AN IN SILICO APPROACH TO MULTI-EPITOPE SUBUNIT VACCINE DEVELOPMENT AGAINST SARS COV-2 USING MEMBRANE GLYCOPROTEIN

By

Rian Rafsan

16336017

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

Fall 2020

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:

_____

Rian Rafsan
16336017

Approval

The thesis titled 'An *in silico* approach to multi-epitope subunit vaccine development Against SARS CoV-2 using membrane Glycoprotein' submitted by Rian Rafsan(16336017) of Summer 2016 has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Bachelor in Biotechnology on the 8th of November 2020.

Examining Committee:

Supervisor:
(Member)

_____

Romana Siddique
Senior Lecturer,Department of Mathematics and Natural Sciences
BRAC University

Program Coordinator:
(Member)

_____

Dr.Iftekhar Bin Nasser
Assistant Professor, Department of Mathematics and Natural Sciences
BRAC University

Departmental Head:
(Chair)

_____

Prof. A. F. M. Yusuf Haider
Chairperson, Mathematics and Natural Sciences BRAC University

Ethics Statement

The thesis was conducted maintaining ethical standards at all regard whatsoever.No animals or human beings were hurt or used in any way and no breach of biosafety & ecological protocol was done by any means.

Abstract

SARS Cov -2, short for *Severe Acute Respiratory Syndrome Coronavirus-2* is a + sense zoonotic retrovirus of the *Coronaviridae* family and is genetically a mutant version of the SARS Cov and is the causal organism of the Covid-19 disease. First reported on the 31st of December, 2019 in the Wuhan, Hubei province of China the disease was soon declared a pandemic by the WHO on the 30th of January 2020. The virus itself is claimed to have originated from snakes, bats and pangolins as potential carriers in the 'wet markets' of Wuhan where live wild animals are sold for consumption under poor hygienic conditions. As such, the environment is ideal for viral growth and transmission particularly via faeces, blood, semen and other fluids of biological origin from the carriers.Patients of this disease are majorly symptomized by mild to moderate fever, runny nose, dried cough and tiredness along with respiratory arrest in extremely severe cases, particularly elderlies.

Although medicinal drugs such as hydroxy-Chloroquine and Remdesivir are claimed to be potential therapeutic candidates against this virus , no credible vaccine has yet been successfully developed that can efficiently combat this disease. Here, we tried to develop a multi-epitope sub-unit vaccine against SARS Cov-2 using the membrane glycoprotein (M) as the primary antigen in our vaccine construct. Our choice of antigen was carried out based on two basic features ;(i)Antigenicity of the protein.(ii)Functionality of the protein.To measure the antigenicity of our proteins we used the Vaxijen 2.0 server.At a minimal threshold set to 0.5 , antigenicity of the potential protein candidates preliminarily selected were measured . A total of 12 proteins' primary structures were primarily retrieved in FASTA format from NCBI website . Among the candidates were 10 non-structural proteins (nsp2-11) and 2 structural proteins - namely the Spike glycoprotein (S) & membrane glycoprotein (M) . The membrane glycoprotein (M) was selected for further analysis.CTL epitopes were then identified using NetCTL 1.2 server.CTL epitopes were analysed for MHC I allele specificity and HTL epitopes corresponding to MHC II alleles respectively using the respective IEDB servers . Unique epitopes were selected and epitopes that were mutually homologous or partially homologous were excluded.Linear B cell epitopes were identified specific to our primary antigen using Bepipred Linear Epitope Prediction 2.0 algorithm at a minimal threshold of 0.5.All the selected epitopes were then added

to the primary antigen using proper linkers to enhance stability of our vaccine . Biochemical analysis was done using the PROTPARAM tool. Next we built a PDB structure of our final vaccine using phyre2 server via homology modelling that was docked with the TLR8 receptor of toll-like receptor family.We further validated the credibility of our vaccine's efficacy by carrying out an *in silico* simulation analysis using C-immsim server . We believe that the vaccine we developed will be a potential candidate to effectively combat SARS Cov-2 . We also believe that further *in vitro* analysis is required to get a clear understanding of this vaccine and its further development.

Acknowledgments

I would like to take this opportunity to first and foremost express my utmost gratitude towards late Professor Ziauddin Ahmed , former Professor and Chairperson , department of Mathematics and Natural Sciences BRAC University for the constant care , support and inspiration he provided me ; particularly towards learning new things . I would also like to thank Professor A B M Yusuf Haider,Chairperson to Department of Mathematics and Natural Sciences and Professor Mahboob Mojumder,Dean to the School of Sciences, BRAC University for their constant support and encouragement towards me. I am sincerely grateful to S.M. Rakib Uz Zaman , lecturer to department of Mathematics and Natural Sciences and Professor Aparna Islam for their cordial support and undaunting encouragement. I would also like to thank Dr. Jamal Mostofa , Assistant Professor to Faculty Of Forensic Medicine , University of Kagawa , Japan for his constructive suggestions in regards to research development.I would like to thank Royvi Rafayet my brother, student of Nuclear Science and Engineering at **Military Institute Of Science & Technology,Dhaka** for his cordial efforts in regards of logistics support.I also thank Arafat Khan Antu for his constant support & inspiration to be laborious towards work.

I would like to specially thank my supervisor Ms. Romana Siddique , Senior Lecturer , Department of Mathematics & Natural Sciences , for her cordial support and laborious efforts towards keeping me constantly motivated and challenging me to reach new heights . I greatly appreciate the patience and courage she has shown , especially towards me throughout my academic life here at BRAC University which is nonetheless worthy of admiration. Lastly , I express my gratitude towards all the faculty members and staff associated with BRAC University who helped make this journey towards the endeavour of knowledge an amazing experience to remember.

Rian Rafsan
Biotechnology Program
Department of Mathematics & Natural Science
BRAC University

Table Of Contents.

List Of Tables.

List Of Figures.

List of Acronyms

SARS CoV-2	Severe Acute Respiratory Syndrome Corona Virus -2
nsp	Non Structural Protein
NK	Natural Killer
MHC	Major Histocompatibility Complex.
CTL	Cytotoxic T cell Lymphocyte.
Covid-19	Corona Viral Disease 2019
DMV	Di-membrane Vesicle
DNA	Deoxy-ribonucleic acid.
RNA	Ribo-nucleic acid.
IFN	Interferon
S protein	Spike protein
M protein	Membrane glycoprotein
N protein	Nucleocapsid protein
ACE-2	Angiotensin Converting Enzyme-2
E protein	Envelope Protein

Chapter 1

Introduction & Literature Review

Before we dive deep into the process of vaccine development we must first understand what is a vaccine? The word 'Vaccine' has originated from the Latin words *Variolae vaccinae* first demonstrated by Edward Jener in 1798 as an attempt to eradicate smallpox in humans . Today , vaccines refer to a specific preparation of biological or semi-biological origin that upon administration to the host boosts up the immune system . This immunity that is innate by the nature of mechanism followed,can either be :

(i)Prophylactic .
(ii)Therapeutic .

Prophylactic vaccines prevent the occurrence of a vaccine , whereas therapeutic vaccines tend to approach the disease from a curative perspective ; such vaccines attempt to eradicate the disease that has already attacked upon its host by training the host's immune cells to effectively fight against the disease . One basic fact to remember is that vaccines are developed to fight against diseases of infectious nature from a pathogenic origin . The pathogen itself can be viral , bacterial , algal or of any other origin . Vaccines are developed targeting these pathogens (whole cell vaccine) or their specific morphological organelles. Vaccine development is itself a time-consuming process of multifarious variety . Hence they are classified into various groups from various perspectives .

