Developing a novel multi-epitope vaccine against Dengue virus targeting the Envelope protein using an Immunoinformatics approach

By

S M NASIFUZZAMAN

17336015

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of

Bachelor of Science in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

AUGUST 2021

Declaration

It is hereby declared that

1. The Thesis report that I submitted is my original work while completing my degree at BRAC University.

2. The report 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 report does not contain material that 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:

S M NASIFUZZMAN

17336015

Approval

The project titled "Developing a novel multi-epitope vaccine against Dengue virus targeting the envelope protein using immunoinformatics approach" submitted by S M NASIFUZZAMAN (17336015) of Summer 2017 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology on 9th August 2021.

Examining Committee:

Supervisor:

(Member) _____

Full Name

Designation, Department

Institution

Program Coordinator:

(Member) _____

Full Name

Designation, Department

Institution

Departmental Head:

(Chair)

_____

Full Name

Designation, Department

Institution

Acknowledgment

First and foremost, I would like to thank Almighty Allah for gifting me knowledge, capability, and the strength to complete my thesis project in the best way possible.

I would like to express my gratitude to Professor Dr. A F M Yusuf Haider, (Chairperson Department of Mathematics and Natural Sciences, BRAC University) and Iftekhar Bin Naser Sir for looking after all the students and teachers under his department and always providing his helping hands whenever needed.

I am deeply grateful to S M Rakib-Uz-Zaman sir for guiding me throughout my thesis project and providing me all the necessary help whenever I needed. His vast knowledge and years of experience has greatly benefitted me in doing my thesis project. I appreciate all the teachers that taught me throughout my undergraduate level and helped me understand the theories, as well as the lab work very well. The knowledge that I gained from their teaching helped me better understand every aspect of my research that I did during my undergrad thesis work.

Table of Content

Table of content

List of Figures

Developing a novel multi-epitope vaccine against Dengue virus targeting the envelope protein using an immunoinformatics approach

Abstract

Every year during the rainy season in Bangladesh severe Dengue infection is widespread throughout the country. Aedes aegypti and Ae. Albopictus are the two vectors that are responsible for the transmission of the virus. Thousands of lives including infants, adults, and old people die and many suffer painful fever due to this infection. In some severe cases, the patients experience a drastic decrease in their platelet count which sometimes causes hemorrhage and eventually leads to death. To protect lives against the Dengue virus, vaccination is an effective method. In our study, we have designed a vaccine against the Dengue virus by reverse vaccinology method. We have developed our vaccine targeting the envelope protein since it is highly conserved and immunogenic. Our study was mainly focused on predicting the right B cell epitopes, Cytotoxic T cell epitopes, and Helper T cell epitopes. These epitopes will be the key component of the vaccine. Stimulation of these cells in the body will help to acquire both Humoral and Cell-mediated immunity. As our study followed a immunoinformatics approach we employed various online bioinformatics tools and servers for epitope prediction.

Keywords: Dengue virus, Envelope protein, Reverse vaccinology, multi-epitope vaccine, B cell, CTL, HTL, Immunoinformatics approach.

Chapter 1

Introduction

Dengue is a viral disease with a 0.8 to 2.5 mortality rate. It has been widespread in almost all the regions of the world health organization. As Dengue is a mosquito-borne disease it is transmitted through female mosquitoes of the species Aedes aegypti and, Ae. Albopictus. The dissemination of dengue is impacted by fast unplanned urbanization, rainfall, temperature, and humidity. The vectors for dengue viruses i.e., the female Aedes mosquitos are also the carriers of Zika and Chikungunya. During the monsoon season in Bangladesh dengue is one of the main causes of hospitalization and death of many children and adults.

The epidemiology of Dengue is quite distinct as it is associated with four serotypes of the virus. Dengue is very often transferred from one place to another by infected travelers. The global and national economy is greatly hampered by dengue because of its adverse effects on human health. Many countries, especially the countries of the tropical region have a high level of dengue occurrence.

Figure 1: Aedes aegypti

The Dengue virus belongs to the Flaviviridae family. DENV-1, DENV-2, DENV-3, and DENV-4 are the four serotypes of the virus. In Bangladesh DENV-2 and DENV-4 are found but in recent years the prevalence of DENV-3 is also observed. It is believed that recovery from infection gives lifelong immunity against that particular serotype. But immunity against other serotypes may not be gained. If a person is infected by different serotypes of the virus, then they are more likely to suffer from severe dengue which has many negative health consequences.

1.1 Incidence of Dengue virus and global distribution

The cases of dengue occurrence have significantly increased in the last few decades. Since a big portion of the cases do not show symptoms or show mild effects and are taken care of at home, the actual numbers of this disease are not recorded.

The incidence of dengue that has been reported to WHO has incremented almost 8 times in recent years. The significant changes are from 505,430 cases in 2000 to over 2.4 million cases in 2010. And in 2019 the number has increased to 5.2 million in 2019. In case of death due to dengue, it is 960 people in 2000 to 4032 people in 2015.

The most dengue burden is seen in Asia which accounts for almost 70 percent of the total global dengue burden despite a risk of infection that exists in almost 129 countries of the world. One scientific study indicates 390 million dengue virus incidences per out of which 96 million leads to the severity of the disease. From another research on the prevalence of dengue, it is reported that about 3.9 billion people are at risk of being infected with the dengue virus.

Figure 2: Aedes albopictus.

1.2 Epidemiology

The explosive outbreak of Dengue is observed only in the past few decades. The virus is now widespread in over 130 countries whereas just before 1970 it was restricted to only 10 countries. In Africa, the Eastern Mediterranean, South-East Asia, and the Pacific the Dengue has taken an Endemic form. The most severity is observed in Asia. Even in Europe, the dengue virus is causing outbreaks in recent years. Bangladesh, Brazil, Cook Islands, Ecuador, India, Indonesia, Maldives, Mauritania, Mayotte (Fr), Nepal, Singapore, Sri Lanka, Sudan, Thailand, Timor-Leste, and Yemen were affected by Dengue in 2020 along with Covid-19. Brazil, Bangladesh, Cook Islands, Colombia, Fiji, Kenya, Paraguay, Peru, and Reunion Island are still facing Dengue in 2021. dengue is the second most diagnosed cause of fever after malaria, among travelers returning from low- and middle-income countries.

The highest number of dengue cases ever reported globally was in 2019. All WHO regions were affected, and dengue transmission was recorded in Afghanistan for the first time and a large number of cases were reported in Bangladesh (101,000), Malaysia (131,000), Philippines (420,000), Vietnam (320,000) in Asia. In 2016 Brazil had approximately 1.5 million dengue cases which is 3 times higher than in 2014. It had 1032 deaths recorded that year.

1.3 Transmission

The main role for spreading the disease is contributed by the female Aedes aegypti mosquito. The other species of the Aedes genus also can carry the virus but their role is not significant. There is a crucial stage in the life cycle of the virus. First, the mosquito bites a Dengue-infected individual. Here the virus is transferred to the mosquito. In the midgut of the mosquito the virus replicates and then spreads to the other tissues and the salivary glands. The period required for the virus starting from ingestion to transmit the disease to another new healthy person is referred to as the extrinsic incubation period (EIP). When the ambient temperature is about 25-28 °C the EIP is around 8-12 days. The extrinsic incubation period is affected by several factors. Virus Genotype, initial viral concentration, the magnitude of daily temperature fluctuations can also cause the EIP to alter. When the mosquito becomes infectious it has the potential to spread the disease until it dies.

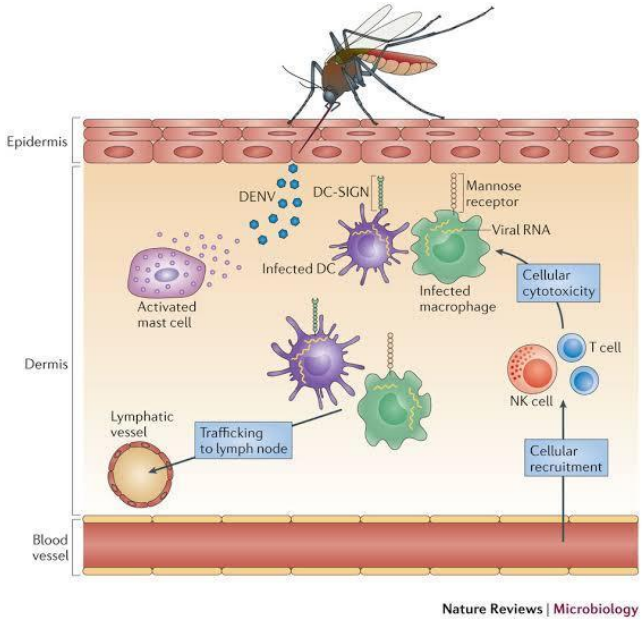


Figure 3: Mode of action of Dengue virus

Individuals who are viremic with DENV can cause the mosquito to be infected. All asymptomatic, pre-symptomatic, and symptomatic patients can spread the virus through a mosquito. This human-to-mosquito transmission of the virus takes place up to 2 days before someone shows any symptoms of dengue to 2 days after the symptoms are gone. The presence of a virus in the blood of the host is called viremia. Usually, people remain viremic for 4-5 days but sometimes it can last

up to 12 days. Human mannose-binding receptor (MR) and DC-SIGN on macrophages as primary receptors used by the Dengue virus to bind to host cells.

Evidence of Maternal transmission (pregnant mother to her baby) is found even though the primary mode of transmission of DENV amongst humans requires mosquito vectors. Although vertical transmission cases are few, the infected victims face serious health issues including fetal distress, low birth weight, and pre-time birth.

1.4 Breeding of mosquito

Considered as the primary carrier of dengue the Aedes aegypti mosquito lives in urban environments and grows mainly in man-made containers. This mosquito feeds quite frequently during the egg-laying period and it mostly feeds at day time. The early morning and in the evening before sunset is the peak biting time of the mosquito. The egg laid by the mosquito remains viable for a few months. The eggs will hatch when they come in contact with water.

Due to the international trade of used tires (a breeding habitat) and other goods like bamboo dengue has spread over 25 countries of the European region and 32 states in the USA. In Asia, Aedes albopictus is the secondary carrier of dengue. As Ae. Albopictus is highly adaptive, its geographical dissemination is mainly due to cold tolerance.

1.5 Signs and symptoms of dengue

Dengue can affect infants, young individuals, and adult people. It is a severe flu-like disease. After the bite, the virus incubates in the host for 10 to 14 days. The actual severe symptoms are observed in patients for 5-7 days. According to the World Health Organization Dengue has been classified into major categories which are Dengue with or without warning signs and Severe Dengue. Physicians treat Dengue patients by observing certain symptoms. If the patient's health condition deteriorates, they are admitted to the clinic for close observation and to minimize the risk of developing severe Dengue.

Dengue can be suspected if a person shows certain symptoms. During the febrile phase is if there is a high fever (40°C/104°F) along with 2 of the following mentioned symptoms then Dengue can be suspected in the person. The Symptoms are

- Pain behind the eyes
- Nausea
- Vomiting
- Swollen glands
- Rash
- Severe headache
- Muscle and joint pains

About 3-7 days after the onset of illness, the patient enters the critical phase. It is at this time the patient may manifest Severe Dengue. Many potentially fatal complications are associated with Severe Dengue like plasma breaking, respiratory distress, and many others.

In this condition, the Doctors observe the following symptoms

- blood in vomit.
- Fatigue
- rapid breathing
- bleeding gums
- severe abdominal pain
- persistent vomiting
- restlessness

Patients showing these symptoms should be under close medical observation for the next 24-48 hours to avoid the risk of death.

1.6 Diagnosis

There are multiple means for detecting the infection of the Dengue virus. Depending upon the time of the test type may vary. If the test is done in an earlier period then virological tests are used. By this method, the presence of a virus can be detected in the patient. Usually, in this case, a Reverse transcriptase-polymerase chain reaction is conducted. Moreover, another commercially available diagnostic test is often used which gives results faster (20 minutes). Here the protein produced by the virus NS1 is detected by the test which confirms the presence of the virus in the host.

Other serological methods are also used to detect the virus. Enzyme-linked immunosorbent assay (ELISA) is employed to find the IgG and IgM antibodies. Just after one week of infection, IgM can be found in the patient's body. Between 2 to 4 weeks from the onset of infection, the highest amount of IgM is present. This IgM will remain in the body for 120 days. IgG takes a bit longer time to grow compared to IgM. This antibody remains in the body for years. IgM indicates a recent infection and IgG reflects a past infection.

1.7 Cure of infection

Currently, there is no cure or specific treatment for Dengue infection. Since high fever, pain is associated with this disease paracetamol and acetaminophen may be administered to minimize the sufferings of the patients. NSAID (non-steroid anti-inflammatory drugs) like ibuprofen, aspirin must be strictly avoided as these cause thinning of blood, and for dengue patients, it may increase the risk of hemorrhage thus worsening the condition. For severe dengue patients, an ample amount of fluid must be supplied. Since during infection the platelet count drastically decreases so platelets can be administered by the recommendation of an authorized doctor.

