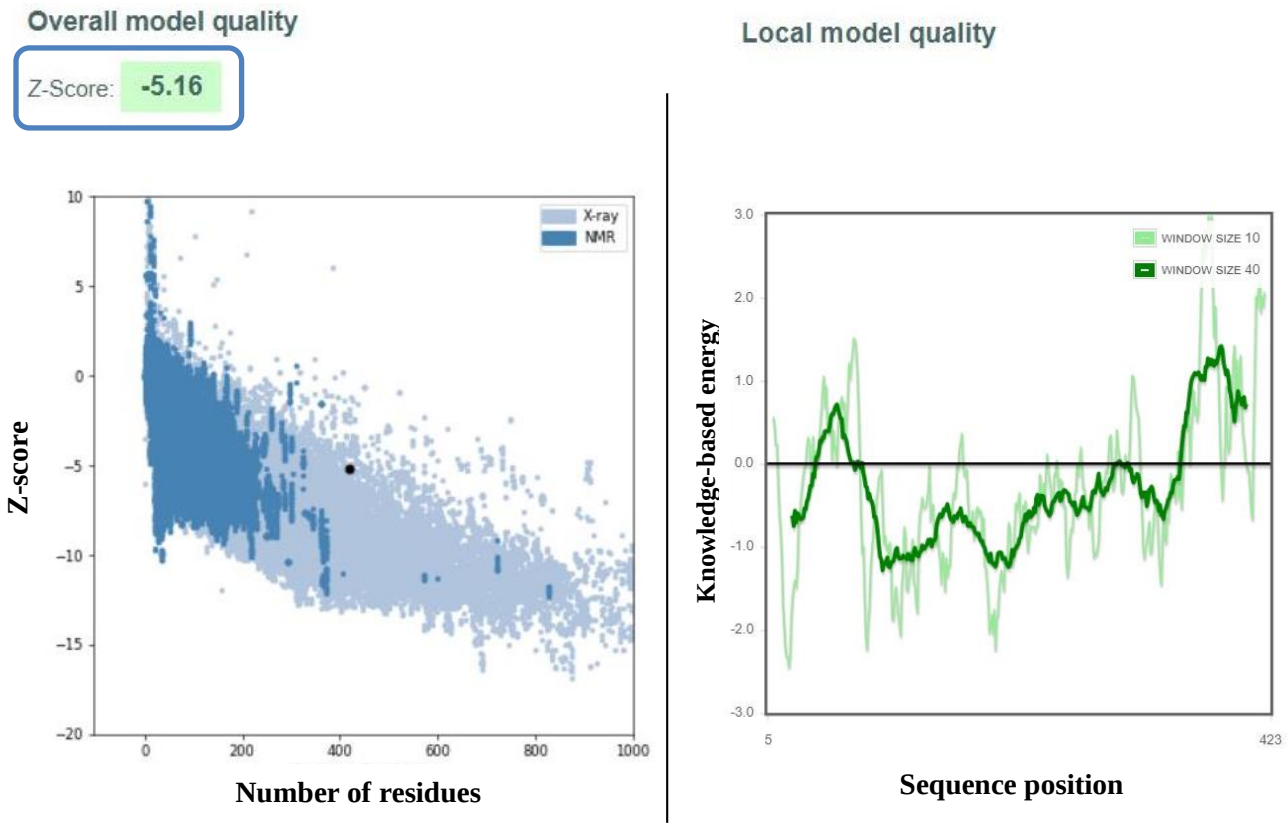


Figure 10: Homology modeling of vaccine via Phyre 2 server (Kelley et al., 2015).

3.12 Z-Score and Ramachandran Plots Evaluation:

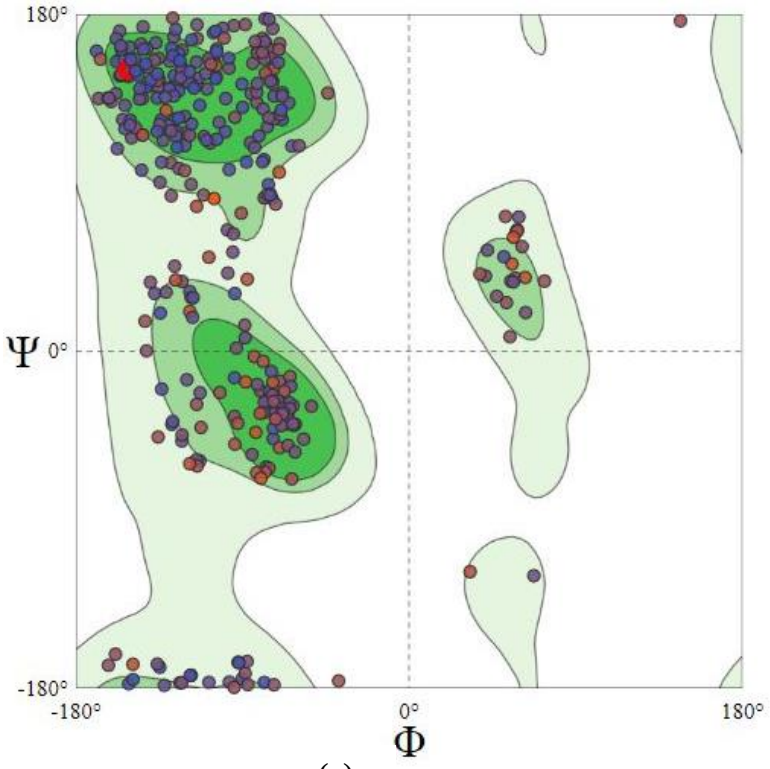
The phyre 2 server generated a pdb file of the final vaccine, which could be opened using Discovery Studio software. The pdb file was also used in further steps of biochemical analysis. Upon submission of the pdb file of the vaccine in ProSAweb, the Z-score versus number of residues graph was obtained on the overall model quality. The Z-score value was -5.16 (figure 11). Another graph was obtained from the server of local model quality which represented the relation between knowledge-based energy and sequence position (Wiederstein & Sippl, 2007).