Vaccines , from the viewpoint of construction, can be classified into 5 basic categories based on the process employed behind their construction as follows :

(i)Live attenuated vaccine.
(ii)Inactivated vaccine.

(iii)Subunit vaccine.

(iv)Toxoid vaccine.

(v)Conjugate polysaccharide vaccine.

(vi)Nucleic acid vaccine.

1.1: Whole pathogen vaccines:

The most traditional of the lot , these vaccines consist of the entire pathogen with their antigenicity reduced. Vaccines of such origin are capable of eliciting strong immune responses against the pathogen by the host. Although most vaccines in clinical use fell in this category , the scope of using this method in vaccine construction is rather narrow. Partly the reason being the whole cell pathogen if introduced in the host body can often result in a very strong immune response within the host system that can actually kill the host or harm the system severely .

1.2: Inactivated Vaccines:

The discovery of inactivated or killed pathogens to induce immunity within host system in the 19th century led to the development of inactivated vaccines through killing of the causal pathogen via heat or chemical treatment.An example of an inactivated vaccine is the vaccine Havrix against hepatitis A developed by NIAID with aid from the US and was licensed in the USA in 1995.

1.3: Live attenuated Vaccines:

Rapid advancements in the field of tissue culture in the 1950s lead to the development of live attenuated vaccines.Such vaccines contain a weakened version of the pathogen with reduced pathogenicity.Examples of such vaccines include the MMR vaccine effective against measles , mumps and rubella and the *Varicella zoster* vaccine against chickenpox and herpes . Capable of inducing strong longer immunity towards specific pathogen upon administration at specific dosage , a major limitation of this method lies in the fact that although it exhibits relative ease of

usage against pathogens of viral origin,antibacterial and antiparasitic vaccines are relatively difficult to be developed owing to their more complex morphological features.

1.4: Chimeric Vaccines:

Vaccines that are classified to be of chimeric origin are composite by nature.They structurally contain an antigenic protein from one organism and a trafficking domain that will transport that antigenic protein to an endosomal compartment.Often the two may be of different origins both chemically and biologically.For example , a vaccine developed by the NIAID consists of dengue virus surface surface protein with zika virus backbone.

1.5: Subunit Vaccines:

Instead of usage of whole cell pathogens,subunit vaccines are built on the backbone of specific antigens that stimulate strong immune response in the host system.To make the immune response stronger , proper adjuvants and linker molecules are also added to enhance the immune response further and make the vaccine more stable.Subunit vaccines tend to pose lower side-effects than the aforementioned vaccine types and are easier to develop making it an ideal method to develop vaccines at a lower risk level.To better understand the development of a subunit vaccine , we can look into the case of pertussis or whooping cough vaccine.Pertussis vaccines were first in the mid 1940s as a combination of diphtheria,tetanus & pertussis (DTP) vaccine.Although functionally effective,these **whole cell** vaccines often caused fever and swelling at injection sites.This resulted in people avoiding the vaccine causing an adverse rise in new infection rates.A new vaccine was thus developed using acellular *B pertussis* organisms. Meaning , they didn't contain whole cells , but rather purified cellular components . This resulted in reduced cases of adverse reactions and fatalities .

1.6: Toxoid Vaccines :

Toxoid vaccines are vaccines composed of inactivated toxin from the causal organism in minor concentration induced to the host that boosts up the host immunity . These toxic 'antigens' elicit

an immune response and consequently provide immunity to the host system . An aspect to be careful of during development of such vaccines is the median lethal dose or LD_{50} value or the lethal dose 50 of the toxin being used. Quantitatively, this property expressed in the unit of $mgkg^{-1}$ of body mass signifies the amount of the toxin needed to kill 50% of a test population . Toxins with very low median lethal doses are difficult to be developed into toxoid vaccines , for example *Clostridium botulinum* with a LD_{50} value of 1.3-2.1 $ngKg^{-1}$.An example of toxoid vaccine is the tetanus vaccine against tetanus caused by *Clostridium tetani*.

1.7: Nucleic Acid Vaccines:

A relatively new approach to vaccine development is the production of nucleic acid vaccines. Genetic matter is introduced within the host system that encodes the antigenic protein against which the antigenicity is sought . Potential advantages of such vaccines include a broad spectrum of pathogens to whom this method is applicable , provision of stronger immune response and long term immune protection to the host as well as excellent stability and ease of scaling up of vaccine productivity at industrial level makes this method ideal for future vaccine development .

Nucleic acid vaccines are mutually variant among themselves ; because various types of nucleic acids can be used to develop the vaccine construct. **DNA plasmid vaccines** for example , are developed by identifying the gene of interest (GOI) of the potent antigenic protein inserted into the plasmid of a bacterial host . Competent bacterial cells that will contain the GOI will produce the antigenic protein that's the primary backbone to the construct. Further purification, quality control and trials will then lead to the development of the final end product i.e. the commercial vaccine. **mRNA vaccines** are also a new generation of vaccines that are developed based on the mRNA template of the antigenic POI of the causal pathogen. Often potentially harmless carriers of viral or bacterial origin of minimal antigenicity are used to carry nucleic acid vaccines within the host system . Example of nucleic acid vaccine is the HPV vaccine while potential candidates are being tested for vaccine development against EBOV and Zika viruses under NIAID supervision .

Chapter 2

Understanding the Genomic & Pathogenic Origins of SARS CoV-2.

2.1: Morphology and Pathogenesis of the SARS CoV-2 :

Superkingdom: Viridae

Phylum: Riboviria

Order: Nidovirales

Sub-order: Cornidovirineae

Family: Coronaviridae

Sub-family: Orthocoronavirinae

Genus: Betacoronavirus

Sub-genus: Sarbecovirus

Species: Severe Acute Respiratory Syndrome CoronaVirus 2

To get a better understanding of SARS CoV-2 we started by studying the virus' taxonomic classification as done by the International Commission On Taxonomy of viruses (ICT) as stated above . As priorly mentioned , SARS CoV-2 is a + sense retrovirus of the *Coronaviridae* family.Based on the rooted or unrooted nature of phylogenetic tree and partial sequencing of their RNA dependent RNA polymerases , members under the sub-order of *Coronavirinae* are further classified into 4 genera ; alpha (α) coronavirus , beta (□) coronavirus , gamma (ɣ) coronavirus and delta (*δ*) coronavirus . (Woo, Lau, Huang & Yuen, 2009). Members of the α & □ coronavirus genera infect only warm blooded animals as hosts to be specific , whereas members of ɣ & *δ* coronavirus genera mostly infect birds . Although some of their members are also seen to possess animal hosts . A phylogenetic tree among different members of the *Coronaviridae* family was studied (Cui, Li & Shi, 2019) as shown in figure 1 followed by a generic classification of different CoVs was also tabulated in table 1.