1.8 Vaccine

The biological process which activates acquired immunity against a specific infectious disease is a vaccine. The vaccine acts as an agent in the body that stimulates the immune system. The surface protein, modified toxin, weakened form of the microorganism that resembles the pathogen is contained within the vaccine. In the case of infectious microorganisms, the vaccine is prophylactic which means it will prepare the body in a way that it will be able to defend itself if any

microorganism enters the body. The vaccine triggers the immune system such that it recognizes the foreign microbes as a threat and then kills it and be ready to encounter them in case of future invasion. Vaccination is the systematic administration of vaccines. To prevent the transmission of infectious diseases, the vaccine is the most efficient tool. In the past, by the virtue of vaccines scientists were able to eradicate smallpox, polio, influenza, measles, and many more from the surface of the earth. According to the world health organization, about 26 licensed vaccines are available against different preventable diseases. A French multinational biotech company Sanofi Pasteur developed Dengvaxia, a vaccine against the Dengue virus. The vaccine was licensed in December 2015. Now it has been approved by the regulatory authorities in more than 20 countries of the world. But the outcome of Dengvaxia was not so promising as some reports claim the individuals that have been given this vaccine without prior infection have shown to develop severe Dengue. That is why this vaccine is currently administered in people dwelling in the endemic territory, from age 9 to 45, and have at least one previous record of infection.

1.9 Reverse Vaccinology

Reverse vaccinology is a new term in the field of vaccine development. Mostly due to vaccine design against Covid-19 this method of vaccine development gained popularity. This novel vaccine-making process was initiated by Whole Genome Sequencing. When the scientists came to know the entire genome of a bacterium or a virus, they started creating new techniques to identify the epitope parts of the proteome of the pathogen and eventually could determine suitable candidates for vaccine development.

Reverse vaccinology is a part of vaccinosis. It involves the use of bioinformatics techniques to screen the whole genome of the microorganism. By using computational calculations, the genes that express epitopes can be detected. These epitopes are the peptides that are part of the virus or bacteria (antigen) where the antibody of the host body ligates. So, if the epitope can be predicted accurately then-candidate vaccine development is much more specific and fast. After that for validation, these peptides can be synthetically produced and injected into animal models and primates to test the efficacy of the vaccine. The first vaccine that was developed by this approach was against the Serogroup B meningococcus to prevent meningitis. Even though there had been many experiments and trials this method of vaccine development was very fast compared to

traditional vaccine development which takes decades to be available for patients. The structure-based vaccines and reverse vaccinology depend on the principles and the ability of the target antigen (proteins, peptides) to stimulate the body to get an antibody response when used as an immunogen.

Multiple series of biological, molecular, physiological, genetic, and other processes must be activated in an individual to build a protective response against a particular microbe. In this process, antigen recognition, processing, presentation, and stimulation of innate, adaptive, and cell-mediated responses must take place. The suppression and repression of a certain gene and protein transcription, expression, secretion, and function are necessary for a person to attain immunity. The formation of genes involved in pattern recognition, attaching, and cell entry, the processing and presentation of cytotoxic T cells and B cells, complement system, and other immunoregulation are crucial in reverse vaccinology.

Advantages of reverse vaccinology are as follows:

- It is a data-driven approach.
- More accurate target molecules can be used (spike protein, envelope protein).
- The knowledge of the genetic sequence and proteomics can be employed.
- Computational calculations and predictions are more accurate.
- The process of designing the vaccine construct is very fast compared to conventional methods.
- The results of this process are more efficient. By using machine learning and artificial intelligence this process one can predict the exact protein sequence or peptides that are antigenic, non-allergic and non-toxic.
- The physio-chemical properties can be tested by bioinformatics servers.
- Using various software protein-protein interactions can also be predicted.
- This method is extremely time-efficient.

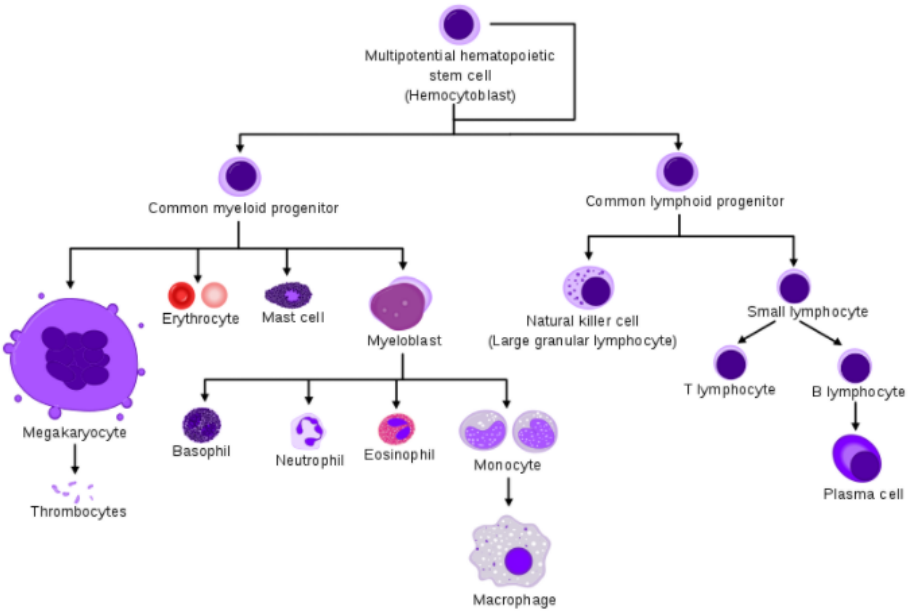


Figure 4: Production and maturation of different immune cells from Multipotent hematopoietic stem cells.

For a vaccine to elicit a proper immune response the immune cells of the body like B cells, T cells, and other immunoregulatory cells must be activated to remove the pathogen. For this proper insight into the mechanism of the immune cell, interactions are extremely crucial.

1.10 Role of B cell to develop immunity

B cells are produced in the bone marrow throughout the life but in the prenatal condition, it is formed in the fetal liver. For the adaptive humoral system, the role of B cells is pivotal. B cells are responsible for inducing the formation of antigen-specific immunoglobulin (Ig) that are directed against invasive microorganisms. For the development of B cells, the rearrangement of the Ig heavy chain occurs first. This starts the D-J recombination that takes place in the common lymphoid progenitors and pro B cells. Both antigen-dependent and -independent phases of selection are experienced by B cells that are strictly regulated by cell signaling. A mature category of B cells called B1 cells are found in antigen-dependent and -independent phases of selection, the peritoneal cavity, and pleural cavities and intestines. In between the secondary lymphoid organs, the mature B cells recirculate to find antigens. If the B cells find antigens it receives assistance from T cells and enters a couple of different developmental possibilities. Initially, the cells may go through plasmacytic differentiation. This leads to the formation of extrafollicular plasmablasts

and forms IgM-secreting plasma cells. These cells provide a rapid initial response to antigen but since they do not have somatically mutated Ig genes, these are short-lived. The other developmental possibility is the formation of a germinal center. Formed in the secondary lymphoid tissues, the germinal center produces long-lived antibody-secreting plasma cells and memory B cells, which can protect against reinfection. The cells will eventually proceed to class-switching recombination. To diversify B cell responses, immunoglobulin class switching to IgG, IgA, and IgE is a major mechanism to match antibodies with antigens. The memory B cells and plasma cells exit the GC along with high-affinity BCRs of switched isotypes.

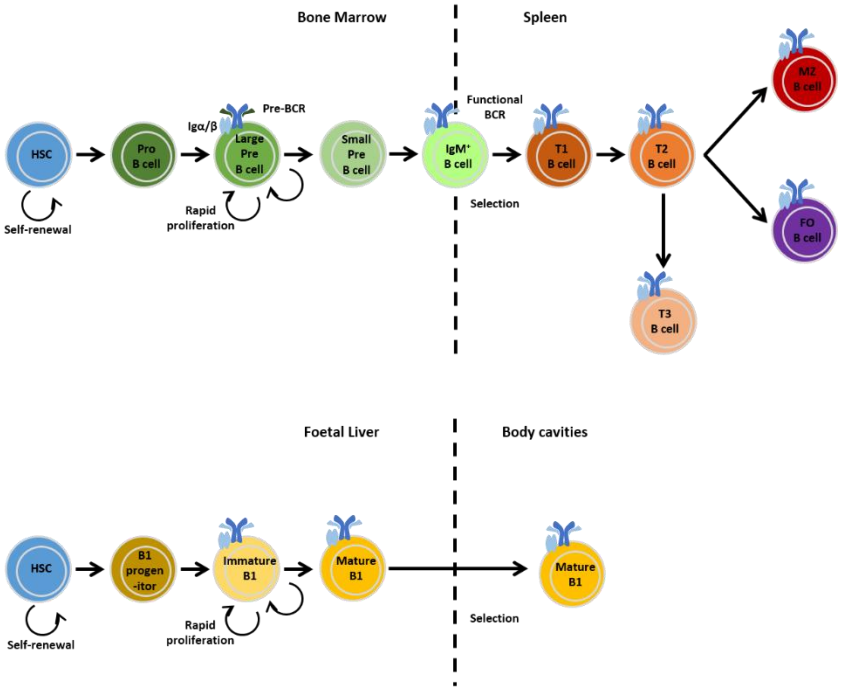


Figure: 5 Maturation of B cells**.**

For the preparation of adaptive CD4+ T cells, B cells have an important contribution. Referred to as B10 Bregs, the Interleukin (IL-)10-secreting B cells with suppressive functions in reducing disease severity in animal models.

1.11 Role of T cell to develop immunity

T cells, a major component of the adaptive immune system, plays an important role in directly killing infected host cells, producing cytokines, activating other immune cells, and regulating the immune response. The originating source of T lymphocytes is the hematopoietic stem cells within the bone marrow. It travels by blood to the thymus from the bone marrow after becoming lymphoid progenitor cells. When the cells reach the thymus it takes part in a selection process in which most of the cells die. The developed T cells called thymocytes and the ones with receptors for self-antigens will receive negative signals and eventually will be removed from the repertoire (thymus).

Every T lymphocyte has a T cell receptor (TCR) that is specific against a particular antigen and they circulate through peripheral lymphoid organs, ready to encounter their cognate antigens and become activated. As people age the thymus atrophies and so produces fewer naïve T lymphocytes over time.

The T cells that do not have encountered any antigen are called a Naïve T cell. It can interact with antigen-presenting cells (APCs), which use MHC molecules to present antigen in the peripheral lymphoid organs. They proliferate and differentiate into one of several effector T lymphocyte subsets when the T lymphocytes recognize their specific antigens. Coreceptors are used by T lymphocytes to bind to MHC molecules. CD4 or CD8 are coreceptors and the CD proteins help to differentiate major groups of effector T lymphocytes. Naïve CD8+ T lymphocytes will be matured to cytotoxic T lymphocytes and CD4+ T lymphocytes will become T helper lymphocytes.

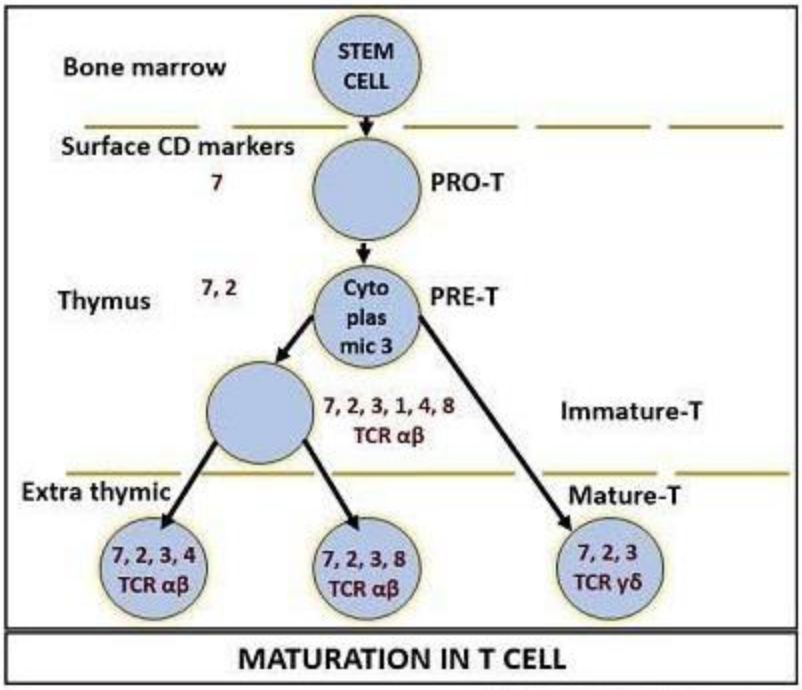


Figure 6: Maturation of T cells

Cytotoxic T cells need several signals from other cells like dendritic cells and T helper cells to be activated and MHC Class I molecules interact with CD8 on the cytotoxic T cells. By releasing cytotoxic granules into the target cells cytotoxic T lymphocytes kill their target cells.