(a) (b)

Figure 11: (a) Z-score vs. number of residues graph and Z-score in ProSAweb server. (b) Local model quality relating knowledge-based energy with sequence position (Wiederstein & Sippl, 2007).

Structure assessment using SWISS-MODEL server generated the Ramachandran plots (figure 12) of the vaccine along with the quality estimation, residue quality, and MolProbity results (K. Gopalakrishnan et al., 2007). According to the MolProbity results, the Ramachandran favored region of the vaccine was 86.33%, the Ramachandran outliers were 0.48% and the rotamer outlier was 0.0%. The QMEANDisCo Global value was 0.72 ±0.05 (figure 12a and 12b).



(a)

MolProbity Results

MolProbity Score	2.72
Ramachandran Favoured	86.33%
Ramachandran Outliers	0.48%
Rotamer Outliers	0.00%
C-Beta Deviations	0

(b)

Figure 12: (a) Ramachandran plots of vaccine generated in SWISS-MODEL, (b) MolProbity results in SWISS-MODEL (K. Gopalakrishnan et al., 2007)

3.13 Molecular Docking of Final Vaccine with Relevant Human Receptor:

In the PATCHDOCK server, 3W3G was inputted as the receptor molecule which represents the crystal structure of human toll-like receptor 8 (TLR8). The final vaccine was inputted as the ligand molecule. The server presented 20 solutions for the receptor-ligand complex. The highest scoring complex (solution 1) had a score of 20126, an area of 2330.60, and an Atomic contact energy (ACE) of 494.07. The transformation of this complex was -1.76 0.75 -2.47 -31.26 253.35 74.31 (figure 13 and 14). Patchdock creates flat, concave, or convex patches from the molecular surface representation. The server then produces possible transformations by matching complementary patches. After the subsequent assessment, root mean square deviation (RMSD) is lastly followed to eliminate solutions that are unessential (Schneidman-Duhovny et al., 2005). Toll-like receptors act as pattern recognition receptors which play a significant function in the immune system by sensing pathogens. Inflammatory cytokines are produced in the body when TLRs are triggered. In presence of a virus, TLR8 in particular promotes interferon production. TLR8 is also triggered by

bacterial RNA. Agonists of TLR8 receptors have been a lucrative target in the case of vaccine adjuvants as well (Cervantes et al., 2012).

Receptor	Ligand	Complex Type	Clustering RMSD	User e-mail
3w3g	Vaccine_viii.pdb	Default	4.0	intizarahmed726@gmail.com

Solution No	Score	Area	ACE	Transformation
1	20126	2330.60	494.07	-1.76 0.75 -2.47 -31.26 253.35 74.31

Figure 13: Highest scoring solution in PATCHDOCK (Schneidman-Duhovny et al., 2005).

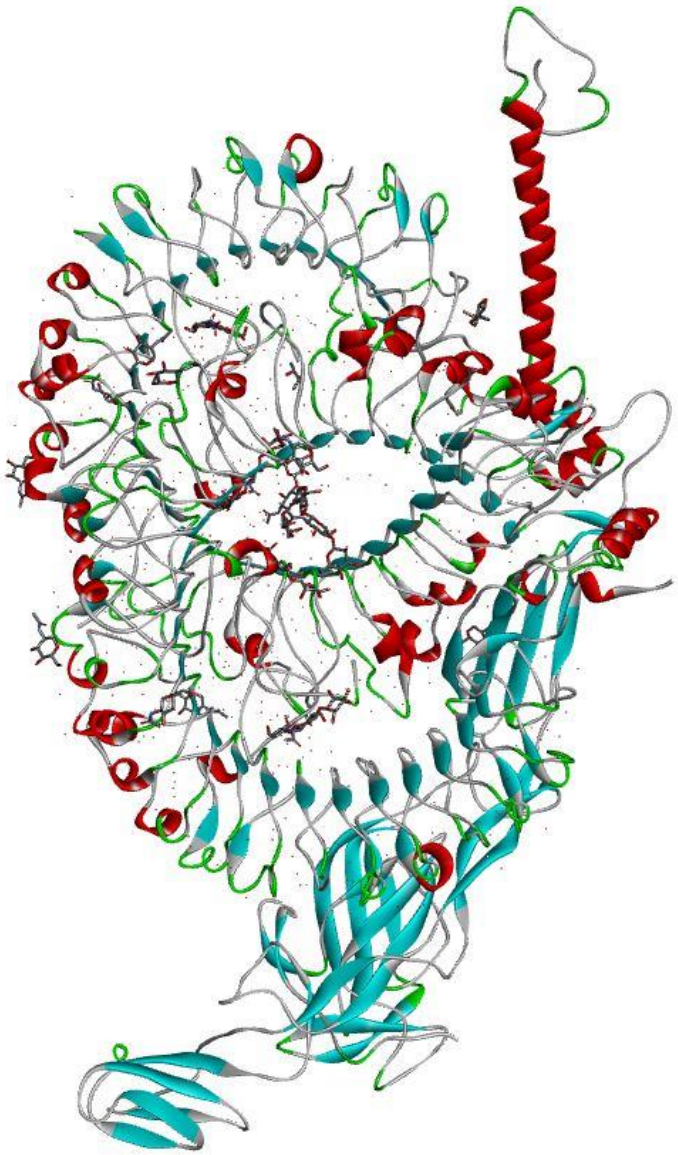
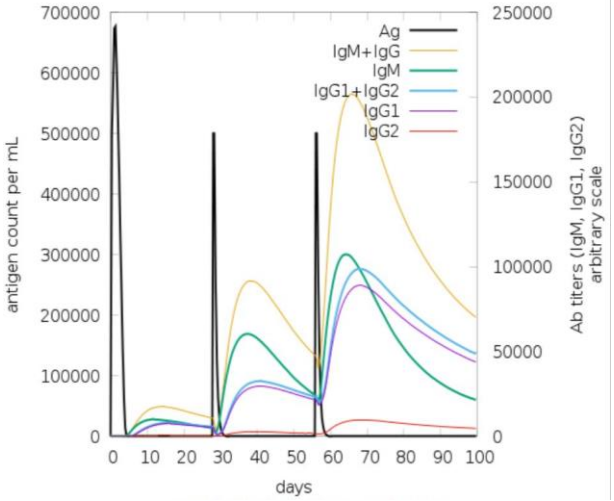