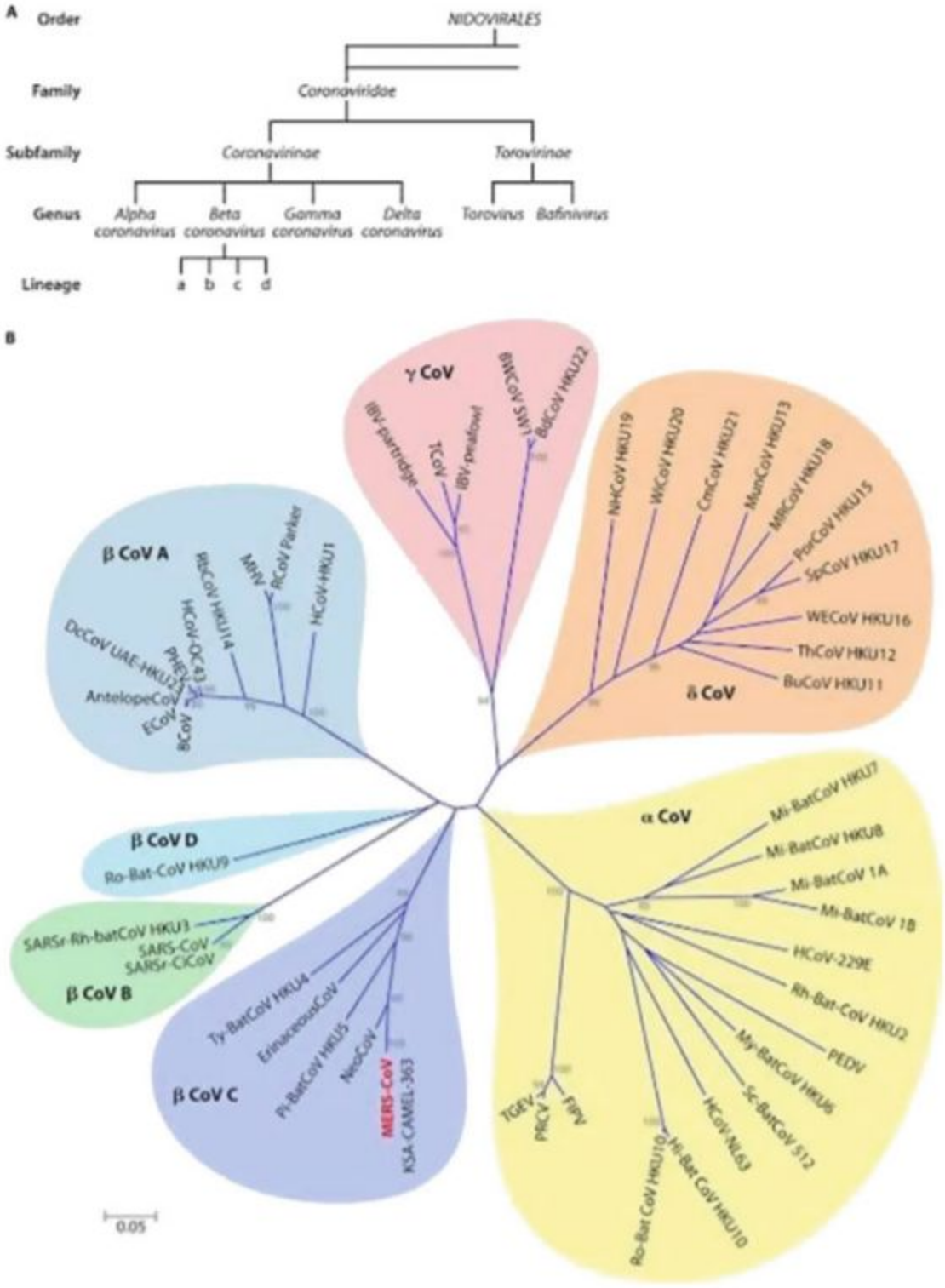


Figure 1: Phylogenetic tree constructed among different members of the Coronaviridae family.(Cui, Li & Shi, 2019)

Genus	Species
Alpha (α) coronavirus .	Human coronavirus 229E , Human coronavirus NL63 , Miniopterus Bat coronavirus 1,Miniopterus Bat coronavirus HKU8 , porcine epidemic diarrhoea virus , Rhinolophus bat coronavirus HKU2 , Scotophilus bat coronavirus 512
(□) coronavirus	Beta (□) coronavirus 1, Human coronavirus HKU1 , murine coronavirus, Pipistrellus bat coronavirus HKU5, Rousettus bat coronavirus HKU9, SARS CoV, SARS CoV-2, MERS CoV, Human coronavirus OC43, hedgehog coronavirus(EriCoV)
Gamma (ɣ) coronavirus	Infectious bronchitis virus (IBV) , beluga whale coronavirus SW1 ,
Delta (*δ*) coronavirus	Bulbul coronavirus HKU11 , porcine coronavirus HKU15

Table 1: Genera wise classification of different coronaviruses.

2.2: Genomic and morphological organization to SARS CoV-2:

SARS Cov-2 has a + sense , linear ssRNA genome of ≈ 30 kbp in size comprising 6-11 ORFs responsible for encoding 9680 polyproteins . (Guo et al , 2020). The first ORF (1a) constitutes around 67% of the viral genome encoding 16 nsps . The rest of the ORFs encode accessory and structural proteins . The genome lacks the hemagglutinin esterase gene but contains two flanking UTRs at 5'-end of 265th and 3'-end of 358th nucleotide . nsp-s of SARS CoV-2 include 2 cysteine

proteases -one that is a pepaine homologue (nsp3) while the other is the core protease and a chymotrypsin homologue consisting of 3 carbon atoms (nsp5) .

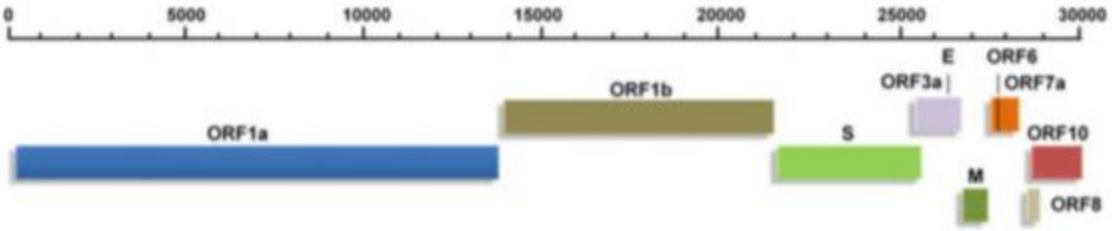


Figure 2: Genomic structure of SARS CoV-2 ssRNA with coding regions of functional polyproteins.

Table 2: Functions of nonstructural proteins encoded by SARS CoV-2 genome.

Name of non-structural protein.	**Function.**
nsp1	IFN signal inhibition , mRNA degradation , translatory inhibition and arrest of the cellular cycle .
nsp2	Function unknown ; associated with RTCs.
nsp3	Pepain homologue; processing of polyproteins at ORFs , ADP -ribose phosphorylation , binding to RNA, host IFN antagonisation and disruption of host innate immune-response.
nsp4	Presumably Double Membrane Vesicle (DMV) formation.
nsp5	Main proteinase M , inhibitor of IFN signalling.
nsp6	Presumably DMV formation.

Name of non-structural protein.	**Function.**
nsp7	ssRNA binding and cofactor for nsp8 & nsp12.
nsp8	Cofactor for nsp7,nsp12 & primase.
nsp9	ssRNA binding;associated with RNA transcription complexes' dimerization.
nsp10	Dodecameric zinc finger;associated with RTCs;nsp16 methyltransferase stimulator & scaffold for nsp14 & nsp16.
nsp11	Unknown.
nsp12	Primer dependent RNA dependent RNA polymerase.
nsp13	Helicase;RNA 5'-triphosphatase.
nsp14	3'→5' exonuclease;guanine-N7-methyl transferase(RNA cap formation.)
nsp15	Endoribonuclease;dsRNA sensor evasion.
nsp16	Ribose-2'-O-methyl transferase (Formation of the RNA cap.

Table 2(Contd.): Functions of nonstructural proteins encoded by SARS CoV-2 genome.

The viral genomic sequence is responsible for encoding the entire poly proteomic structure of the virus . While the non-structural proteins are responsible for the smooth coordination of the entire

system , the physical framework behind a virus' functionality is made out of the structural proteins building its morphological structure.