Due to the difference of cytokines, they synthesize and release CD4+ helper T cells (Th) that can be divided into two subgroups. T-helper cells can differentiate into many subgroups i.e.Th1, Th2, Th17, and other regulatory T cells and have a wide range of effector functions. For T helper cells to be activated they need to be presented with peptide antigens on MHC Class II molecules. The functions of CD4+ T cells are to release cytokines, activate other immune cells, and assist B cells to produce antibodies. IL-3 and TNF are secreted from both TH1 and Th2. Th1 triggers a cellular immune response and takes part in the inhibition of macrophage activity as well as stimulates B cells to form IgM. On the other hand, Th2 mediates the differentiation and proliferation of mast cells (IL-3, IL-4). It is also involved in the differentiation and proliferation of eosinophilic leukocytes (IL-5). The main task of Th2 is to stimulate the humoral immune system, promote B cell proliferation, and induce antibody production (IL-4).

1.12 Envelope protein of Dengue virus

The genetics of the Dengue virus is composed of a single-stranded RNA genome of positive polarity. The RNA is translated in the cytoplasm in the form of a single polyprotein. Being processed by cellular and viral proteases, the polyprotein produces three structural and seven nonstructural mature proteins. The capsid protein, the glycosylated envelope protein, and the precursor of membrane protein are the three structural proteins. The virus assembles in the endoplasmic reticulum. The RNA-C complex and the endoplasmic reticulum lumen are attached and thus it acquires the lipid bilayer, envelope protein, and the pre-membrane protein. The formation of the mature viral particle occurs by Furin-mediated proteolysis of prM in the trans-Golgi network that triggers rearrangement and also influences the homodimerization of envelope protein. The interaction of envelope protein and host cell receptors leads the Dengue virus to an endocytic pathway for entry into the host cell. The acidic condition in the endosome causes a significant conformational change in the envelope protein that induces the fusion of viral and host cell membranes. This mechanism leads to the release of the viral genome into the cytoplasm of the host cell. Initially, we considered a combination of capsid protein, pre-membrane protein, and envelope protein as the target molecule against which the vaccine will be designed. But considering the functionality and role of the envelope protein, we select only the envelope protein as the targeted sequence. The functions of envelope protein are a fusion of virus membrane with host endosome membrane, virus-host interaction, viral attachment to host cells.

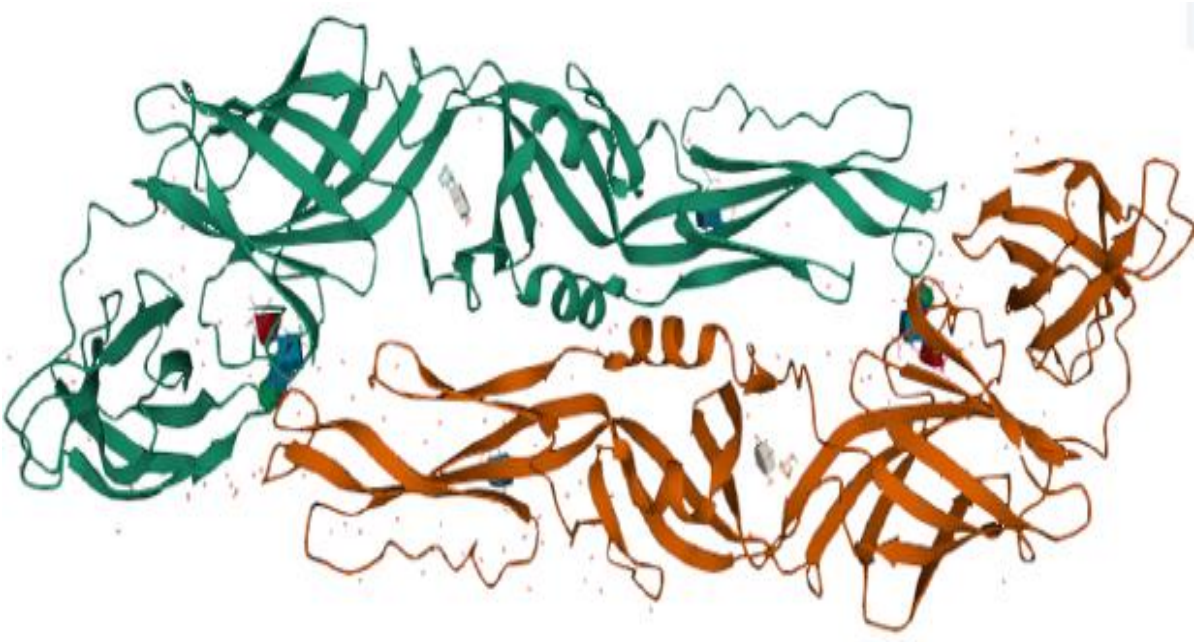


Figure 7: Three-dimensional structure of envelope protein of Dengue virus from PDB

Chapter 02

Methodology

A literature review was done extensively to design the protocol. UniProt and NCBI both were used to analyze the whole proteome. I checked the sequence from NCBI but used UniProt sources as it is only dedicated to Protein Database. The envelope protein sequence of Dengue virus of protein ID D6MQ38 has been retrieved. There are four serotypes of the virus, and in Bangladesh, the serotypes DENV 2, DENV3, DENV4 are mostly found and causing diseases, so we retrieved the sequence of envelope proteins of those serotypes as well. Amongst the sequences, we tried to collect strains from Singapore, Malaysia, Thailand, and Bangladesh. Initially, sequences of 10 strains were selected.

Multiple sequence alignment was done to check and ensure that the sequence that will be targeted is conserved among all strains and serotypes. In this step, Muscle was employed to do the multiple sequence alignment. ChimeraX 1.2.5 version was downloaded and installed on my computer. The purpose of this step was to visualize the target protein in different configurations. Vaxigen 2.0

server was used to check the antigenicity of those sequences. The sequence was antigenic so we proceeded to the next step. Next, we worked with Cytotoxic T cell epitope prediction. NetCTL 1.2 server was employed to find the Cytotoxic T cell epitope. All supertypes A1, A2, A3, A24, A26,B7,B8,B27,B39,B44,B58 and B62 were tested in the server. All the parameters are set to default except the score option. We selected the combined score option. Then only the epitope peptide sequences were selected. The epitope peptides were marked by E.

Again, Vaxigen 2.0 server was used to check antigenicity. These sequences were tested separately. Both antigenic and non-antigenic sequences were determined. We chose only the antigenic sequences. The selected antigenic sequences were saved in a new file in Notepad. Next, the selected antigenic sequences were tested whether they are toxic or non-toxic. For this purpose, the ToxinPred server was employed. The chosen sequence was pasted in the box and then we clicked the run analysis option keeping other parameters as default settings.

We found many sequences to be toxic so we meticulously chose only the non-toxic sequences. Next, the immunogenicity test was done with the help of IEDB (http://tools.iedb.org/immunogenicity/). Here the MHC I was immunogenicity was tested. The result was downloaded and saved in EXCEL format.

The high score epitopes were selected as the high score has a greater probability to give a better immune response. Now the Helper T cell epitope was predicted. We used the (http://tools.iedb.org/mhcii/) IEDB MHC II server to run the data. The length that was selected was 15 and the rest of the parameters were as default settings. We have also given the email address to get the result. The result of the test was saved in excel format.

The epitopes with lower percentile rank were chosen. Next, the allergenicity test was done with the help of AlgPred: http://crdd.osdd.net/raghava/algpred/. The protein sequences were tested individually. Many of the sequences were allergenic, so we eliminated them. Only the non-allergenic sequences were selected for further research.

Again, the selected sequences were tested individually for antigenicity. This time it is for HTL epitopes. So, Vaxigen 2.0 server was employed. The non-antigenic sequences were removed. The antigenic sequence was saved in a new file of Notepad.

Next, the newly selected sequences were tested for toxicity. ToxinPred server (https://webs.iiitd.edu.in/raghava/tox) was used to choose only the nontoxic sequences. After that, we have to eliminate the overlapping sequences with the help of the IFNepitope server (https://webs.iiitd.edu.in/raghava/tox) The positively marked sequences are non-overlapping.

We used ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/index.html) to determine the B cell epitopes. The sequence was previously checked as antigenic. Then we checked toxicity by ToxinPred server (https://webs.iiitd.edu.in/raghava/tox). and Allergenicity by AlgPred: (http://crdd.osdd.net/raghava/algpred). And the selected sequence that passed the criteria and parameters was saved in a Notepad file.

Now we chose 10 CTL epitopes and 3 HTL epitopes that have fulfilled all the criteria and parameters. The molecular weight of the vaccine molecule should be 30 to 70 kilo Dalton. We added Adjuvant GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRK, upstream of the sequence.

We join the adjuvant and B cell epitope sequence with an EAAAK linker. The B cell epitope sequences are linked by GPGPG linkers. B cell epitope and HTL epitope are linked by GPGPG linker.

The 3 HTL epitope sequence linkers are also joined by GPGPG linkers. The HTL and CTL are joined by the GPGPG linker. For the rest of the CTL epitope sequences, AAY linkers are used to ligate them in the vaccine construct.

The final vaccine constrict is saved in a new Notepad file. Next, I employed the ProtParam tool of ExPASy server to determine the physicochemical properties of our newly designed vaccine. The result and analysis are saved in a separate word file.

For the secondary structure prediction, PSIPRED server was employed. The result was saved for the interpretation of the structure of the protein.

Chapter 03

Result:

The envelope protein sequence was conserved and no significant mutation was observed among the selected strains. The chosen sequence was computed in Vaxigen server and it was revealed that the sequence was antigenic. This indicated that the sequence should be a good target to design a vaccine candidate. The B cell epitope was predicted by ABCPred and 54 epitopes show in the result and only10 epitopes with the highest score was selected for the next step.

The IEDB server predicted the CTL and HTL epitopes. From the result of the data run 39 CTL epitopes were predicted and only 24 highest score epitopes were selected. In case of HTL epitopes the sequences of the highest percentile rank were chosen.

All of these chosen sequences were subjected to allergenicity, antigenicity, and immunogenicity tests. The sequences that passed the criteria were further selected based on their score.

The B cell epitopes, CTL epitopes, and HTL epitopes were ligated with linker protein sequences. An adjuvant was introduced in the vaccine construct to elicit better immune response.

The following is the table showing the selected epitopes in my vaccine construct.

PROTEIN SEQUENCE TYPE	**Sequence**	**Linker**
Adjuvant	GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRK	EAAAK
B CELL EPITOPE	HGTILIKVEYKGEDAP	GPGPG
B CELL EPITOPE	GRLITANPVVTKKEEP	GPGPG

B CELL EPITOPE	TTMAKNKPTLDIELQK	GPGPG
B CELL EPITOPE	SGATTETPTWNKKELL	GPGPG
HTL EPITOPE	GARRMAILGDTAWDF	GPGPG
HTL EPITOPE	ARRMAILGDTAWDFG	GPGPG
HTL EPITOPE	RMAILGDTAWDFGSV	GPGPG
CTL EPITOPE	IVIGIITLY	AAY
CTL EPITOPE	VVQYENLKY	AAY
CTL EPITOPE	KLELKGMSY	AAY
CTL EPITOPE	ALFSGVSWI	AAY
CTL EPITOPE	WIMKIGIGV	AAY
CTL EPITOPE	IVIGIITLY	AAY
CTL EPITOPE	KLELKGMSY	AAY

CTL EPITOPE	VTFKNAHAK	AAY
CTL EPITOPE	WMVHRQWFF	AAY
CTL EPITOPE	QYENLKYTV	

Table 1: Demonstration of B cell, CTL, and HTL epitopes that are chosen for the vaccine design.