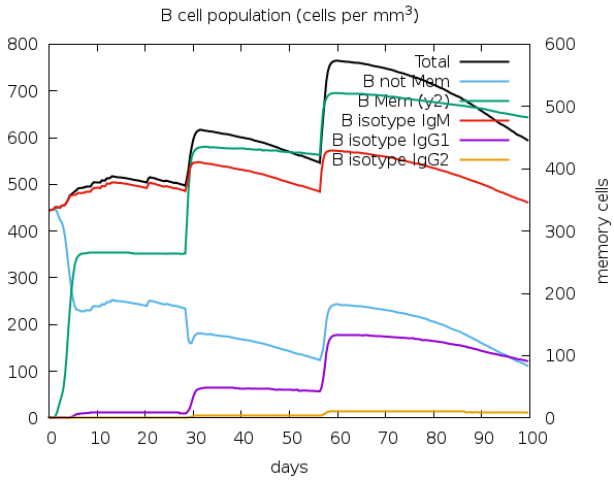
Figure 14: Molecular docking of vaccine and receptor in PATCHDOCK (Schneidman-Duhovny et al., 2005).

3.14 Immune Simulations:

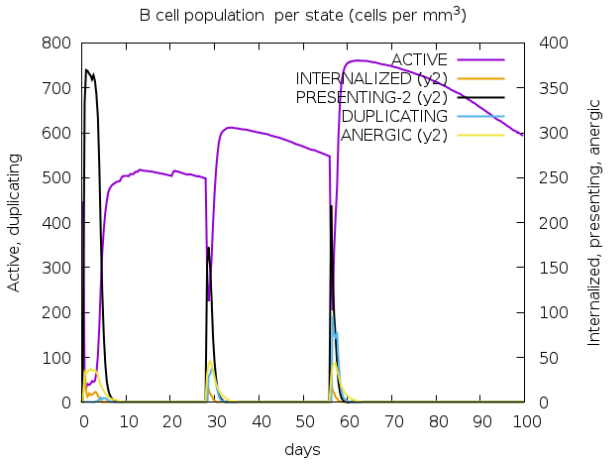
Multiple doses, including booster doses, are usually necessary to maintain a level of the vaccine in the body which would be sufficient to create the desired sustained immune response. The immune response generated in the body needs to be assessed by observing levels of specific antibodies or immunoglobulin with respect to the time and dosage regimen of the vaccine. Moreover, the escalation and reduction of different immune cells in response to the vaccine doses need to be assessed. These parameters in response to the vaccine were predicted and represented graphically by the C-IMMSIM server as shown in the figure 15 (a-j) (Rapin et al., 2010).



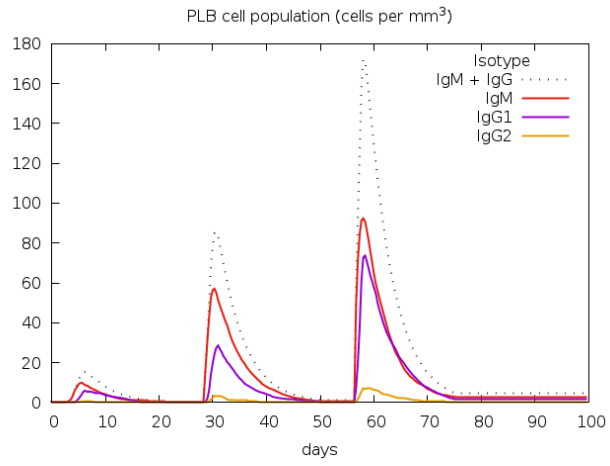
(a) Antigen count per mL and antibody titers (Rapin et al., 2010)



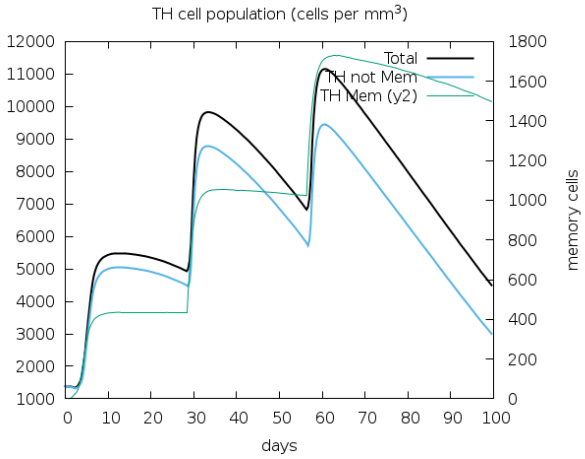
(b) Total count of B lymphocytes and memory cells (Rapin et al., 2010)