4 major structural proteins are found to compose the composite proteomic structure ; they are the spike protein(S) , membrane glycoprotein(M), nucleocapsid protein(N) and the envelope protein(E) respectively .

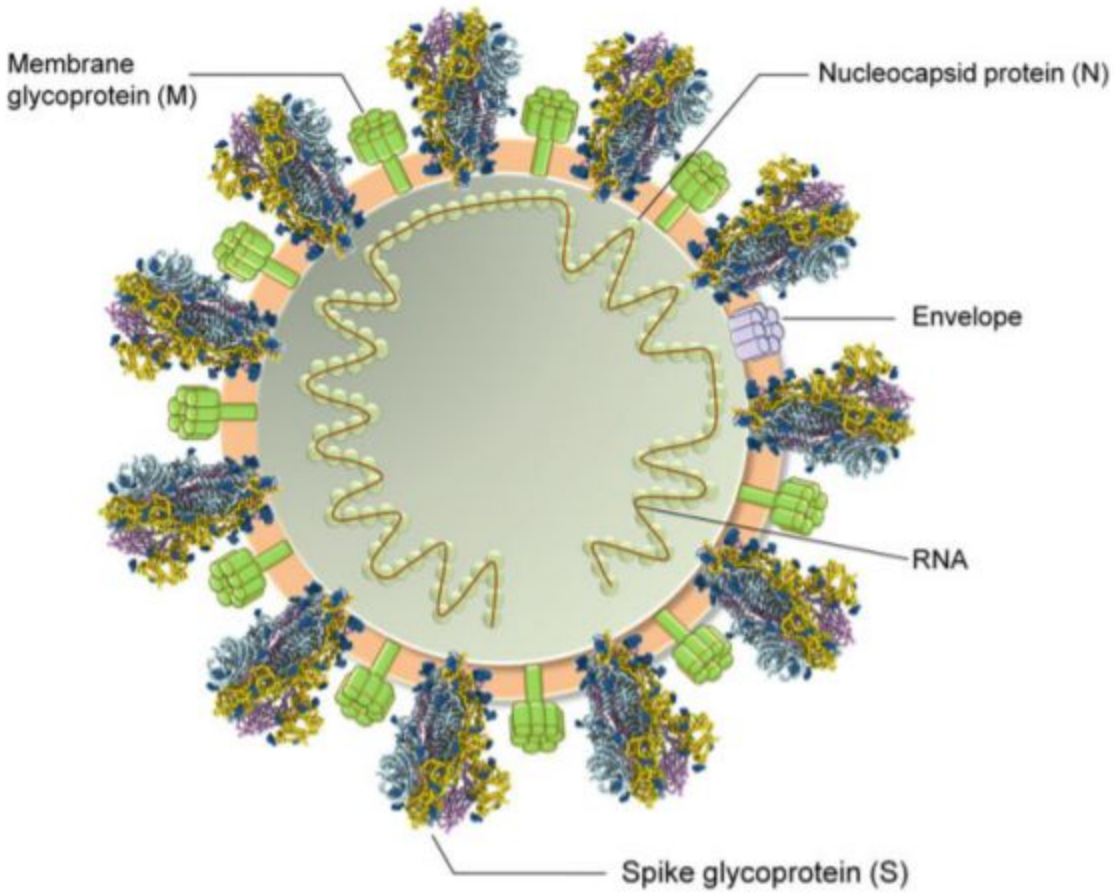


Figure 3: Labelled diagram of SARS Cov-2 showing position of structural proteins in the viral proteome .

The S, M & E proteins are all embedded within the lipid bilayer encapsulating the N protein that contains within itself the viral RNA . Among the structural proteins, S binds to the Angiotensin

Converting Enzyme -2 (ACE-2) receptor and thus causes viral infection. The S protein can also be proteolytically cleaved at S1(685 aa) and S2 (588 aa) subunits respectively and is a potential drug target (Zhang, Penninger, Li, Zhong & Slutsky, 2020).While E & M are functionally responsible for morphogenesis , assembly & budding , S is a fusion protein.

2.3: Entry, replication & Pathogenesis of SARS Cov-2 within the host system:

Entry of Cov into the target host cell depends on the binding affinity during the interaction between receptor and corresponding ligand. Homologous to its predecessor SARS CoV, SARS CoV-2 uses ACE-2 receptor and TMPRSS2 serine proteases for the purpose of internalization and priming of S protein respectively (Hoffmann et al., 2020). Studies have further revealed that S protein of SARS Cov-2 has 10-20 times more affinity towards ACE-2 receptors than SARS Cov (Wrapp et al. 2020). Binding between S and the ACE-2 receptor forms a protein-ligand complex . This complex formation results in conformational change of the S protein leading to the fusion between M and E protein , followed by an entry into the host cell via endosomal pathway (Coutard et al. 2020; Matsuyama and Taguchi 2009). Viral RNA is released into host cell cytoplasm. Translation of the RNA generates replicase polyproteins pp1a and pp1b respectively that are further cleaved under the influence of viral proteases. Replication of SARS Cov-2 involves ribosomal frameshifting .Viral assembly occurs via interaction between RNA , proteins in the endoplasmic reticulum (ER) and Golgi complex formation.Once formed, the complexes are packed into vesicles and released via exo-cytosis.

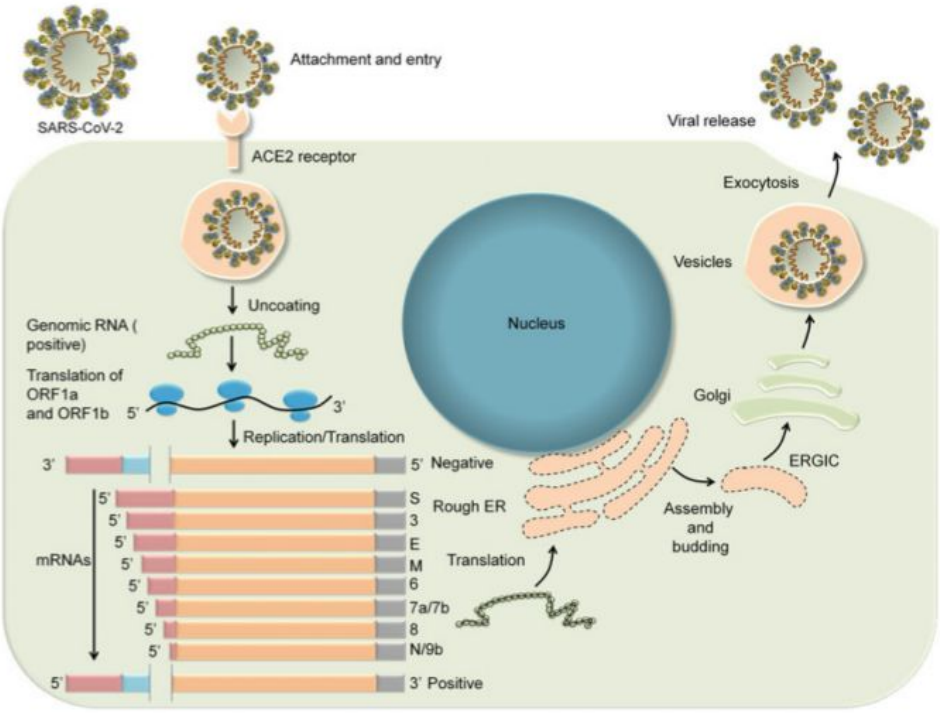


Figure 4: Stepwise mechanism of SARS CoV-2 replication within the host system.