Vaccine Sequence

10 20 30 40 50 60
GIINTLQKYY CRVRGGRCAV LSCLPKEEQI GKCSTRGRKC CRRKEAAAKH GTILIKVEYK

70 80 90 100 110 120
GEDAPGPGPG GRLITANPVV TKKEEPGPGP GTTMAKNKPT LDIELQKGPG PGSGATTETP

130 140 150 160 170 180
TWNKKELLGP GPGGARRMAI LGDTAWDFGP GPGARRMAIL GDTAWDFGGP GPGRMAILGD

190 200 210 220 230 240
TAWDFGSVGP GPGVVQYENL KYAAYKLELK GMSYAAYALF SGVSWIAAYW IMKIGIGVAA

250 260 270 280 290 300
YIVIGIITLY AAYKLELKGM SYAAYIVIGI ITLYAAYVTF KNAHAKAAYW MVHRQWFFAA

310
YQYENLKYTV

Number of amino acids: 310

Molecular weight: 33435.86

Theoretical pI: 9.49

Amino Acid Composition

Amino acid	Number of Residues	Percent Composition
Alanine	37	11.9%
Arginine	14	4.5%
Aspartic Acid	7	2.3%
Asparagine	8	2.6%
Cysteine	6	1.9%
Glutamic Acid	6	1.9%
Glutamine	14	4.5%
Glycine	43	13.9%
Histidine	3	1.0%
Isoleucine	22	7.1%
Leucine	21	6.8%
Lysine	24	7.7%
Methionine	8	2.6%
Phenylalanine	7	2.3%
Proline	20	6.5%
Serine	8	2.6%
Threonine	19	6.1%
Tryptophan	8	2.6%
Tyrosine	20	6.5%
Valine	15	4.8%
Pyrrolysine	0	0 %
Selenocysteine	0	0%

Table 2: Amino acid composition of the epitopes that are chosen for the vaccine construct.

Total number of negatively charged residues (Asp + Glu): 22

Total number of positively charged residues (Arg + Lys): 38

Atomic composition:

Carbon	**C**	**1525**
Hydrogen	H	2367
Nitrogen	N	403
Oxygen	O	415
Sulfur	S	14

Chemical Formula: $C_{1525}H_{2367}N_{403}O_{415}S_{14}$

Total number of atoms: 4724

Extinction coefficients:

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

Ext. coefficient 74175

Abs 0.1% (=1 g/l) 2.218, assuming all pairs of Cysteine residues form cystines.

Ext. coefficient 73800

Abs 0.1% (=1 g/l) 2.207, assuming all Cysteine residues are reduced.

Estimated half-life:

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

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 29.60

This classifies the protein as stable.

Aliphatic index: 80.06

Grand average of hydropathicity (GRAVY): -0.105

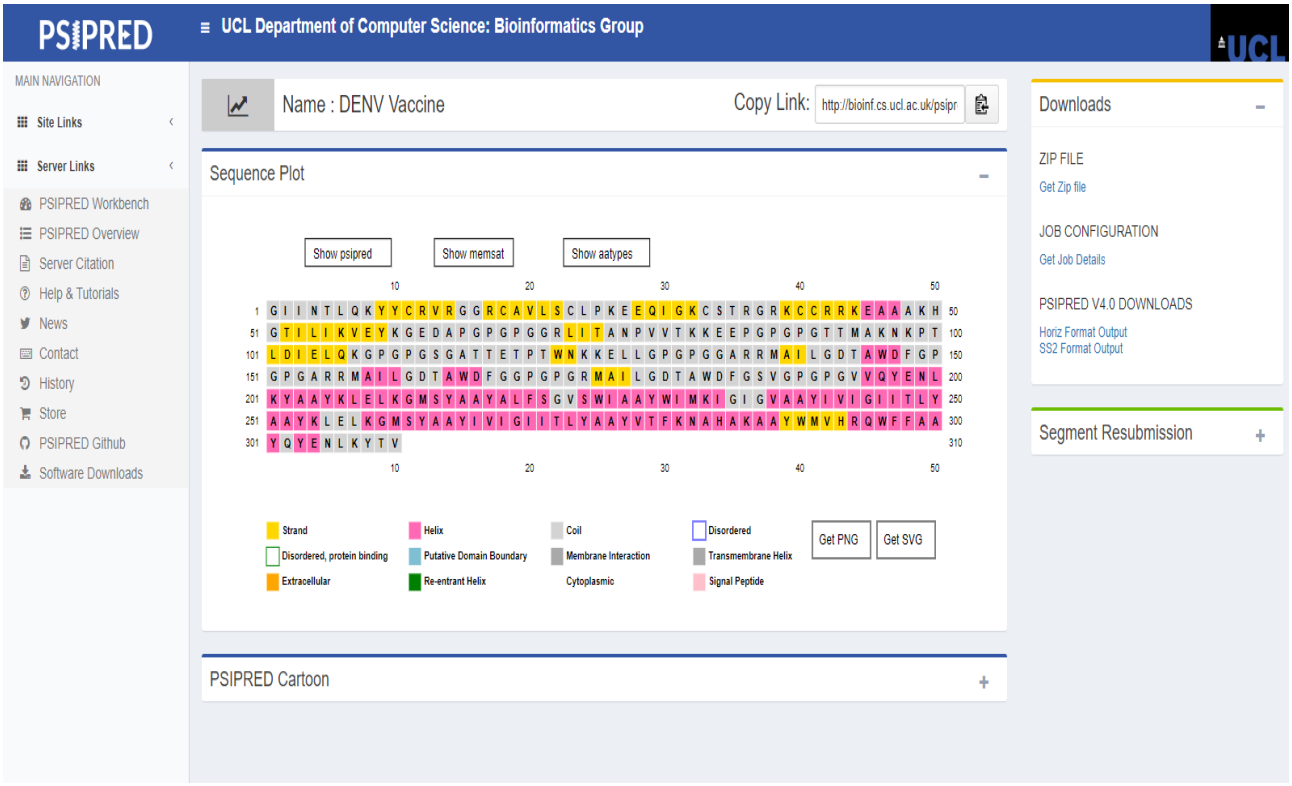


Figure 8: Secondary structure prediction by PSIPRED

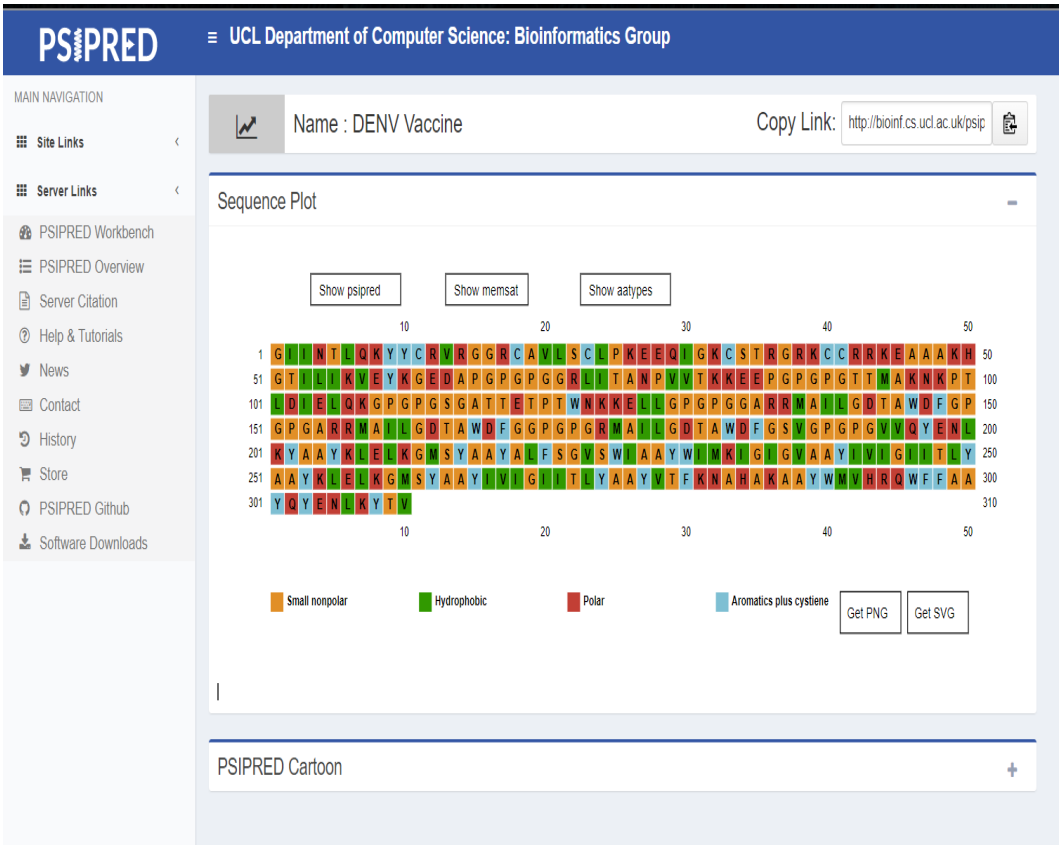


Figure 9: Characteristics of the Amino Acids of the epitopes of the vaccine. The yellow color indicates nonpolar; the green color indicates hydrophobic; the red color shows polar Aromatics plus cysteine.

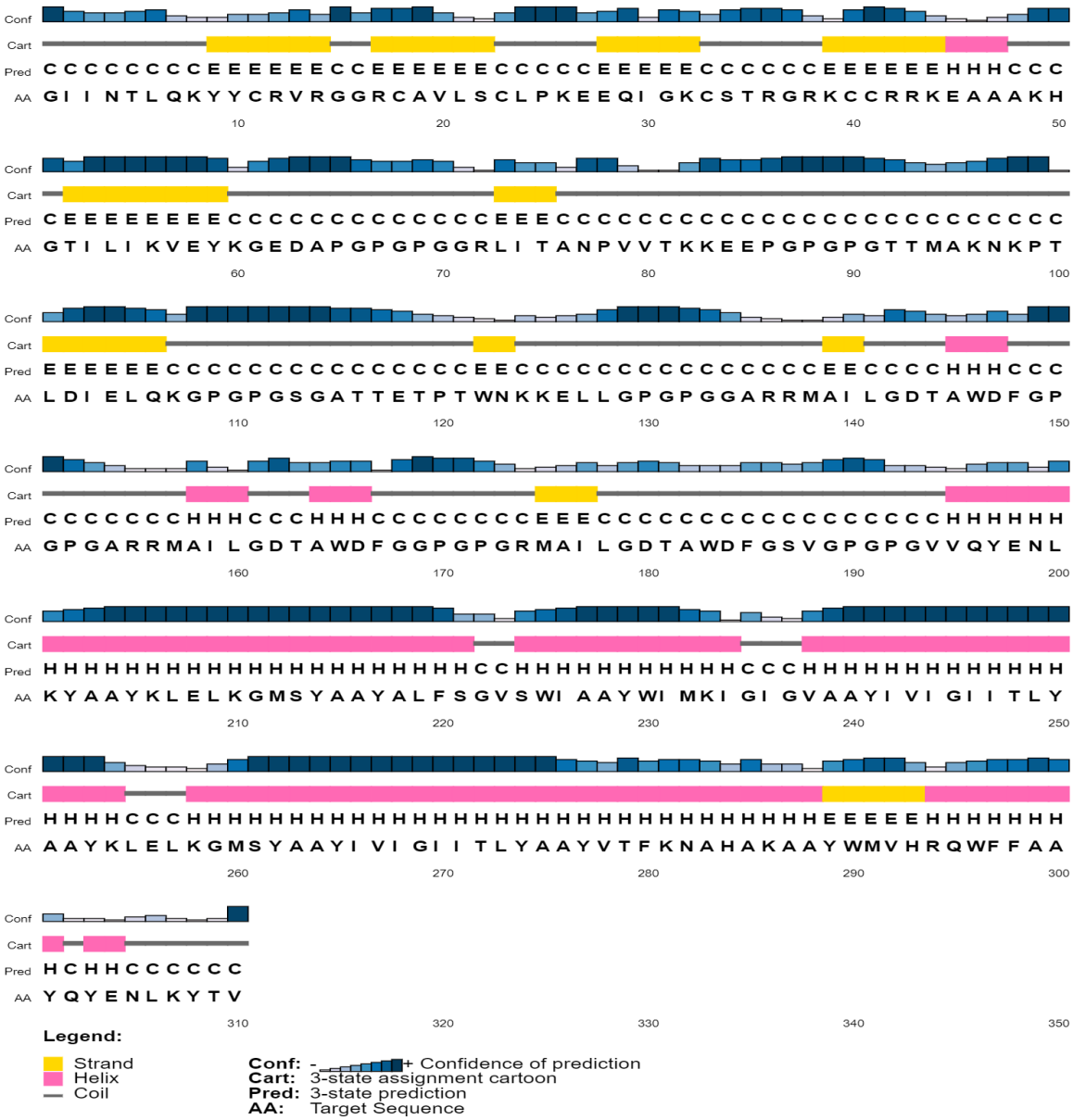


Figure 10: Image of vaccine showing the Strand, helix and coil parts of the epitopes. Here the yellow part shows strand; pink shows helix and gray line indicates coil.

Chapter 04

Discussion

Compared to drugs for the management of viral diseases, vaccines have been more efficient and an improved option. Every year millions of infections of the dengue virus across the globe have brought unbearable pain in the life of human beings. Due to Dengue infection, many lost lives and also jeopardize the health system and economy in infected countries in a catastrophic manner. According to many health experts, mass vaccination along with vector control can bring positive outcomes in reducing the infection rate of such a viral disease. In this thesis project, we attempted to design a novel vaccine against the Dengue virus by the combination of bioinformatics and immunoinformatics.