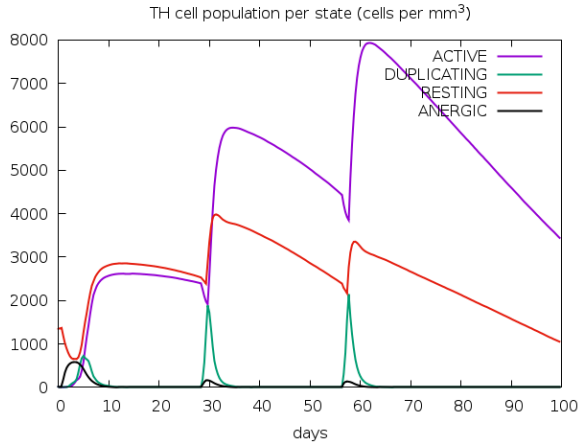
(c) Entity-state of B cell population (Rapin et al., 2010)



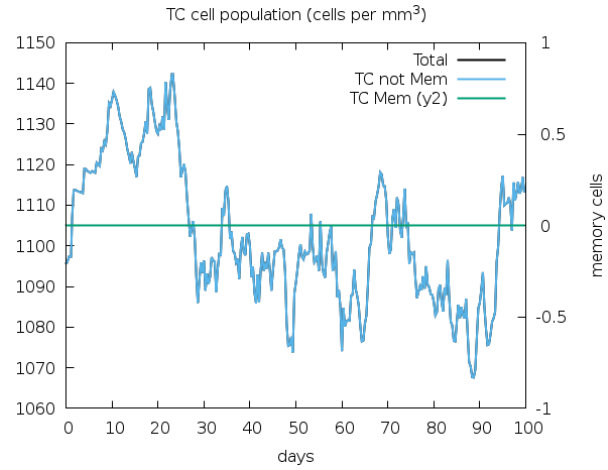
(d) Number of plasma B cells according to their isotypes (Rapin et al., 2010)



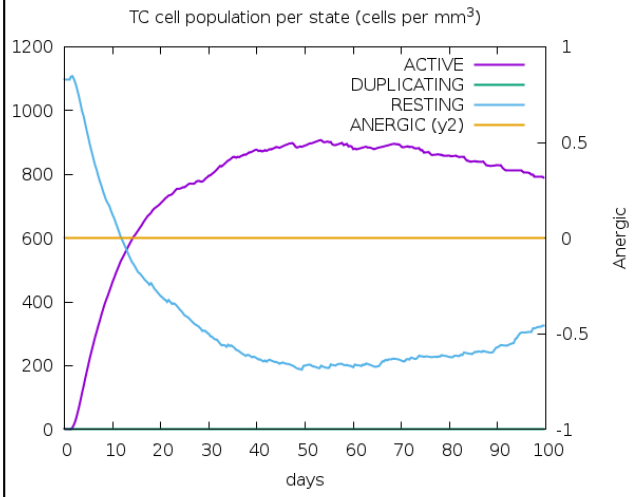
(e) Number of CD4 T helper lymphocytes (Rapin et al., 2010)



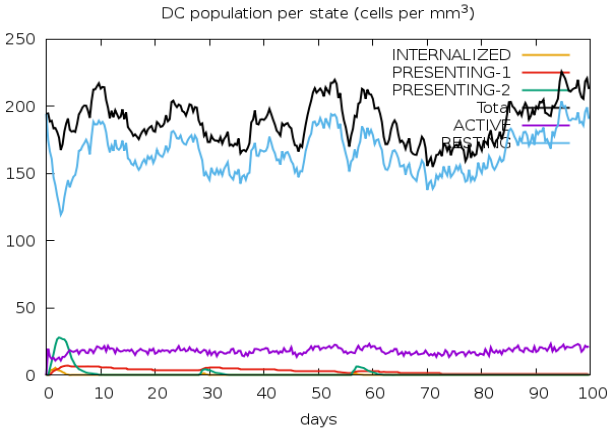
(f) Entity state of CD4 Helper T cells (Rapin et al., 2010)



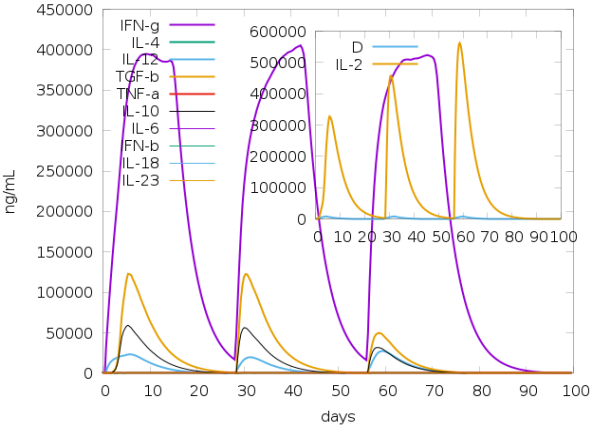
(g) Number of CD8 T cytotoxic (TC) cells (Rapin et al., 2010)



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



(i) Total number of dendritic cells and their states (Rapin et al., 2010)



(j) Levels of cytokines including IFN-γ, IL-4, IL-10 (Rapin et al., 2010)

Figure 15 (a-j): Immune simulations of vaccine via C-IMMSIM (Rapin et al., 2010)