2.4: Transmission Cycle of SARS CoV-2:

SARS CoV-2 is, as mentioned earlier, is a virus of zoonotic origin ; implying a pathogen that is primarily specific to infect animal hosts, but under specific conditions can infect humans. Scientists now believe that like previous members of the *Coronaviridae* family , SARS CoV-2 has also experienced a 'Zoonotic spillover' ; a term referring to the fact that the pathogen was transmitted to humans from a vertebrate animal , in this particular case, from bats. Although, the underlying mechanism behind such a spillover is still under study and still quite a mystery to the scientific community, it's quite certain that phylogenetic characteristics and host susceptibility does play an important role in this regard (Plowright et al. 2017).To get a clear understanding of

the underlying mechanism behind SARS CoV-2 transmission , the process was observed from 3 different perspectives.

(i)Transmission among animals.

(ii)Transmission from animals to humans.

(iii)Transmission among humans.

2.4.1:Transmission among animals:

In 2005, two mutually independent research groups reported the existence of novel CoVs associated with SARS CoV. Termed as SARS related CoVs or SARS CoV related viruses , these viruses were claimed to be found in species of horseshoe bats(*Rhinolophus)* (Lau et al. 2005; Li et al. 2005).These studies lead to an understanding within the scientific community that bats may have been the primary host for SARS CoVs , whereas with civets acting as intermediary carrier in between. It was also confirmed beforehand that the CoV genome is often prone to genetic recombination (Lai and Cavanagh 1997). They had also suggested a highly probable event of genetic recombination to occur in the viral genome within the bats in the similar or between different communities.The evolution of a new genera of specimen of SARS CoV was also speculated via genetic recombination among bats and passed on to farmed civets via fecal-oral routes before being passed on to humans through the wet markets of Guangdong. (Cui, Li & Shi, 2019)

Phylogenetic studies on novel CoV specimens also suggest the occurrence of cross-specimen transmision mostly induced by transient spillovers. High recombination frequencies for CoVs in bats points them as a vital and an ideal reservoir for their evolution as well(Banerjee, Kulcsar, Misra, Frieman & Mossman, 2019).

2.4.2:Transmission from animals to humans:

Currently, scientists confirm > 95% genomic homology between SARS CoV-2 & bat CoV; with an implication that bats are indeed the most probable primary host to this virus (Perlman 2020; Zhou et al. 2020). Other animals reported to probably have acted as potential reservoirs for this pathogen include snakes (Ji et al. 2020) , pangolins (*Manis javanica*) & minks (Lam et al.2020).

As mentioned earlier, the spread of SARS CoV-2 is presumably a consequence of zoonotic spillover.The process itself is supposedly carried out by the induction of a chain of events facilitating the virus to infect human hosts. Animal-human transmission is itself influenced by numerous factors such as disease dynamics in human hosts,level of viral exposure and susceptibility of the human race to the virus in a particular region.These factors can be explained as a tri-stage mechanism depicting the viral transmission process; **The primary stage** defines the pathogenic pressure on human host system. That is, the amount of virus interacting with the human hosts at a particular instant influenced by viral prevalence, dispersal from animal hosts followed by survival, development & distributional features. I.e. to sum up how the virus will adapt itself within the secondary host system outside the primary host, in this case preferably the bats.**The secondary stage** is defined by factors such as behaviour between human host and the vector. These factors define the probabilistic chances of viral exposure, entry routes and viral dosage.**The tertiary stage** of pathogenesis is influenced by genetic, physiological and immunological factors of the affected individual along with the stage of pathogenic development within the individual host system(Plowright et al. 2017). These stagewise development altogether creates a virtual barrier towards the gradual evolutionary process of this virus that is further emphasized and necessitated by the mechanism of spillover inducing this pathological event in the first place.

2.4.3:Transmission among humans:

SARS CoV-2 is an airborne virus and uses routes similar to flu and common cold pathogens for transmission to human hosts. Media of pathogenic transmission in this case includes air droplets via coughing, sneezing ,vomiting or even stools as a means to transmission through the fecal-oral route (Gu et al. 2020; Holshue et al. 2020).

Commencement of viral infection occurs via interaction between specific host receptors and fusion with M proteins. Reports also suggest that the RBD of the S protein binds to the ACE2 receptor of the human host and is responsible for human-human transmission (Jaimes et al. 2020; Wan et al. 2020). Although information and reports regarding SARS Cov-2 transmission among

humans from pets are inadequate at this moment , a high possibility of such an event to occur in the near future does exist.

2.5:Understanding the pathogenesis of SARS CoV-2 from a physiological perspective:

(Li et al., 2020) hypothesized that the viral pathogenesis of SARS CoV-2 leading to gradual stagewise development of COVID-19 in a healthy individual by the process of viral sepsis. They formed this hypothesis by studying alveolar cells of affected patients at both early and late stages of viral attack.

Clinically, patients were observed to exhibit typical manifestations of shocks, extreme cold and weak peripheral pulse rate despite severe hypertensive situations. Metabolically patients often showed symptoms of acidosis, impsired liver and kidney functionality. Patients diagnosed with COVID-19 are a perfect fit to the features mentioned in Sepsis-3 International Consensus (Singer et al.,2016). Critical pathogenic infection of bacterial or fungal origin were found negative in 76% of patients' blood and lower respiratory tract cultures implying that viral sepsis would be terminologically appropriate for critically affected patients.

From an immunological perspective, severe cytokine storms were observed in COVID-19 affected patients exhibiting elevated levels of IFN-ɣ induced protein-10 , IL-1β , granulocyte colony inducing factor , monocyte chemoattractant protein-1 and macrophage inflammatory protein -1α.This is a significant immunopathological feature of SARS CoV-2 pathogenesis homologous to H_1N_1;i.e. Common flu.(Li et al., 2020) also state that based on previous studies they state that lung epithelial cells,macrophages and dendritic cells all express cytokines at varied degrees using receptor based mechanisms.Members of the toll-like receptor family play a prominent role in this regard;like TLR-7,TLR-8 etc.Still the underlying cause responsible behind this massive cytokine storm is still a mystery itself.Anti-IL-6R monoclonal antibodies are also presumed to induce a primary response against COVID-19 by neutrophil mediated viral clearance. IL-6 & IL-6 R deficit also tend to induce flu infection and finally death in mice(Dienz et al.,2012) while use of corticosteroids against viral infections among physicians is still a controversial issue till this day(Russel et al.,2020;Shang et al.,2020).