To have a better understanding and observation of the envelope protein we used a biomolecular image visualization program called Chimera X. It is developed at the University of California, San Francisco by the Resource for Biocomputing, Visualization, and Informatics. By the use of Chimera X automatic identification of atoms, measurements: distances, hydrogen addition, and partial charge assignment angles, surface area, volume, amino acid rotamer libraries, protein Ramachandran plot, protein contact map can be done along with many other structural analyses. High-resolution images, interactive shadows, silhouette edges, multicolor backgrounds, standard molecular representations (sticks, spheres, ribbons, molecular surfaces), visual effects including depth-cueing are the features of Chimera X that allowed us to create images of our target DENV envelope proteins.

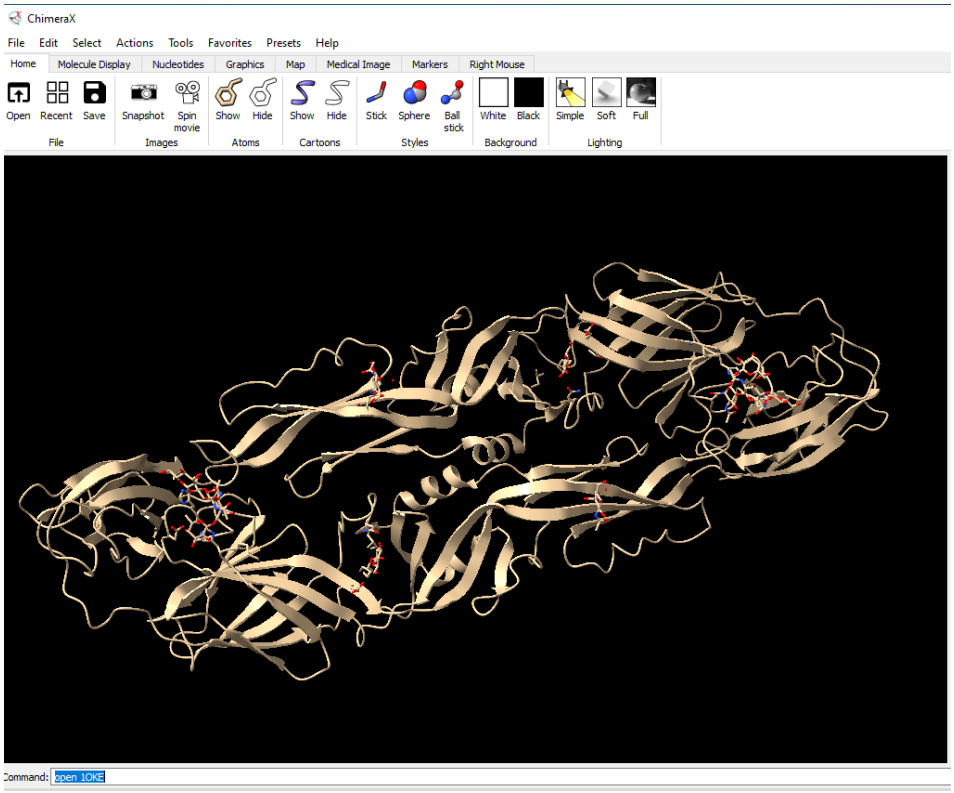


Figure 11: Image of envelope protein PDB ID 1OKE by CHIMERA X in Ribbon form.

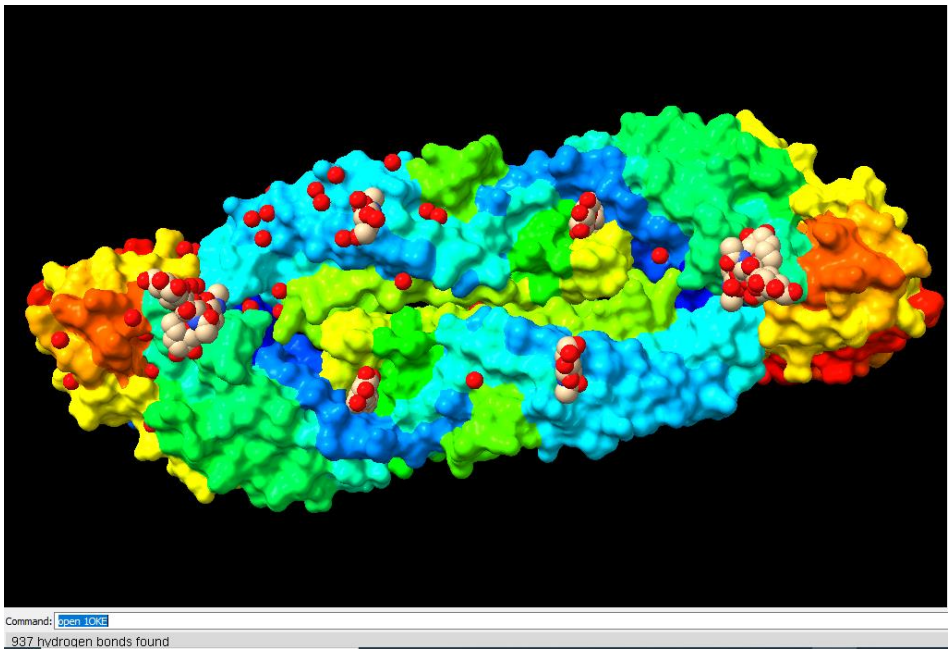


Figure 12: Chimera X (envelope protein) Dengue virus PDB ID 1OKE; Rainbow molecular display indicating 937 hydrogen bonds.

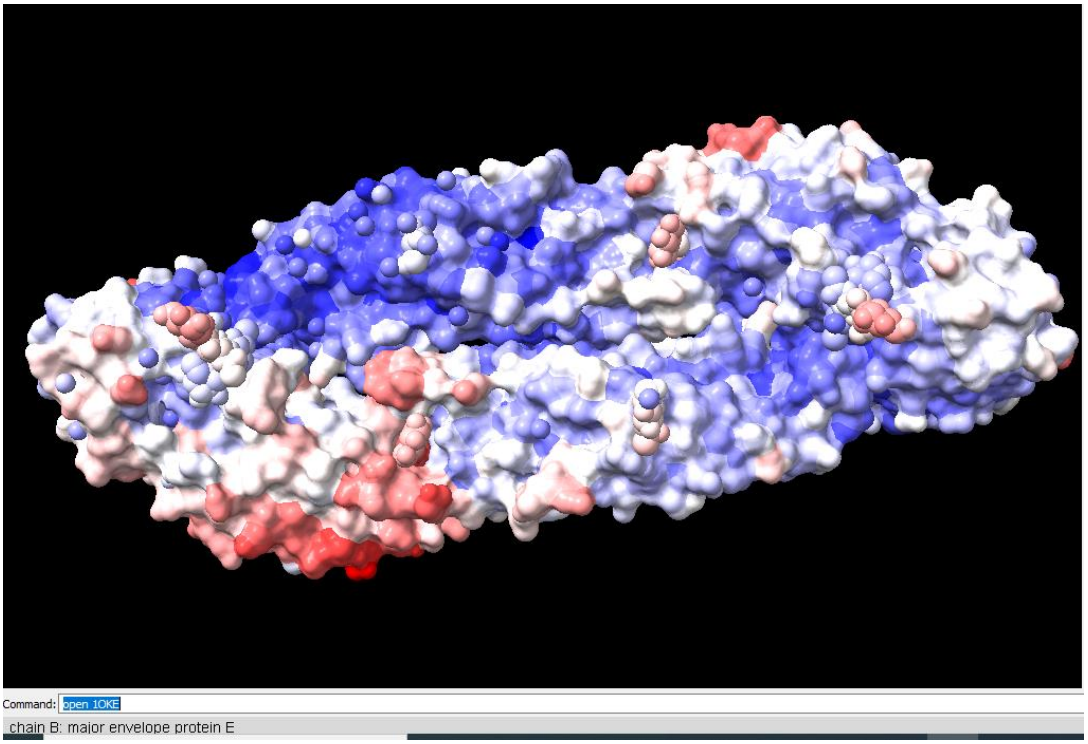


Figure 13: Chimera X (envelope protein) Dengue virus PDB ID 1OKE; Electrostatic molecular display

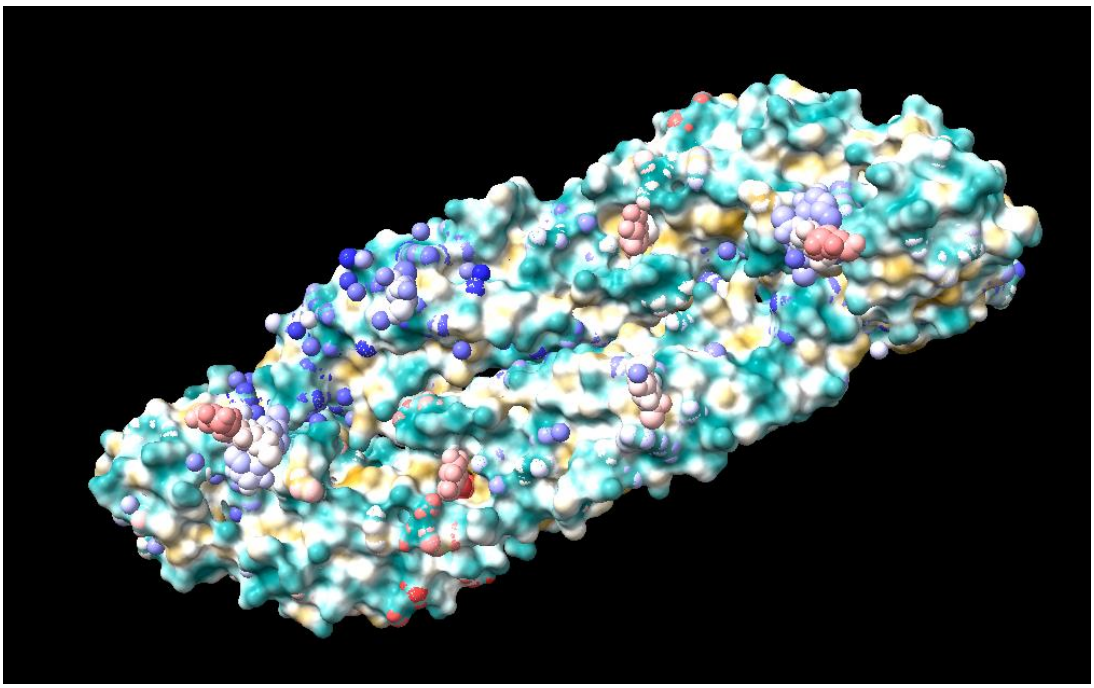


Figure 14: Chimera X (envelope protein) Dengue virus PDB ID 1OKE; Hydrophobic molecular display

The sample sequences that we chose were covering a large geographical region. Since Asia has the highest amount of Dengue infection per year our sample sequences were chosen based on the sequences collected from Asian countries. From the samples that we have collected, we performed a multiple sequence alignment test with the help of MUSCLE. MUSCLE is the abbreviated form of Multiple Sequence Comparison by Log- Expectation. The result of multiple sequence alignment revealed that all the selected sequences do not have a significant mutation. One out of four hundred ninety-three amino acid sequence mutation is observed. We worked with the part of the sequence that is conserved in all twenty strains.

First, we checked whether the whole sequence through vaxigen 2.0 server for the antigenicity test. For alignment-independent prediction of protective antigens, VaxiGen is the first server. This server can be utilized separately or in combination with alignment-dependent prediction protocols. The classification of antigens is exclusively dependent on the physicochemical properties of the protein. The sequence length is not a major factor for antigenicity determination for this server. Virus, Bacteria, and tumor sequences can be used for the prediction of antigenicity. Models derived by ACC pre-processing of amino acids properties are contained in the server. Derived from a principal component analysis (PCA) of 29 experimental or calculated physicochemical properties of the twenty naturally occurring amino acids the z descriptors are highly condensed. The descriptors have three principles where z1 works on hydrophobicity, z2 on steric properties, and z3 for the polarity of the amino acids. z descriptors are good predictors of MHC binding peptides and act for the main physicochemical properties important for the recognition of antigens. Irrelevant information, such as sequence length, and amplifying the class-discriminating properties were eliminated by the use of ACC (Auto cross-covariance) transformations. Protein sequences are uploaded as single or multiple files in plain or fasta format respectively and the results page reports antigen probability (as a fraction of unity) for each protein and a statement of antigen status.