Antibodies play important role in acquired immunity. From Figure (a), the rise in antigen count per mL can be observed following the first immunization. Although the antibodies IgM and IgG rise slightly in the initial period within the first 28 days, the antibodies rise significantly after approximately 60 days due to repeated administration of and completion of the vaccine doses. Production of antibodies indicated that the vaccine was predicted to create a desired immune response. B lymphocytes are the immune cells that create memory cells in the body alongside producing antibodies. These memory cells created from B cells render the body the ability to recognize the same pathogen in the future and provide protective immune function quicker. Thus the production of capable memory cells is essential for effective immunization. The graph showed the total count of B lymphocytes and memory cells following vaccination according to different isotypes in figure (b). The graph in figure (c) differentiated and represented the B cells per entity state. It illustrated the number of active B cells, number of B cells presenting on Class ii, B cell count which internalized the antigen, and count of B cells undergoing duplication and anergy. The number of plasma B lymphocytes (PLB) was shown in graph (d) including their isotypes IgG1, IgG2, IgM, and IgM + IgG. The count of CD4 T helper lymphocytes following each vaccination dose was predicted in figure (e). The number of Helper T cells predicted to be active, to undergoing duplication, be in resting state or be in anergy state was simulated by C-IMMSIM in figure (f). Number of CD8 T cytotoxic (TC) cells (total and memory) along with their active, duplicative, resting and anergic states were simulated as shown in figure (g) and (h). The active, resting, antigen-presenting, and internalized dendritic cells were differentiated in the graph shown in figure (i). PRESENTING-1 here indicated DC cells presenting on MHC class I molecules whereas PRESENTING-2 here denoted DC cells presenting on MHC class II molecules. The danger signal and interleukin-2 level were represented in the inset figure of the graph as shown in figure (j).

Chapter 4

Discussion:

The primary protein selected for vaccine construction in this study was the E2 protein of CHIKV. The CTL, HTL, and B cell epitopes for vaccine construction were searched based on this primary protein. This E2 protein has a significant role in the life cycle of CHIKV. A phylogenetic study conducted on a 2017 CHIKV strain collected from Bangladesh divulged that the virus belonged to Indian Ocean Lineage (IOL). The genotype of the CHIKV was familiar to preceding outbreaks, but two specific mutations were observed in these viruses. One of the mutations was valine to alanine mutation of the E2 protein at 264 position. The mutated virus entered Bangladesh in 2015 and the most severe chikungunya virus outbreak emerged in 2017 (Phadungsombat et al., 2020). In this study, the E2 protein was also found to have the highest antigenicity score (0.5679) in VaxiJen v2.0 server compared to the other screened primary proteins of CHIKV (figure 2). Since this E2 protein was predicted as a probable-antigen and it also plays a vital role in the lifecycle of CHIKV, it was a suitable protein to proceed with further in the vaccine design.

It is known that cytotoxic T lymphocytes have a crucial function in the immune response of the body as these are able to selectively kill target cells that possess specific antigens. Such target cells are killed in apoptosis or lysis mechanism due to the action of CTLs (Charles A Janeway et al., 2001). Therefore incorporation of CTL epitopes in the vaccine can help to stimulate active immunity against CHIKV. CTLs can also aid in the development of immunological memory after vaccination (Kumar et al., 2021). In this experiment, 11 cytotoxic T lymphocyte (CTL) epitopes were obtained initially for the E2 protein sequence via the NetCTL-1.2 server applying a threshold of 0.75 (figure 3). These 11 CTL epitopes were further screened for their respective alleles, toxicity, antigenicity, and allergenicity.

By utilizing bioinformatics tools, peptide binding of protein sequences with human leucocyte antigens HLA-A and HLA-B can be predicted with substantial precision. In this study, MHC I alleles specific to these CTL epitopes were obtained using NetMHCpan-4.1 server. Among the 11 CTL epitopes, alleles were found for 6 epitopes (table 2). In case of cell-mediated immunity, peptide binding to the major histocompatibility complex or MHC molecules is a very important step. This is because the T-cell receptors can recognize this complex of peptide and MHC class I

molecule and consequently stimulate an immune response in the body. T-cell receptors of CTLs are able to recognize such complexes (Lundegaard et al., 2010). Hence the prediction of MHC I allele binding was performed in this study.

The in-silico prediction of toxicity of each CTL epitope was conducted in the ToxinPred server (figure 5). All CTL epitopes were predicted as non-toxin, thus no CTL epitope was excluded based on this criteria. Four out of the six CTL epitopes were predicted to be probable antigens in the VaxiJen v2.0 server. This indicated that most of these epitopes could have antigenic properties helpful for generating an immune response. Moreover, three out of the six epitopes were identified as probable allergens in the AllerTOP v. 2.0 server. However, the probable allergen CTL epitopes were not excluded yet since the final vaccine could still be non-allergenic.

Finding effective HTL epitopes for the selected protein sequence is essential in designing a potential subunit vaccine. From the NetMHCIIpan 4.0 Server, 62 strong binding HTL epitopes were found for the E2 protein of CHIKV. Activation of helper T lymphocytes (HTLs) can facilitate successful vaccination. HTLs or $CD4^+$ cells have a significant role in the immunity of the body as these cells are responsible for triggering almost all other defense cells of the immune system. It is these HTLs that stimulate cytotoxic T cells to kill pathogens and that stimulate B cells to generate antibodies. HTLs also trigger effector cells and macrophages. These helper T cells are activated when CD4 proteins on their surface bind to MHC class II molecules (Britannica, n.d.).

Helper T cells can be subdivided into T_H1 and T_H2 cells, where each group secretes different types of cytokines. Interleukin 4 (IL-4) and interleukin 10 (IL-10) cytokines are both mainly secreted by T_H2 cells. Whereas interferon-gamma (IFN-γ) is mainly secreted by T_H1 cells. The 62 HTLs epitopes obtained in this study were screened to predict which HTLs induce the cytokines IL-4, IL-10, and IFN-γ simultaneously (figure 7, 8, 9). Through the screening, 7 such HTL epitopes were found (table 4). Only these 7 HTLs were selected for further vaccine design. The ability of these HTL epitopes to induce IL-4, IL-10, and IFN-γ will contribute to several functions in the final vaccine. It is known that IL-4 acts on B cells and helper T cells, which facilitates the proliferation and maturation of B cells and antibody class switching. IFN-γ acts on macrophages, B cells, and endothelial cells, which aids in the activation of macrophages and different MHC

genes. Thus the ability of HTL epitopes in the vaccine to induce IL-4 and IFN-γ will help to attain a desired immune response. On the other hand, IL-10 has an inhibitory effect. It inhibits macrophages and suppresses the immune system (Alberts et al., 2002). Therefore, the ability of HTL epitopes to induce IL-10 will help to limit the excessive immune response that may be harmful to the human host (Iyer & Cheng, 2012).