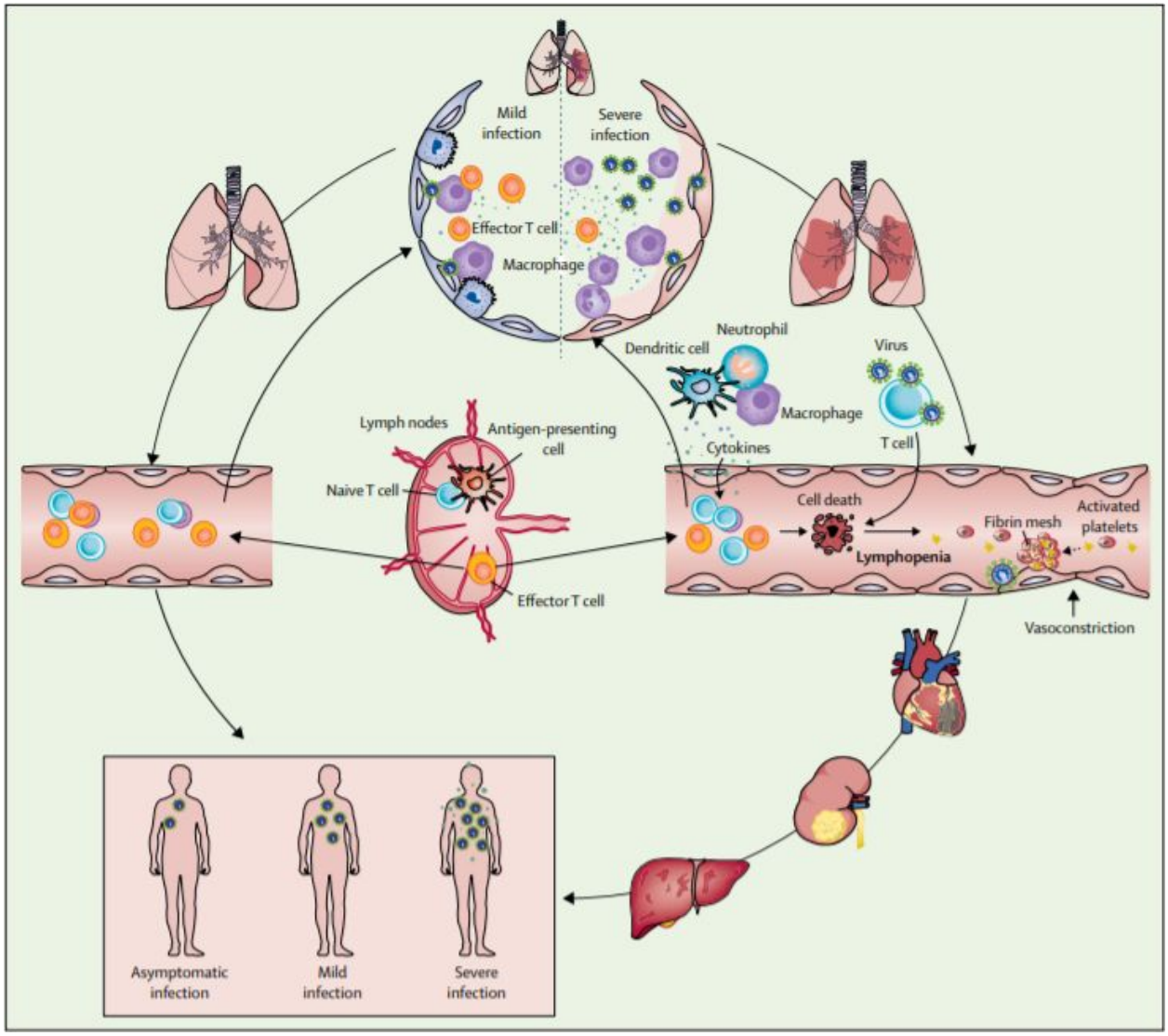


Figure 5: Schematic diagram representing Viral Sepsis of SARS CoV-2 within human host.

The dysregulated immune system goes through an immuno-suppressive stage post pre-inflammatory stage. This stage exhibits reduced cell count of $CD4^+$ and $CD8^+$ T cell lymphocytes leading to secondary opportunistic pathogenic infections in COVID-19 patients particularly at the later stage of attack and is correlative to the degree of infection severity(Liu et al.,2020).The condition itself is called lymphopenia.All these findings along with the presence of SARS CoV-2 RNA in blood samples of severely affected collectively led the hypothetical assumption that RBD of SARS CoV-2 S protein is capable of promoting T cell killing.Furthermore, the virus lacks ACE-2 receptor in the case of T cell infection post cytokine

storm. This points to the existence of an alternative mechanism to infection without the usage of ACE-2 receptors.

Based on the study of cellular biopsy and autopsy results of COVID-19 patients, it was understood that proper utilization of both innate and adaptive immunity led to the faster recovery of COVID-19 patients mostly at the early infection stages. In severe cases , interruption in the air-blood barrier was seen ultimately leading to acute respiratory and pulmonary arrest.Cells of both endothelial and epithelial types were attacked leading to a large amount of plasma components to be exodated in the alveolar cavity resulting in acute inflammation. With time, the degree of inflammation increases with time, particularly due to the fact that neutrophils and macrophages become hyper chemotactic to remove viral exudates. T cell lymphocytes that has become dysfunctional also contribute to adaptive immune failure followed by infiltration by macrophages and internal organ failure like lungs and kidneys.Altogether these hyperinflammatory actions being carried out by our 'guard cells' ultimately results in severe viral sepsis and host death.

2.6:Global situation of the COVID-19 pandemic & Evaluation of potential scope to overcome the situation:

As mentioned earlier, COVID-19 has been declared a global pandemic 30th January 2020. With a tentative mortality rate of around 3% as declared by WHO, the efficacy factor behind the pathogenicity of this virus lies within the fact that this virus can remain dormant within the host system for about 2 weeks.Furthermore, the degree of onset of attack depends on multifarious dimension of factors.A very important aspect in this context is the host immunity.As it was previously discussed, symptoms of COVID-19 in a 'seemingly healthy host' at the primary stage is basically homologous to those found in common flu infected patients.These include uneasiness, mild to moderate fever, cough etc, leading to severe respiratory arrest to multi-organ failure in the most severe attacks.Adaptive immune failure via promotion of acute inflammation within host system is a notable mechanism of viral pathogenesis of SARS CoV-2 within the host system.Individuals with weaker immunity such as infants and senior citizens are highly prone to

COVID-19.Simply put, to end this pandemic situation,it's essential to find a way to eradicate this disease.Various approaches from various perspectives are being experimented upon laboriously by the scientific community with this end in view.

Here, we tried to develop a multi-epitope subunit vaccine using *in silico* against SARS CoV-2 targeting its membrane glycoprotein (M) as a potential candidate in this regard.Furthermore, we developed this vaccine with a view to promote innate immunity to healthy individuals against the virus.

Chapter 3

Materials and Methods

3.1: Brief Overview to Applied Methodologies:

As mentioned earlier in the abstract, here we have attempted an *in silico* approach to develop a multi-epitope subunit vaccine against SARS CoV-2.In this respect, we tried to replicate the procedural methodologies applied by (Shey et al., 2019) and (Nain et al.,2019) to develop vaccines against onchocerciasis and *Elizabethkingia anophelis* respectively.Despite this fact, we modified our methodologies from certain aspects with a view to optimize the end-product of our vaccine in terms of stability and efficacy of our developed vaccine.

Procedurally speaking, instead of performing a complete proteome driven search approach to find the most suitable antigen as done by (Shey et al., 2019) and (Nain et al.,2019) based on antigenicity of a protein . We took a rather simplistic approach in that respect. We didn't go for whole proteome based screening; instead, we initially classified the proteins present in the proteome in two large categories.

(i)Structural Proteins.
(ii)Non-structural Proteins(nsp).

NSPs are proteins that are responsible for effective functionality the complete viral proteome.NSP 9 for example is the viral replicase of SARS CoV-2.Next we picked a viral protein to act as the primary antigen and the potential backbone of our vaccine. As mentioned earlier, a subunit vaccine is developed around a viral antigen instead of the whole viral proteome as in the case of a whole cell vaccine.We chose a structural viral protein that was highly antigenic , so that even if the antigenicity is lowered during vaccine preparation, the immune response stimulated in the host is sufficiently strong enough to promote innate immunity against SARS CoV-2 in the host , and yet the response is not vitally strong to kill the host in the process.Secondly, another purpose behind the selection of a structural protein as the primary

protein was that, a non-structural protein is often capable of exhibiting prolonged dormancy , as in the case of HIV. The SARS CoV-2 is also capable of exhibiting dormancy within a host for a period of 14 days or 2 weeks before the host begins to carry COVID-19 symptoms.Furthermore, the protein we select should be sufficiently long in the primary structure so it's physically stable and still not potentially large enough with a significantly high molar mass that the final vaccine exhibits non-specificity during administration to the patient.Considering all this facts, the membrane glycoprotein (M) of SARS CoV-2 was used as the primary antigen in our vaccine after checking its antigenicity on *Vaxijen 2.0* server at a minimal threshold of 0.5 and screening from other structural proteins.