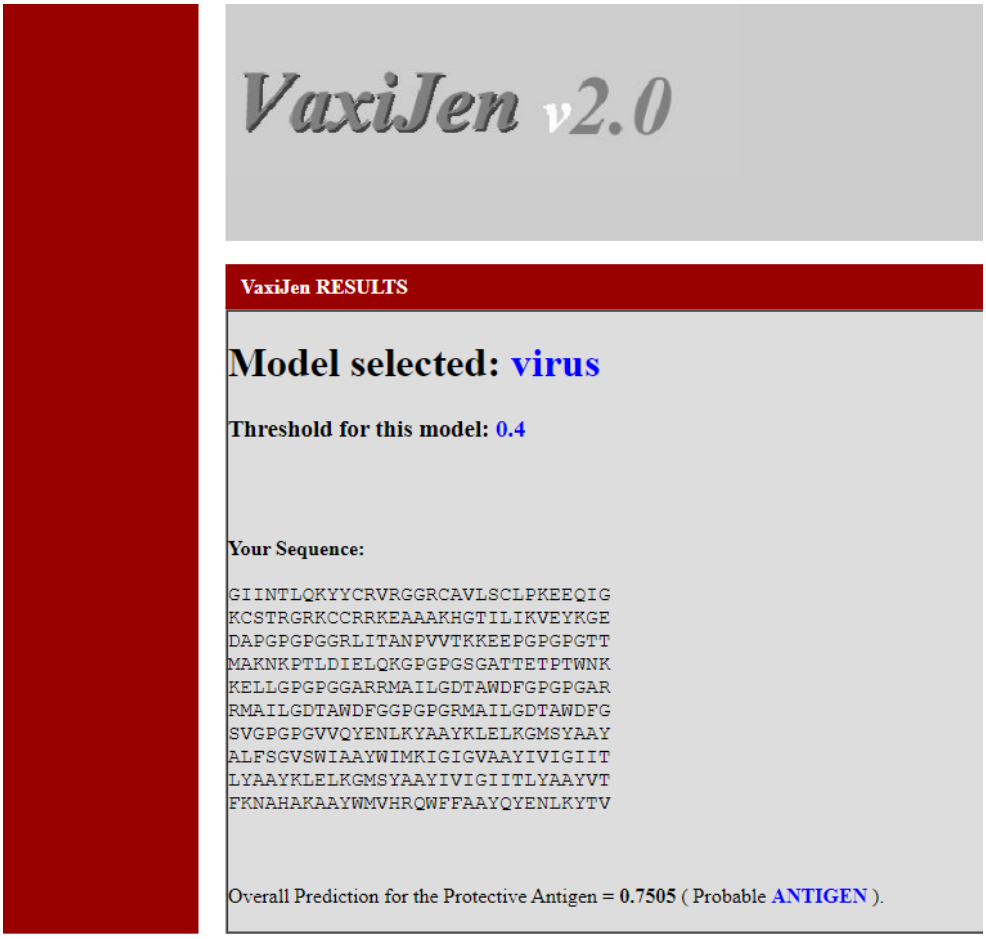


Figure 15: Result of epitope sequence indicating that it is antigenic in Vaxigen server.

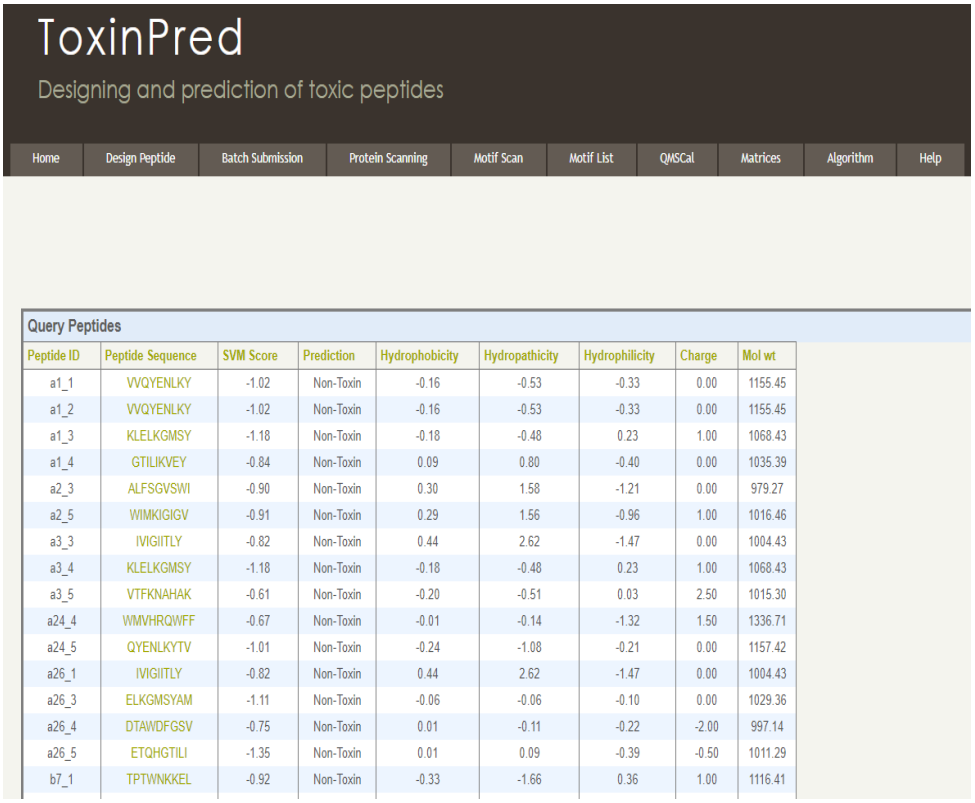
For the prediction of CTL epitopes, we employed NetCTL version 1.2, which is the updated form of 1.0. version. This updated version, NetCTL 1.2 has been trained on a set of 886 known MHC class I ligands and expands the MHC class I binding prediction to 12 MHC supertypes including the supertypes A26 and B39. The accuracy of the MHC class I peptide binding affinity is significantly improved compared to the earlier version and also the prediction of proteasomal cleavage has been advanced and is similar to the predictions received by the NetChop-3.0 server. The methodology involves the prediction of peptide MHC class I binding, TAP transport efficiency, and proteasomal C terminal cleavage. NetCTL 1.2 servers can carry on predictions of CTL epitopes up to 12 MHC class I supertype. MHC class I binding and proteasomal cleavage is performed using artificial neural networks and the TAP transport efficiency is calculated using a weight matrix. The prediction of the proteasomal cleavage is carried out with the assistance of

NetChop neural networks trained on C terminals of known CTL epitopes. The initial calculation of the neural network predicting MHC/peptide binding is shown as a log-transformed value which is interconnected to the IC50 values in nM units. The IC50 value denotes the certain concentration of a drug that decreases the activity (or binding) of another drug to an enzyme by 50% and under certain conditions, it can be used to demonstrate the affinity of the enzyme inhibitor. The marks from the three individual predictions are unified as a weighted sum with a relative weight on peptide/MHC binding of 1. In our data run we selected the combined score, weight on C terminal cleavage was 0.15, weight on TAP transport efficiency was 0.05, and threshold for epitope identification was 0.75. In the result sheet, the higher combined score epitope sequences were selected. Only the sequences showing E which stands for probable epitope were chosen for further analysis.

The vaccine that we have designed must be non-allergenic. So we checked our sequences to see whether they are allergenic or not. We employed AlgPred to carry on the allergenicity test. In the case of most vaccines, it is delivered to the body by injection and the most allergic responses take place in the mucosal membrane surface. Allergic reactions affect a large population across all geographic locations. Both intrinsic and extrinsic factors contribute to the development of allergic reactions and the mechanism of allergic reactions occurs by a series of complex reactions and interactions between the immune cells and the allergens. IgE antibodies are elicited when type 1 hypersensitivity reactions are mediated by certain allergens and the symptoms are rhinitis, asthma, and atopic eczema. More severe reactions such as acute and fatal anaphylactic shock can also occur due to allergic reactions. From the amino acid sequence of the protein, the Algpred server can predict allergenic proteins. Readseq program is used by the server and any standard format of protein sequences like EMBL, GCG, and FASTA or as plain text format is accepted by the server. BLAST search on ARPs, Hybrid approach, SVM-based method using the amino acid composition of protein, motif-based approach, and scanning of IgE epitopes are the options of method predicting allergens that can be selected by the researchers. We selected the SVM-based method using the amino acid composition of protein for our prediction purpose as the SVM module based on amino acid composition is slightly better than the dipeptide composition. The results are given

as potential allergens or non-allergenic along with prediction scores. After the test, we selected only the non-allergenic sequences and eliminated other potential allergenic sequences.

The toxicity profile of protein in a vaccine is of remarkable importance. The epitopes that we selected for our vaccine construct were nontoxic. If the vaccine is of toxic proteins it will lead to the production of a poisonous product rather than a beneficial drug. ToxinPred was employed to predict the toxicity of the selected epitopes. The ToxinPred is a server that uses an *in silico* method to predict and design toxic/non-toxic peptides. Consisting of 1805 peptides (<=35 residues) in the main Dataset, the server can also be used to design peptides of desired toxicity, protein scanning as well as a QSM calculator. While doing our toxicity test we uploaded our selected epitope sequences. We selected an SVM based method for prediction and physicochemical properties like hydrophobicity, charge, hydropathicity were displayed in the result. We chose only the non-toxic epitopes for the further research of our thesis project.