Incorporating B cell epitopes in the vaccine is also necessary to elicit a substantial immune response. 14 peptides were found as B cell epitopes from the Bepipred Linear Epitope Prediction 2.0 tool (table 5). Since B cells have fundamental functions in producing antibodies and generating memory cells that recognize pathogens upon future infections, B cell epitopes are crucial in vaccine design. It has been observed that B cell epitopes composed of short residues of 7 to 12 amino acids in length cannot form complexes. This is why B cell epitopes having a short length should be excluded (Parvizpour et al., 2020).

After identifying 6 CTL epitopes, 7 HTL epitopes, and 14 B cell epitopes, different vaccine candidates were assembled. Appropriate linkers were used to connect the epitopes in accordance with available literature. Eight vaccine candidates (vaccines i to viii) were assembled with varying combinations of CTL, HTL, and B cell epitopes (table 10). After comparing the in-silico biochemical analysis of these vaccines (table 11), vaccine no. viii was chosen as the final vaccine as it showed the best results.

The stability of a vaccine is a crucial parameter. The final vaccine was predicted as stable in the ProtParam server. Moreover, it had the lowest molecular weight compared to other candidates which was desired. The negative GRAVY value of this vaccine suggested that the vaccine had hydrophilic characteristics (figure 12). The low GRAVY value indicated increased solubility of the vaccine. The antigenicity of the vaccine was remarkably improved compared to the preliminary primary protein, which is suitable for generating an adequate immune response in the body. Alongside creating a sufficient desired immune response, an ideal vaccine must not induce an allergic reaction in the body. The final vaccine was not predicted as an allergen when tested in AllergenOnline (figure 13). A vaccine must not have toxic properties. The final vaccine designed in this experiment was not identified as any toxic entity when its toxicity was predicted in the

T3DB server (figure 9). Homology modeling of the vaccine via Phyre 2 illustrated the 3D structure of the vaccine with high confidence and coverage (figure 10). The molecular docking of the vaccine with human TLR8 receptor was shown with the Patchdock server (figure 13 and 14). Z-score analysis using ProSAweb and Ramachandran plots of the vaccine using SWISS-MODEL showed promising results (figure 11 and 12). However, the Ramachandran favored region could be improved further to reach greater than 90 percent.

Finally, the immune response to the vaccine was simulated via C-IMMSIM. The server predicted positive results in the levels of immune cells in the body with respect to subsequent doses of the vaccine. IgG antibodies protect against viral infections and its production requires some time after vaccination. In C-IMMSIM, the level of antibodies IgG and IgM elevated significantly after approximately 60 days, indicating a desired antibody response (figure 15a). Production of B memory cells is vital in immunization, as these memory cells will aid in the quick recognition of CHIKV if encountered in the future. The level of B memory cells elevated considerably after each vaccine dose (figure 15b). The rise of plasma B cells was also observed (figure 15c). This indicated the maturation and differentiation of B cells into plasma cells which will secrete antibodies against the vaccine antigen. The predictive response of helper T cells was positive in C-IMMSIM (figure 15e and 15f), these helper T cells will stimulate other immune cells to generate an immune response. The number of active T regulatory lymphocytes rises to a peak after the first vaccine dose, which gradually lowers over time. These T regulatory lymphocytes are helpful to counteract excessive immune response and restore homeostasis in the body. An increase in the cytokines IL-2, IL-10, IL-12 and IFN-γ after each vaccine dose administration was evident from the predictions in C-IMMSIM (figure 15j). However, there was no indication of an increase in IL-4 in the simulation, which needs to be addressed with further experimentation.

Chapter 5

Conclusion

To conclude, in spite of ongoing studies since the inception of the chikungunya virus, an effective vaccine for the virus does not exist. Although infection caused by chikungunya virus is usually not fatal, it can cause debilitating symptoms. This infection is detrimental especially for elderly or comorbid patients. An effective vaccine can offer stronger protection against the virus if there is a re-emergence in the future. In this study, in-silico methods helped to design a multi-epitope vaccine for the chikungunya virus by targeting its E2 envelope glycoprotein. The vaccine was constructed step-by-step in a systematic way, by identifying crucial epitopes using reliable computational tools. From the biochemical analysis conducted with in-silico tools as well, it was evident that the designed vaccine for immunization against the chikungunya virus showed promising results. However, further experimentation is needed to address the limitations of this study. The results obtained for the vaccine from computational methods may not translate similarly when tested in in-vivo physiological conditions. Since this study was conducted with an in-silico approach, it must be considered that further in vitro and in vivo studies are essential to confirm the safety and efficacy characteristics of the designed vaccine. Such systematic in-silico design of vaccines can accelerate the discovery of effective vaccines which may provide immunity against the serious chikungunya infection.