After selection of primary antigen, we took the FASTA formatted sequence of our selected primary antigen as input in the *Net CTL 1.2* server for detection cytotoxic T-cell lymphocytic epitopes.CTL epitopes were sorted out based on parameters of MHC-I binding affinity, combined score , TAP efficiency, weighted C-terminal cleavage efficacy at a minimal threshold of 0.75 overall combined score. Using respective IEDB servers for MHC I and MHC II allele selection respectively , we chose the corresponding peptides using percentile rank as measure of binding affinity to the antigen.The lower the percentile rank, stronger the binding affinity of the peptide to the antigen and higher the antigen presenting efficacy.For MHC I peptide selection , the peptides got as output from *Net CTL 1.2* servers were used as input, whereas the primary antigen (i.e. the M protein) was taken as input selecting 15 as default length(in aa residue length).For the selection of MHC I epitopes, a maximal threshold of 2% and for MHC II epitopes, a maximal threshold of 0.5 % were taken respectively.For MHC II peptides a lower percentile rank was chosen because the peptides were highly scattered within a smaller interval than in MHC I molecules. Homologous and nearly homologous epitopes were screened out in every case and repeated peptides were screened out.Capacity of the MHC II epitopes to induce IL-4 and IL-10 production were measured using IL-4 pred and IL-10 pred servers to see if they are capable of inducing proper immune response within the host and thus boost immunity.Finally, to complete epitope identification for our vaccine construct, B cell epitopes specific to our primary antigen were selected using Biopred Linear Epitope Prediction 2.0

algorithm.This algorithm was carried out at a 0.5 threshold.All these epitopes identified were then catenated in a linear primary protein structure of our vaccine using proper linkers at definite intervals to promote stability and expression efficacy.After the primary structure of the vaccine construct was completed,its biochemical analysis analysis was conducted by using the PROTPARAM tool along with that of the primary antigen as well. For this purpose, we used a FASTA sequence of protein as the input.Further analysis was carried out on a PDB formatted file of our vaccine. This file contained a 3D structure of our vaccine developed on the principle of homology modelling using Phyre 2 server.Once converted, the PDB model was used to plot a Ramachandran plot to predict a mathematically probabilistic model of the vaccine.Furthermore a residue-wise Z-score vs residue curve was plotted with the mean Z-score calculated.Next, performed molecular docking between TLR-8 receptor of toll like family of receptor and our vaccine in PDB format.

The last step marking the ending of our research work involved determining the allergenicity of our vaccine and conducting an *in silico* simulation of our vaccine to see if our vaccine is capable of boosting host immunity to host against SARS CoV-2.For this purpose the *AllergenOnline* server and C-imsim online immuno-simulator were respectively used.A schematic diagram of our workflow is given below:

3.2:Selection Of Primary antigen:

To explain the applied methods and the online tools at every step of the research, it's better to take a step back and start from the basics.Our goal here was to develop a subunit vaccine against SARS CoV-2.Now, unlike a whole cell vaccine where the entire proteome of a pathogen is introduced in the host with reduced pathogenicity, subunit vaccines are developed utilising specific proteins in the pathogenic proteome. Along with the antigenic protein, specific epitopes are also added at specific amounts to optimize the pathogenicity of the final vaccine.Linkers are also added in this regard, which are fusion proteins used as enhancers of stability and expression of protein etc.The pathogenicity of the end product is usually lower than that of the crude protein derived from the pathogen , yet at an optimal level to enhance sufficient immune response within the host.

In this step we screened the proteome of SARS CoV-2 for a suitable protein candidate that will be highly antigenic enough to boost a strong immune response in the host, yet won't be potentially lethal to the host. We screened the proteins available looking for an ideal candidate such that the protein will potentially be a structural protein so it can be easily used as an identifier for the pathogen.Secondly targeting a non-structural protein would not be practically feasible as the SARS CoV-2 can remain latent in host system for up to 2 weeks.

Initially we studied the morphological features of the virus and identified the potential proteins that can be used as the potential backbone of the vaccine.Next we detected the potential antigenicity of the proteins.The primary structure of the proteins were obtained from NCBI website located at https://www.ncbi.nlm.nih.gov/ and protein sequences our desired POI were extracted in FASTA format.These sequences were then used as input for pathogenicity determination using the *Vaxijen 2.0* server located at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html.The server was set at a threshold of 0.5 and virus chosen as the target organism.

3.3:Identification of potential Cytotoxic T-cell Lymphocytic Epitopes:

Cytotoxic T-cell lymphocytes are highly efficient immune cells with TCR expressed on their surface capable of promoting immune response via inflammation aided by chemokine and cytokine production.Hence, T-cell epitopes act as a potential booster for vaccine mediated protection in the host.To detect CTL epitopes we used the *NetCTL 1.2* server located at http://www.cbs.dtu.dk/services/NetCTL/.This server detects CTL epitopes specific to a particular antigen it takes as input in the FASTA file format.CTL epitopes are identified based on criteria of TAP efficacy, MHC I binding affinity and weight on C-terminal cleavage.The epitopes can be compared against 12 MHC supertypes.While the MHC I binding affinity and proteasomal cleavage capability are calculated using artificial neural networks, TAP transportation efficacy are calculated by using weighted matrix system.Placing weights on C-terminal cleavage, TAP transport efficacy at 0.15 and 0.05 respectively while the threshold for epitope prediction was set at 0.75.

3.4:Identification of MHC I Alleles:

For the identification of MHC I alleles we used the specific IEDB server located at http://tools.iedb.org/analyze/html/mhc_binding.html.In this step of research, we used the output of NetCTL server as input that were obtained in the former step.Percentile rank is used as a measure to signify the degree of binding affinity.The lower the percentile rank, the stronger the binding affinity between epitope and allele.The version of the server used here is 2009-09-01B.Epitopes with percentile rank ≤ 2 were selected.It's highly probable that the detected MHC I peptides may be homologous to CTL epitopes detected earlier.Such epitopes were excluded from further insertion to the vaccine construct.For the reference allele set used for this step we used

3.5:Identification of MHC II Alleles:

MHC II alleles are essential to promote humoral and cell mediated immunity in the host boosting its adaptive immunity further.For the purpose of MHC II peptide identification,the server at http://tools.iedb.org/mhcii/ was used. The primary antigen was used as the input here with a default length of peptide set at 15.Similar to the previous step,percentile rank was used as the potential indicator of binding affinity between the epitope and the antigen.However, this time the maximal possible threshold of percentile rank was taken to be ≤ 0.5 .This was done because the peptides were more varied in nature than MHC I epitopes.

3.6:Identification of IL producing capability of MHC II peptides(HTL epitopes):

In this step we primarily tried to detect the IL4 and IL10 inducing capability of the MHC II peptides. IL4 pred and IL10 pred servers were used for these respective purposes.These capabilities were further analysed during the immuno-simulation step carried out at the later part of this research work.

3.7:Identification Of B-cell Epitopes:

B-cell epitopes are capable of inducing humoral immunity by secreting immunoglobulins capable of neutralizing antigens upon binding with their B-cell receptors(BCR).Such epitopes can either be :

(i)Linear or continuous.

(ii)Conformational or discontinuous.

However, for vaccine designing, linear epitopes are usually preferred owing to ease of primary structure modification.Hence, linear B cell epitopes were identified in this case.B cell epitopes corresponding to our primary antigenic sequence were identified using Bepipred Linear Epitope Prediction 2.0 algorithm via the server located at http://tools.iedb.org/bcell/.Epitopes were identified at a threshold of 0.5 .