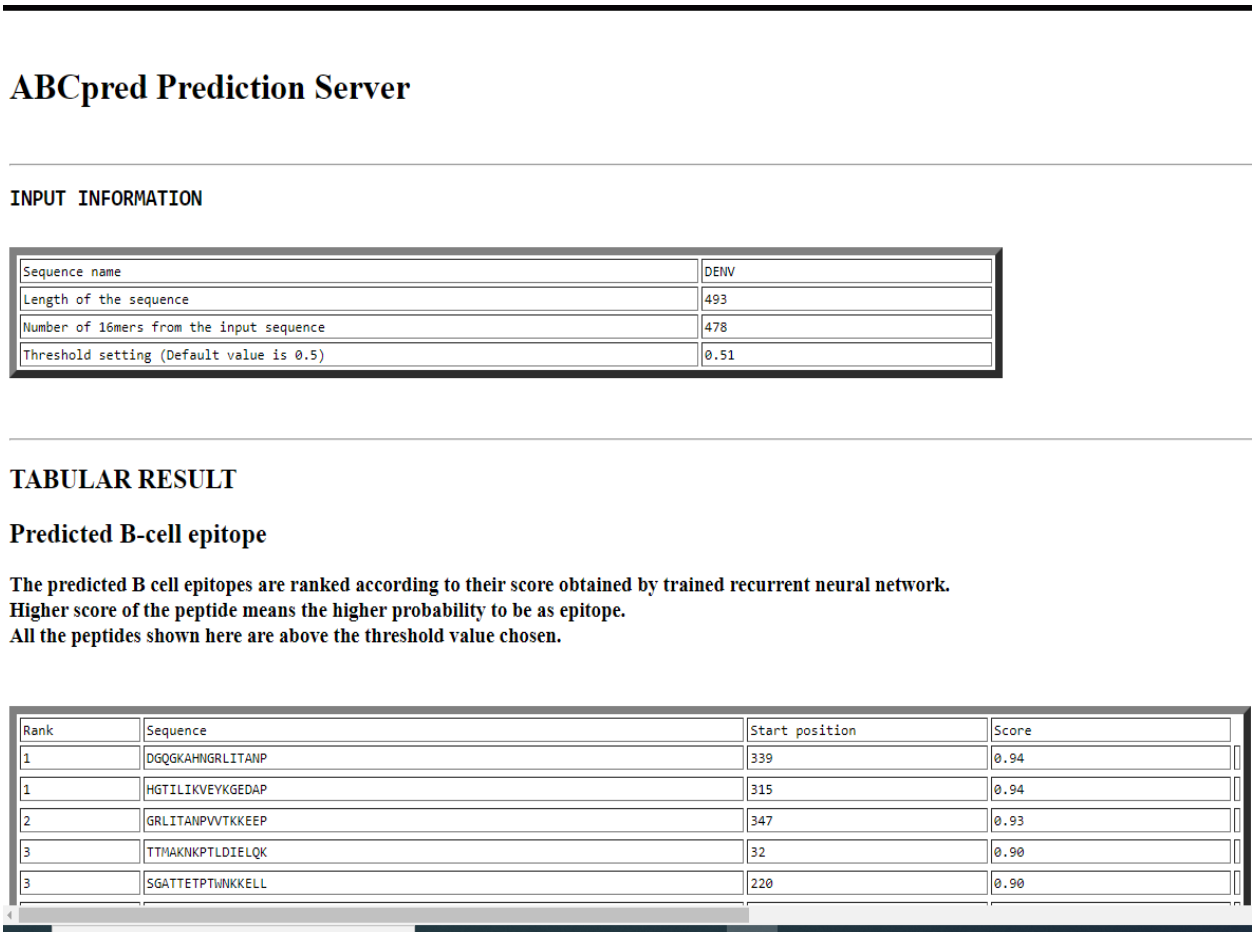


Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
a1_1	VVQYENLKY	-1.02	Non-Toxin	-0.16	-0.53	-0.33	0.00	1155.45
a1_2	VVQYENLKY	-1.02	Non-Toxin	-0.16	-0.53	-0.33	0.00	1155.45
a1_3	KLELKGMSY	-1.18	Non-Toxin	-0.18	-0.48	0.23	1.00	1068.43
a1_4	GTILIKVEY	-0.84	Non-Toxin	0.09	0.80	-0.40	0.00	1035.39
a2_3	ALFSGVSWI	-0.90	Non-Toxin	0.30	1.58	-1.21	0.00	979.27
a2_5	WIMKIGIGV	-0.91	Non-Toxin	0.29	1.56	-0.96	1.00	1016.46
a3_3	IVIGIITLY	-0.82	Non-Toxin	0.44	2.62	-1.47	0.00	1004.43
a3_4	KLELKGMSY	-1.18	Non-Toxin	-0.18	-0.48	0.23	1.00	1068.43
a3_5	VTFKNAHAK	-0.61	Non-Toxin	-0.20	-0.51	0.03	2.50	1015.30
a24_4	WMVHRQWFF	-0.67	Non-Toxin	-0.01	-0.14	-1.32	1.50	1336.71
a24_5	QYENLKYTV	-1.01	Non-Toxin	-0.24	-1.08	-0.21	0.00	1157.42
a26_1	IVIGIITLY	-0.82	Non-Toxin	0.44	2.62	-1.47	0.00	1004.43
a26_3	ELKGMSYAM	-1.11	Non-Toxin	-0.06	-0.06	-0.10	0.00	1029.36
a26_4	DTAWDFGSV	-0.75	Non-Toxin	0.01	-0.11	-0.22	-2.00	997.14
a26_5	ETQHGTILI	-1.35	Non-Toxin	0.01	0.09	-0.39	-0.50	1011.29
b7_1	TPTWNKKEL	-0.92	Non-Toxin	-0.33	-1.66	0.36	1.00	1116.41

Figure 16: Result of our multiepitope vaccine in ToxinPred.

ABCPred was utilized for the prediction of B cell epitopes. The Dataset of this server is obtained from BCIPEP. 700 unique experimentally proved continuous B cell epitopes are contained in the server which can be used to predict B cells of virus, bacteria, protozoa, and fungi. Artificial neural network (ANN), an information-processing paradigm following the densely interconnected, parallel structure of the mammalian brain process information is used by ABCPred in the epitope selection process. Our sample sequences were viruses and the predicted epitopes were further tested for its antigenicity, toxicity allergenicity, and immunogenicity.



ABCpred Prediction Server

INPUT INFORMATION

Sequence name	DENV
Length of the sequence	493
Number of 16mers from the input sequence	478
Threshold setting (Default value is 0.5)	0.51

TABULAR RESULT

Predicted B-cell epitope

The predicted B cell epitopes are ranked according to their score obtained by trained recurrent neural network.
Higher score of the peptide means the higher probability to be as epitope.
All the peptides shown here are above the threshold value chosen.

Rank	Sequence	Start position	Score
1	DGQGKAHNGRLITANP	339	0.94
1	HGTILIKVEYKGEDAP	315	0.94
2	GRLITANPVVTKKEEP	347	0.93
3	TTMAKNKPTLDIELQK	32	0.90
3	SGATTETPTWNKKELL	220	0.90

Figure 17: B cell epitopes predicted by ABCPred server.

In reverse vaccinology, the epitopes that are selected for the vaccine construct must be immunogenic. The peptide sequences should induce an immune response within the body. This immunogenicity test was carried on by the IEDB ANALYSIS RESOURCE which is an online server. IEDB is an immune epitope database site that includes MHC I and MHC II binding predictions, predict regions of proteins that are likely to be recognized as epitopes in the context of a B cell response. Both the results of CTL epitopes and HTL epitopes immunogenicity are saved in excel format. The peptide sequences of the higher percentile rank are chosen for further analysis. The peptide sequences of higher percentile rank have more probability of eliciting an immune response.

H1 | *fx* | percentile_rank

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	allele	seq_num	start	end	length	method	peptide	percentil	percentil	comblib_	comblib_	comblib_	comblib_	smm_ali	smm_ali
2	HLA-DRI	1	232	246	15	Consens	KELLVTF	0.14	0.14	-	-	-	-	LVTFKN/	23
3	HLA-DRI	1	231	245	15	Consens	KKELLV	0.14	0.14	-	-	-	-	LVTFKN/	23
4	HLA-DRI	1	230	244	15	Consens	NKKELL	0.18	0.18	-	-	-	-	LVTFKN/	24
5	HLA-DRI	1	229	243	15	Consens	WNKKEL	0.18	0.18	-	-	-	-	LLVTFKI	24
6	HLA-DRI	1	233	247	15	Consens	ELLVTFI	0.2	0.2	-	-	-	-	LVTFKN/	25
7	HLA-DRI	1	126	140	15	Consens	EGKVVC	0.36	0.36	-	-	-	-	VVQYEI	311
8	HLA-DRI	1	406	420	15	Consens	GARRM/	0.36	0.36	ARRMAI	0.01	0.01	0.01	GARRM/	58
9	HLA-DRI	1	407	421	15	Consens	ARRMAI	0.38	0.38	ARRMAI	0.01	0.01	0.01	ILGDTAV	60
10	HLA-DRI	1	127	141	15	Consens	GKVVQY	0.43	0.43	-	-	-	-	VVQYEI	304
11	HLA-DRI	1	125	139	15	Consens	IEGKVV(	0.52	0.52	-	-	-	-	VVQYEI	354
12	HLA-DRI	1	42	56	15	Consens	DIELQKT	0.58	0.58	TEATQL/	7979.18	64	64	LQKTEA	15
13	HLA-DRI	1	43	57	15	Consens	IELQKTE	0.58	0.58	TEATQL/	7979.18	64	64	LQKTEA	15
14	HLA-DRI	1	40	54	15	Consens	TLDIELC	0.64	0.64	LQKTEA	21222.54	71	71	LQKTEA	16
15	HLA-DRI	1	409	423	15	Consens	RMAILG	0.65	0.65	ILGDTAV	11.18	19	19	ILGDTAV	59
16	HLA-DRI	1	408	422	15	Consens	RRMAIL(	0.65	0.65	ILGDTAV	11.18	19	19	ILGDTAV	60

Table 3: Percentile rank of the peptide sequences after using IEDB server.

After selecting the T cell and B cell epitopes we added an adjuvant at the upstream of the epitope sequences. Vaccine adjuvants are molecules or compounds that have intrinsic immunomodulatory properties. When the adjuvant is administered in conjunction with an epitope-based vaccine it effectively potentiates the host antigen-specific immune responses. To achieve a higher, earlier,

and longer-lasting immune response adjuvants are used in vaccines. Typically, aluminum salts like aluminum hydroxide, aluminum phosphate are used in vaccines. Cytokine adjuvants are more recently used adjuvants in bacterial vaccines. In epitope-based vaccines, GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRK is used as adjuvants as it shows promising characteristics. It is allergenic and non-allergenic. The physicochemical properties are the molecular weight of 21806.21 Daltons and theoretical pH is 9.28. The instability index is 27.48 which indicates that the adjuvant is stable and the aliphatic index is 80.79. The estimated half-life of the adjuvant in E. coli is less than 10 hours (in vivo). This peptide adjuvant is also in research use in hepatitis C, hepatitis G, and HIV vaccines.

In the vaccine construct the adjuvant and B cell epitopes are added by EAAAK linkers. This linker peptide is mainly of alpha-helix structure which gives it a rigid property. EAAAK linkers can assist the two parts in a fusion protein to maintain the distance between each other and at the same time function independently. Another function of the EAAAK linker is to prevent interruption between two peptides during protein folding. GPGPG linkers were used in ligating B cell epitopes and helper T cell epitopes. We also use AAY linkers for the joining of CTL epitopes. The AAY linker assists in the expression of peptide sequences.

The physiochemical properties of a vaccine need to match and fulfill certain parameters. We have utilized the Expasy server and ProtParam tool for this purpose. Expasy stands for Export Protein Analysis system. The ProParam is a computational tool that can compute the various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user-entered protein sequence. It is operated by the SIB Swiss Institute of Bioinformatics. Created in 1993 Expasy is integrative that gives information exchange to over 160 Databases and software tools.

With the help of this server, several physicochemical parameters can be computed such as the molecular weight, theoretical pI, and amino acid composition. The atomic composition, extinction coefficient, estimated half-life, instability index can also be calculated by the server. The aliphatic

index and grand average of hydropathicity (GRAVY) are important features of an epitope-based vaccine which are also computed by the server. The raw form of a sequence or Swiss-Prot/TrEMBL accession number or ID is accepted in the submission box of the server.

The extension coefficient is an important feature of a peptide sequence that will be used as a vaccine. It is the measurement of light absorption of a substance or a chemical. The ratio of maximum to the minimum transmission of a beam of light that passes through a polarization optical train is referred to as the Extension coefficient. It is the calculation of how strongly a molecule can captivate light at a specific wavelength. The symbol of the molar extinction coefficient is (ε). For both of the calculations our vaccine peptide sequence showed, Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water. Ext. the coefficient is 74175. Here, Abs 0.1% (=1 g/l) 2.218, assuming all pairs of Cysteine residues form cystines. And

Ext. the coefficient is 73800 when Abs 0.1% (=1 g/l) 2.207, assuming all Cysteine residues are reduced.

When a pair of cysteine molecules are joined by a disulfide bond then the Cysteine amino acid is formed. Both calculations show the vaccine peptide sequence has optimal light absorbance capacity. Edelhoch method was used by the server for calculating the extinction coefficients which is based on the assumption that the ε of a denatured protein in 6 M guanidine-HCl can be calculated from the number of the tyrosine, tryptophan, and cysteine residues. Blocked amino acid analogs were in use as model substances to represent the situations in proteins. Tryptophan was replaced by N-acetyl-L-tryptophan amide and in the case of Tyrosine, glycyl-L-tyrosylglycine was used. The values were determined in pH 6.5, 6.0 M guanidium hydrochloride, and phosphate buffer of 0.02 M concentration. To the overall extinction coefficient calculation, Tryptophan contributes more compared to Tyrosine and Cysteine, even though Tryptophan extinction coefficient is not much sensitive to the environment. In the calculation of the server, it was assumed that the protein did not have other chromophores that absorb light at 280 nm which indicates that the concentration of conjugated proteins like the catalase, hemoglobin, or peroxidase that contain prosthetic groups, absorbing in the near UV and visible portions of the spectrum were not analyzed.

The mass of a given molecule is the molecular mass or molecular weight of the substance. The optimum molecular weight of a vaccine should be 30 kilo Daltons to 70 kilo Daltons. Our final vaccine had a molecular weight of 33 kilo Daltons. Therefore, the newly designed vaccine has a molecular weight of an ideal vaccine. The number of amino acids in our vaccine construct is 310. The total number of negatively charged residues Aspartate and Glutamic acid combined is 22 residues in our newly designed vaccine. The positively charged residues are of Arginine and Lysine. The total number of positively charged residues in our Dengue vaccine is 38.

The residues of the amino acid sequence protrude or hang over the surface. This is a characteristic of peptides or proteins. The extent to which the amino acid sequence protrudes from the surface of a protein can be calculated. The highly convex regions in proteins and the identification of protruding is crucial in the analysis of interfaces in protein-protein complexes. With the help of protrusion index (PI) quantification of overhanging peptides on protein, the surface is carried out in the Expasy server. The ideal theoretical PI is greater than 7. From the Expasy server, the theoretical PI is 9.49. The resultant outcome is greater than 7 which fulfills the criteria of an ideal vaccine. It also indicated that the newly designed vaccine is basic in nature.

When proteins are synthesized in cells they denature after carrying out their function. Half of the time required for a protein in a cell to disappear after its synthesis is termed as half-life of protein. It is a unique and important feature of proteins. The half-life of a protein was predicted with the help of ProtParam, Expasy server. The "N-end rule" is used by the server to perform this calculation. The N terminal residue is identified here which determines the half-life of the protein. The calculations of this server are not possible for N-terminally modified proteins. The identity of the N-terminal residue of a protein plays a significant role in calculating its stability in vivo. This rule was established from proven experiments. In those experiments, the metabolic fate of artificial beta-galactosidase proteins with different N-terminal amino acids engineered by site-directed mutagenesis was observed. The half-lives of the designed beta-gal proteins were more than 100 hours to less than 2 minutes, depending on the nature of the amino acid at the amino terminus, which was surprisingly different. The half-life also varied between the model of an experiment where yeast was in vivo, mammalian reticulocytes in vitro, and Escherichia coli in vivo. Moreover, it is observed in eukaryotes that the association of a destabilizing N-terminal residue of an internal

lysine, targets the protein to ubiquitin-mediated proteolytic degradation. Thus for humans, yeast, and E. coli, the prediction could be given by the server.