References

Abdelnabi, R., Neyts, J., & Delang, L. (2015). Towards antivirals against chikungunya virus. *Antiviral Research*, *121*. https://doi.org/10.1016/J.ANTIVIRAL.2015.06.017

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Helper T Cells and Lymphocyte Activation*. https://www.ncbi.nlm.nih.gov/books/NBK26827/

Anwar, S., Mourosi, J. T., Khan, M. F., Ullah, M. O., Vanakker, O. M., & Hosen, M. J. (2020). Chikungunya outbreak in Bangladesh (2017): Clinical and hematological findings. *PLoS Neglected Tropical Diseases*, *14*. https://doi.org/10.1371/JOURNAL.PNTD.0007466

Ayyagari, V. S., Venkateswarulu, T. C., Abraham Peele, K., & Srirama, K. (2020). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of Biomolecular Structure & Dynamics*, *40*. https://doi.org/10.1080/07391102.2020.1850357

Battisti, V., Urban, E., & Langer, T. (2021). Antivirals against the Chikungunya Virus. *Viruses*, *13*. https://doi.org/10.3390/V13071307

Bhuiyan, M. A., Tasnim Quayum, S., Ahammad, F., Alam, R., Samad, A., Nain, Z., Nainu, F., Tallei, T. E., Ratulangi, S., Moreno, S., & Brun, A. (2020). Discovery of potential immune epitopes and peptide vaccine design - a prophylactic strategy against Rift Valley fever virus. *F1000Research*, *9*. https://doi.org/10.12688/f1000research.24975.1

Britannica. (2022). *helper T cell |Britannica*. Retrieved October 5, 2022, from https://www.britannica.com/science/helper-T-cell

Burt, F. J., Chen, W., Miner, J. J., Lenschow, D. J., Merits, A., Schnettler, E., Kohl, A., Rudd, P. A., Taylor, A., Herrero, L. J., Zaid, A., Ng, L. F. P., & Mahalingam, S. (2017). Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. *The Lancet. Infectious Diseases*, *17*. https://doi.org/10.1016/S1473-3099(16)30385-1

Castiglione, F., Deb, D., Srivastava, A. P., Liò, P., & Liso, A. (2021). From Infection to Immunity: Understanding the Response to SARS-CoV2 Through In-Silico Modeling. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/FIMMU.2021.646972/BIBTEX

Cavrini, F., Gaibani, P., Pierro, A. M., Rossini, G., Landini, M. P., & Sambri, V. (2009). Chikungunya: an emerging and spreading arthropod-borne viral disease. *Journal of Infection in Developing Countries*, *3*. https://doi.org/10.3855/JIDC.169

Cervantes, J. L., Weinerman, B., Basole, C., & Salazar, J. C. (2012). TLR8: the forgotten relative revindicated. *Cellular & Molecular Immunology*, *9*. https://doi.org/10.1038/cmi.2012.38

Charles A Janeway, J., Travers, P., Walport, M., & Shlomchik, M. J. (2001). *T cell-mediated cytotoxicity*. https://www.ncbi.nlm.nih.gov/books/NBK27101/

Chevillon, C., Briant, L., Renaud, F., & Devaux, C. (2008). The Chikungunya threat: an ecological and evolutionary perspective. *Trends in Microbiology*, *16*. https://doi.org/10.1016/J.TIM.2007.12.003

Da Silva Pessoa Vieira, C. J., Da Silva, D. J. F., Kubiszeski, J. R., MacHado, L. C., Pena, L. J., De Morais Bronzoni, R. V., & Da Luz Wallau, G. (2020). The Emergence of Chikungunya ECSA Lineage in a Mayaro Endemic

Region on the Southern Border of the Amazon Forest. *Tropical Medicine and Infectious Disease*, *5*. https://doi.org/10.3390/TROPICALMED5020105

DeFilippis, V. R. (2022). Chikungunya Virus Vaccines: Platforms, Progress, and Challenges. *Current Topics in Microbiology and Immunology*, *435*. https://doi.org/10.1007/82_2019_175/TABLES/1

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*. https://doi.org/10.1155/2013/263952

Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, *8*. https://doi.org/10.1186/1745-6150-8-30/TABLES/9

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-4/TABLES/2

Duhovny, D., Nussinov, R., & Wolfson, H. J. (2009). *Efficient Unbound Docking of Rigid Molecules*.

First identified outbreak of Chikungunya in Bangladesh. Health and Science Bulletin. http://dspace.icddrb.org/jspui/bitstream/123456789/4865/1/2009-ICDDRBHealthScienceBulletin-Vol7%281%29-English.pdf

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*. https://doi.org/10.1385/1-59259-890-0:571

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*. https://doi.org/10.1002/MNFR.201500769

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS*, *30*. https://doi.org/10.1002/ELPS.200900140

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*. https://doi.org/10.1371/JOURNAL.PONE.0073957

Haque, F., Rahman, M., Banu, N. N., Sharif, A. R., Jubayer, S., Shamsuzzaman, A., Alamgir, A., Erasmus, J. H., Guzman, H., Forrester, N., Luby, S. P., & Gurley, E. S. (2019). An epidemic of chikungunya in northwestern Bangladesh in 2011. *PLoS ONE*, *14*. https://doi.org/10.1371/JOURNAL.PONE.0212218

Iyer, S. S., & Cheng, G. (2012). Role of Interleukin 10 Transcriptional Regulation in Inflammation and Autoimmune Disease. *Critical Reviews in Immunology*, *32*. https://doi.org/10.1615/CRITREVIMMUNOL.V32.I1.30

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*. https://doi.org/10.1093/NAR/GKX346

K. Gopalakrishnan, G. Sowmiya, S. S. Sheik, & K. Sekar. (2007). Ramachandran Plot on The Web (2.0). *Protein & Peptide Letters*, *14*. https://doi.org/10.2174/092986607781483912

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*. https://doi.org/10.1038/nprot.2015.053

Kori, P., Sajjan, S. S., & Madagi, S. B. (2015). In silico prediction of epitopes for Chikungunya viral strains. *Journal of Pharmaceutical Investigation*, *45*. https://doi.org/10.1007/S40005-015-0205-0/FIGURES/2

Kril, V., Aiqui-Reboul-Paviet, O., Briant, L., & Amara, A. (2021). New Insights into Chikungunya Virus Infection and Pathogenesis. *Annurev-Virology*, *8*. https://doi.org/10.1146/ANNUREV-VIROLOGY-091919-102021

Kumar, N., Admane, N., Kumari, A., Sood, D., Grover, S., Prajapati, V. K., Chandra, R., & Grover, A. (2021). Cytotoxic T-lymphocyte elicited vaccine against SARS-CoV-2 employing immunoinformatics framework. *Scientific Reports*, *11*. https://doi.org/10.1038/s41598-021-86986-6

Labadie, K., Larcher, T., Joubert, C., Mannioui, A., Delache, B., Brochard, P., Guigand, L., Dubreil, L., Lebon, P., Verrier, B., De Lamballerie, X., Suhrbier, A., Cherel, Y., Grand, R. Le, & Roques, P. (2010). Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. *The Journal of Clinical Investigation*, *120*. https://doi.org/10.1172/JCI40104

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*. https://doi.org/10.1186/1471-2105-8-424/TABLES/3

Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A. C., Neveu, V., & Wishart, D. S. (2010). T3DB: a comprehensively annotated database of common toxins and their targets. *Nucleic Acids Research*, *38*. https://doi.org/10.1093/NAR/GKP934

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*. https://doi.org/10.1186/1471-2172-9-8

Lundegaard, C., Lund, O., Buus, S., & Nielsen, M. (2010). Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology*, *130*. https://doi.org/10.1111/J.1365-2567.2010.03300.X

Mahalingam, S., Tharmarajah, K., & Zaid, A. (2017). Chikungunya: vaccines and therapeutics. *F1000Research*, *6*. https://doi.org/10.12688/F1000RESEARCH.12461.1

McFee, R. B. (2018). Selected mosquito-borne illnesses—Chikungunya. *Disease-a-Month*, *64*. https://doi.org/10.1016/J.DISAMONTH.2018.01.009

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*. https://doi.org/10.1038/srep42851

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 : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *61*. https://doi.org/10.1016/J.MEEGID.2018.03.007

Parvizpour, S., Pourseif, M. M., Razmara, J., Rafi, M. A., & Omidi, Y. (2020). Epitope-based vaccine design: a

comprehensive overview of bioinformatics approaches. *Drug Discovery Today*, *25*. https://doi.org/10.1016/J.DRUDIS.2020.03.006

Phadungsombat, J., Imad, H., Rahman, M., Nakayama, E. E., Kludkleeb, S., Ponam, T., Rahim, R., Hasan, A., Poltep, K., Yamanaka, A., Matsee, W., Piyaphanee, W., Phumratanaprapin, W., & Shioda, T. (2020). A Novel Sub-Lineage of Chikungunya Virus East/Central/South African Genotype Indian Ocean Lineage Caused Sequential Outbreaks in Bangladesh and Thailand. *Viruses*, *12*. https://doi.org/10.3390/V12111319

Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H. (2012). ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, *40*. https://doi.org/10.1093/NAR/GKR859

Prevention of Chikungunya infection. (2022). Centers for Disease Control and Prevention. https://www.cdc.gov/chikungunya/prevention/index.html

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*. https://doi.org/10.1371/JOURNAL.PONE.0009862

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*. https://doi.org/10.1093/NAR/GKAA379

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., Peters, B., Nielsen, M., & Nielsen, M. (2020). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *Journal of Proteome Research*, *19*. https://doi.org/10.1021/ACS.JPROTEOME.9B00874

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*. https://doi.org/10.1093/NAR/GKI481

Silva, L. A., & Dermody, T. S. (2017). Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. *The Journal of Clinical Investigation*, *127*. https://doi.org/10.1172/JCI84417

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*. https://doi.org/10.1002/PROT.340170404

Solignat, M., Gay, B., Higgs, S., Briant, L., & Devaux, C. (2009). Replication cycle of chikungunya: A re-emerging arbovirus. *Virology*, *393*. https://doi.org/10.1016/J.VIROL.2009.07.024

Tahir Ul Qamar, M., Bari, A., Adeel, M. M., Maryam, A., Ashfaq, U. A., Du, X., Muneer, I., Ahmad, H. I., & Wang, J. (2018). Peptide vaccine against chikungunya virus: Immuno-informatics combined with molecular docking approach. *Journal of Translational Medicine*, *16*. https://doi.org/10.1186/S12967-018-1672-7/FIGURES/7

Vairo, F., Haider, N., Kock, R., Ntoumi, F., Ippolito, G., & Zumla, A. (2019). Chikungunya: Epidemiology, Pathogenesis, Clinical Features, Management, and Prevention. *Infectious Disease Clinics of North America*,

33(. https://doi.org/10.1016/J.IDC.2019.08.006

Van der Heijden, M. W., & Bol, J. F. (2014). Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins. *Archives of Virology*, *147*. https://doi.org/10.1007/S00705-001-0773-3

Weger-Lucarelli, J., Aliota, M. T., Wlodarchak, N., Kamlangdee, A., Swanson, R., & Osorio, J. E. (2016). Dissecting the Role of E2 Protein Domains in Alphavirus Pathogenicity. *Journal of Virology*, *90*. https://doi.org/10.1128/JVI.02792-15/ASSET/81B5BF24-320B-4DC1-9FC9-B9F1FF55FBCE/ASSETS/GRAPHIC/ZJV9990913490008.JPEG

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*. https://doi.org/10.1093/NAR/GKM290

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*. https://doi.org/10.1093/NAR/GKU1004