The vaccine sequence will be reverse translated and it will be cloned in E. coli. So this parameter holds huge importance. The standard half-life for E. coli in vivo should be more than 10 hours. The newly designed Dengue vaccine has an estimated half-life of more than 10 hours in E. coli in vivo which fulfills the criteria. The N-terminal of the sequence was in Glycine and the estimated half-life in mammalian reticulocytes in vitro is more than 30 hours and in yeast, in vivo, it is more than 20 hours.

The stability of a protein is of huge significance in vaccine designing. It can be done with the help of the instability index which gives an approximate value of the stability of a protein in a test tube. The instability index (II) was calculated by the statistical analysis of 12 unstable and 32 stable proteins. From the analysis, it is found that the expression of certain dipeptides is significantly different in the unstable proteins compared with those in the stable ones. A weight value of instability to each of the 400 different dipeptides (DIWV) is assigned by the authors of this method. It is possible to compute an instability index (II) using these weight values.

A peptide sequence whose instability index is less than 40 is considered stable whereas a value above 40 predicts that the protein may be unstable. The instability index (II) is computed to be 29.60 which classifies the protein as stable. Therefore, the protein sequence of the epitope-based Dengue vaccine that we designed is well constructed.

The amino acids Alanine, Valine, Isoleucine, and Leucine have aliphatic side chains. These side chains occupy a certain volume within the molecule. In a protein, the relative volume occupied by the aliphatic side chains is termed the aliphatic index of that protein. For the increase of thermostability of globular proteins, the aliphatic index is considered a positive factor.

The value generated after the calculation from ProtParam, Expasy server for our vaccine is 80.06. For a protein sequence to be thermostatically stable its Aliphatic index must be above 70. Therefore, the epitope-based Dengue vaccine is thermostatically well stable.

There are hydrophobic and hydrophilic side chains in amino acids. The number that represents the hydrophilic and hydrophobic properties of an amino acid is termed as hydropathy index. The

hydrophobic amino acids have higher values. Isoleucine (4.5) and valine (4.2) are the most hydrophobic amino acids. The hydropathy values are used in the calculation of the Grand Average of Hydropathy. The sum of hydropathy values of all the amino acids divided by the number of residues present in the sequence can determine the GRAVY value. For the newly constructed vaccine against Dengue virus, the Grand Average of Hydropathy value is – 0.105.

The chemical composition of the designed Dengue virus contains Carbon, Hydrogen, Nitrogen, Oxygen, and Sulphur. And the chemical formula is $C_{1525}H_{2367}N_{403}O_{415}S_{14}$. The total number of atoms in the vaccine is 4724 atoms.

To investigate the protein structure of our vaccine we used the PSIPRED, PSI-blast-based secondary structure PREDiction. From the primary protein sequence, it can determine beta-sheets, alpha helices, and coils with the help of artificial neural network machine learning. Web server, as well as software, is available for PSIPRED. Two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST are incorporated in the system which makes its prediction very accurate. It was developed by the bioinformatics group, University College London. Generating a sequence profile then predicting the initial secondary structure, and filtering the predicted structure are the three stages of the prediction algorithm. The neural network is fed with a window of 15 acids for each amino acid in the given sequence. Helix, Strand, and Coil are the forms of protein that are shown in the result sheet. In our vaccine, the percentage of α helices is 34%. 17.0% β sheets are contained in our vaccine which has been calculated by PSIPRED. Other components like Random Coil and β-turn are 49.0% in the vaccine.

Chapter 05

Conclusion

The Dengue virus has caused unprecedented damage in recent years in Bangladesh along with many other countries in Asia and South America. The newly designed vaccine has a high prospect in protecting the lives of human beings in Bangladesh and abroad. The vaccine has been targeted against the Envelope protein of the Dengue virus which is highly conserved and immunogenic. Due to the ability to perform a crucial task during an immune response, B cell, cytotoxic T cell, and T helper cell were employed in the study to achieve humoral as well as cell-mediated immunity. Considering many parameters and factors like antigenic, immunogenic, non-allergenic, non-toxic, and number of MHC-I & II binding properties we chose 10 CTL epitopes, 3 HTL epitopes, and 4 B cell epitopes in the vaccine construct. Even though we have accomplished many steps in terms of designing a vaccine, we still need to work on many other validation procedures in the future. Also known as CD283, Toll-like receptor 3 (TLR3) is an important protein that plays a significant role in pattern recognition and stimulation of the innate immune system. Molecular docking could be done with the help of the ClusPro 2.0 server to assess the binding affinity of the Toll-like receptor 3 and the vaccine. For the visualization of the interactions, we can use Pymol which is a molecular visualization software. This will enable us to understand in depth the vaccine we have constructed. The compactness and the stability of the TLR-vaccine complex will be analyzed in a 50 ns molecular dynamics simulation. In addition, the average RMSF value and the average radius of gyration should also be calculated for validation of the vaccine. For future analysis, we also need to run molecular simulations which can be achieved by GROMACS. By this simulation, we should also calculate the hydrogen bond and other interactions. Furthermore, iMODS analysis will have to be performed which will facilitate the exploration modes and generate feasible transition pathways between two homologous structures which will signify the stability of the vaccine-TLR3 complex interface. In Silico cloning is also required for validation of the vaccine. For this purpose, the JCat server can be utilized which will allow us to observe the GC content of the vaccine. The vaccine needs to be tested in various models to check the immune response. Lab rats, rabbits can be used for the initial animal model test. By performing various

immunoassays from the blood sample of the animal models we will be able to determine the efficiency of the vaccine. Next primates can be injected with the vaccine before the human trial. Three clinical phases will have to be passed by the vaccine. If the vaccine passes all these steps, then it can be commercially produced and it will be ready for vaccination. Therefore, this Dengue vaccine will be able to give protection against Dengue infection and thus save human lives and lessen sufferings due to Dengue Virus.

References

1. Pulendran, B., S. Arunachalam, P. & O'Hagan, D.T. Emerging concepts in the science of vaccine adjuvants. Nat Rev Drug Discov 20, 454–475 (2021). https://doi.org/10.1038/s41573-021-00163-y.

2. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. BMC Bioinformatics. NetCTL-1.2:Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. Oct 31;8:424. 2007.

3. Sudipto Saha, G. P. S. Raghava, AlgPred: prediction of allergenic proteins and mapping of IgE epitopes, Nucleic Acids Research, Volume 34, Issue suppl_2, 1 July 2006, Pages W202–W209, https://doi.org/10.1093/nar/gkl343.

4. Victor Hugo Urrutia-Baca, Ricardo Gomez-flores, Myriam Angélica De La Garza-Ramos, Patricia Tamez-guerra, Daniela Guadalupe Lucio-sauceda, and María Cristina Rodríguez-padilla Immunoinformatics Approach to Design a Novel Epitope-Based Oral Vaccine Against Helicobacter pylori Journal of Computational BiologyVol. 26, No. 10.

5. Pulendran, B., S. Arunachalam, P. & O'Hagan, D.T. Emerging concepts in the science of vaccine adjuvants. Nat Rev Drug Discov 20, 454–475 (2021). https://doi.org/10.1038/s41573-021-00163-y.

6. Maryam Enayatkhani, Mehdi Hasaniazad, Sobhan Faezi, Hamed Gouklani, Parivash Davoodian, Nahid Ahmadi, Mohammad Ali Einakian, Afsaneh Karmostaji & Khadijeh Ahmadi (2021) Reverse vaccinology approach to design a novel multi-epitope vaccine candidate .

7. Saha, S and Raghava G.P.S. (2006) Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. Proteins,**65(1)**,40-48 PMID: 16894596.

12. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268. PMID: 30976793; PMCID: PMC6602479.

13. Thawornchai Limjindaporn, Wiyada Wongwiwat, Sansanee Noisakran, Chatchawan Srisawat, Janjuree Netsawang, Chunya Puttikhunt, Watchara Kasinrerk, Panisadee Avirutnan, Somchai Thiemmeca, Rungtawan Sriburi, Nopporn Sittisombut, Prida Malasit, Pa-thai Yenchitsomanus, Interaction of dengue virus envelope protein with endoplasmic reticulum-resident chaperones facilitates dengue virus production,Biochemical and Biophysical Research Communications, Volume 379, Issue 2, 2009, Pages 196-200,

14. B. Sun and Y. Zhang, "Overview of orchestration of CD4+ T cell subsets in immune responses," Advances in Experimental Medicine and Biology, vol. 841, pp. 1–13, 2014.View at: Publisher Site | Google Scholar.

15. L. Malherbe, "T-cell epitope mapping," Annals of Allergy, Asthma & Immunology, vol. 103, no. 1, pp. 76–79, 2009.View at: Publisher Site | Google Scholar.

16. R. K. Ahmed and M. J. Maeurer, "T-cell epitope mapping," Methods in Molecular Biology, vol. 524, pp. 427–438, 2009.View at: Publisher Site | Google Scholar

17. Anderson, J.R. and R. Rico-Hesse, Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. American Journal of Tropical Medicine and Hygiene, 2006. 75(5): p. 886-892.

18. Ye, Y.X.H., et al., Wolbachia Reduces the Transmission Potential of Dengue-Infected Aedes aegypti. PLOS Neglected Tropical Diseases, 2015. 9(6): p. e0003894.

19. Duong, V., et al., Asymptomatic humans transmit dengue virus to mosquitoes. Proceedings of the National Academy of Sciences of the USA, 2015. 112(47): p. 14688–14693.

20. Nguyen, N.M., et al., Host and viral features of human dengue cases shape the population of infected and infectious Aedes aegypti mosquitoes. Proceedings of the National Academy of Sciences of the United States of America, 2013. 110(22): p. 9072-9077.

21. Sinhabahu, V.P., R. Sathananthan, and G.N. Malavige, Perinatal transmission of dengue: a case report. BMC Research Notes, 2014. 7(795).

22. Basurko, C., et al., Maternal and foetal consequences of dengue fever during pregnancy. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2009. 147(1): p. 29-32.

23. Pouliot, S.H., et al., Maternal Dengue and Pregnancy Outcomes A Systematic Review. Obstetrical & Gynecological Survey, 2010. 65(2): p. 107-118.

24. Trpis, M., et al., DIEL PERIODICITY IN LANDING OF AEDES-AEGYPTI ON MAN. Bulletin of the World Health Organization, 1973. 48(5): p. 623-629.

25. C. P. Toseland, D. J. Clayton, H. McSparron et al., "AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data," Immunome Research, vol. 1, no. 1, p. 4, 2005.View at: Publisher Site | Google Scholar.

26. J. D'Amaro, J. G. A. Houbiers, J. W. Drijfhout et al., "A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs," Human Immunology, vol. 43, no. 1, pp. 13–18, 1995.View at: Publisher Site | Google Scholar.

27. B. Peters, W. Tong, J. Sidney, A. Sette, and Z. Weng, "Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules," Bioinformatics, vol. 19, no. 14, pp. 1765–1772, 2003.View at: Publisher Site | Google Scholar.

28. V. Brusic, G. Rudy, G. Honeyman, J. Hammer, and L. Harrison, "Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network," Bioinformatics, vol. 14, no. 2, pp. 121–130, 1998.View at: Publisher Site | Google Scholar.

29. P. Donnes and O. Kohlbacher, “Integrated modeling of the major events in the MHC class I antigen processing pathway,” Protein Science, vol. 14, no. 8, pp. 2132–2140, 2005.View at: Publisher Site | Google Scholar.

30. Allen TM, O’Connor DH, Jing P, Dzuris JL, Mothé BR, Vogel TU, Dunphy E, Liebl ME, Emerson C, Wilson N, et al. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. Nature. 2000;407:386–390.

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

32. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6, 1948-1954. [PubMed: 6049437]

33. Yang, Y., Sun, W., Guo, J., Zhao, G., Sun, S., Yu, H., Guo, Y., Li, J., Jin, X., Du, L., Jiang, S., Kou, Z., & Zhou, Y. (2015). In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. Human vaccines & immunotherapeutics, 11(3), 795–805. https://doi.org/10.1080/21645515.2015.1012017.

34. Kumar, J., Qureshi, R., Sagurthi, S. R., & Qureshi, I. A. (2020). Designing of Nucleocapsid Protein Based Novel Multi-epitope Vaccine Against SARS-COV-2 Using Immunoinformatics Approach. International journal of peptide research and therapeutics, 1–16. Advance online publication. https://doi.org/10.1007/s10989-020-10140-5.