3.8:Construction Of the Final primary structure of vaccine construct:

The final construct of the vaccine was prepared based on the primary antigen which was collected from NCBI website located at https://www.ncbi.nlm.nih.gov/.The sequence obtained in FASTA format was taken and the primary structure of the protein also had the MHC I, MHC II and B cell epitopes added enhanced by proper linkers.Once completed, the final vaccine construct was ready to be analysed further from biochemical, immunological and statistical research.

3.9:Biochemical Analysis of designed Vaccine:

In this step we observed the biochemical properties of the vaccine constructed using the protparam tool under Expasy server.The server located at https://web.expasy.org/protparam/ uses primary structure of a protein, for this purpose the vaccine as an input and it will analyse its various biochemical parameters, these include the protein's molar mass, its extinction coefficient, theoretical PI,estimated half life,aliphatic index,aa composition,instability index and grand average of hydropathicity(GRAVY).The vaccine was also compared with respect to the crude protein as a control that was primarily obtained from the ncbi website.

3.10:Homology modelling of Vaccine construct to obtain 3D model:

Homology modelling was done to obtain a 3D structure of the vaccine construct from the primary structure prepared earlier based on entries in the PDB website located at https://www.rcsb.org/.In this step we used the phyre2 server located at

http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index.The model builds a 3D model of a primary protein structure with reasonably high confidence and also expresses homology of the obtained structure with respect to the used templates in terms of percentage.The PDB structure obtained is essential for analysis carried out in later steps.

3.11:Statistical Analysis Of the obtained Tertiary PDB structure and Ramachandran Plotting :

This step was carried out with the purpose of analyzing the tertiary structure obtained via homology modelling before further work with the protein.Here two basic functions were performed on the protein's tertiary model.First,the Ramachandran plot was formulated using the Rampage server located at http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php.This plot gives a probable location of the residues based on angles plotted on a quadrant system.This plot is a mathematical estimation of the distribution of aa in a particular protein.The input was the PDB file obtained from the phyre2 server while we got the plot and its interpretation as the output.

Furthermore, we plotted a z-score versus residue curve and calculated the mean z-score as a quality indicator for our tertiary model.

3.12:Molecular Docking Analysis between Vaccine construct & TLR-8 receptor:

Molecular docking is an *in silico* approach done in order to estimate the binding affinity between a receptor and its corresponding ligand resulting in what is called a protein-ligand complex.For this step we used the patchdock server located at https://bioinfo3d.cs.tau.ac.il/PatchDock/index.html.The server performs the desired docking function using an algorithm based on object recognition and image segmentation method used in image processing.The underlying algorithm consists of 3 steps:

3.12.1:Representation Of Molar Surface:

Here the surface was firstly analyzed followed by a segmentation algorithm used to identify the geometric patches.These patches that imply the concave, convex or plain surfaces that are filtered to obtain the 'hotspots'.

3.12.2:Surface Patch Matching:

A hybrid algorithm composed of geometric hashing & pose-clustering techniques are used in this step to match patches detected in the previous step.Convex patches are matched with concave patches while plane patches are matched with any kind of patches.

3.12.3:Filtering & Scoring of Docked complex:

The last step is the performance of the scoring function on the ligand complex at a particular transformation or geometric orientation.

The completion of the aforementioned steps completes the docking algorithm quantified by the scoring function,thus completing this step of the research.

3.13:*In silico* simulation of the final vaccine on human host:

The final step of this work is conducting an *in silico* simulation of the immune response within a human host.The simulation was conducted using the C-Immsim server located at http://kraken.iac.rm.cnr.it/C-IMMSIM/index.php?page=1.The server takes FASTA sequence of the proteinous vaccine as an input with dosage and simulation steps inserted as well as parameters.The simulation was conducted at 300 steps.A tri-dosage system was administered in the form of injection at steps 1,84 & 168 respectively.Actually the vaccine was aimed to be administered thrice at 28 days interval.

Step 1=8 hrs.

Step 84=672 hrs=Day 28

Step 168=1344 hrs=Day 56

Adjuvant and antigen (i.e. the vaccine parameters were set at 100µg & 1000µg respectively.

As an output we obtained various graphs that represented the immune response in a human host upon vaccine administration.The graphs represent the immune status of the host in terms of immune cell concentration, immunoglobulin concentration as well as antigen concentration with time.The immuno-simulation step carried out is an *in silico* approach to check the efficacy of the vaccine computationally before going on to analysis in living systems.

3.14:Checking the final vaccine construct for host allergenicity:

The finalized vaccine construct is then checked whether or not it's a potential allergen.For this purpose we had used the Allergen Online server located at http://www.allergenonline.org/.The server takes in FASTA formatted files of a protein as input and gives homologous protein sequences that can act as allergen as output,along with the allergenic potential particle name it resembles.This step is a necessary to minimize potential side-effects that can occur at later steps of the research.

3.15:Final Remarks on applied methodology:

On a conclusionary remark regarding our applied methodology,it must be said that the procedure we used in our research strictly involved the usage of *in silico* tools.More specifically,we had used online servers in every step of our research.Furthermore, we don't claim with 100% confidence that our end product is a efficient vaccine that will completely eradicate COVID-19 disease.However, we do believe that it can be a potential vaccine candidate and further analysis is required in this respect, particularly in wet lab and clinical levels.

Chapter 4

Result Interpretation & Analysis

In this part of our research work,we are going to chronologically present the results we obtained at every step of conducting the experiment.Furthermore, we will interpret the obtained result so the reader finds ease in assessing the quality of work done in this regard.As well as that, we will also try to identify potential scopes for future work that remain hidden within this work.

4.1:Antigenicity Prediction:

For the purpose of antigenicity prediction we had used *Vaxijen 2.0* server.The E protein,S protein, M protein & N protein among structural proteins and nsp9 and nsp6 were chosen primarily for antigenicity prediction ; these proteins were tested for antigenicity at a threshold of 0.5 with virus as target organism.The GI number , protein name and FASTA sequence along with their antigenicity values are tabulated below:

Accession number & version	**Protein Name**	**Number of aa present**	**Antigenicity**
YP_009742613.1	NSP6	290	0.5657
YP_009742616.1	NSP9	113	0.6476
QIC53213.1	Spike Protein(S)	1273	0.4676
QJA17755.1	Membrane glycoprotein(M)	222	0.5158
QIH45060.1	Nucleocapsid Protein(N)	419	0.5093
QJQ39994.1	Envelope Protein(E)	75	0.4398

Table 3: Proteins primarily analysed with their aa length and relative antigenicity.

Identification was calculated with ‘virus’ selected as the target pathogen.NSP9 is typically the viral replicase;i.e. The RNA dependent RNA polymerase responsible for RNA replication of

SARS CoV-2 and shows the highest antigenicity among the proteins used.NSP6 on the other hand is capable of binding to σ-1 receptor that is also a target of the drug chloroquine ("Chloroquine's use to treat COVID-19 is backed by US government, but many questions remain", 2020), a class of drugs previously considered effective against the virus.

Despite these facts, the membrane glycoprotein (M) was chosen as the feasible target in this regard, particularly due to the fact that it's firstly a structural protein and can easily be used as a vaccine target.Secondly, the M protein has also been used previously in development of vaccines against various kinds of flu viruses.Hence, the M protein was chosen as the primary antigen after thorough screening.Primary structure of our primary antigen is given below:

MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPVTLACFVLAAV
YRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVI
GAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDH
SSSSDNIALLVQ

A screenshot of providing input on the *Vaxijen 2.0* server and the corresponding output are also as follows:

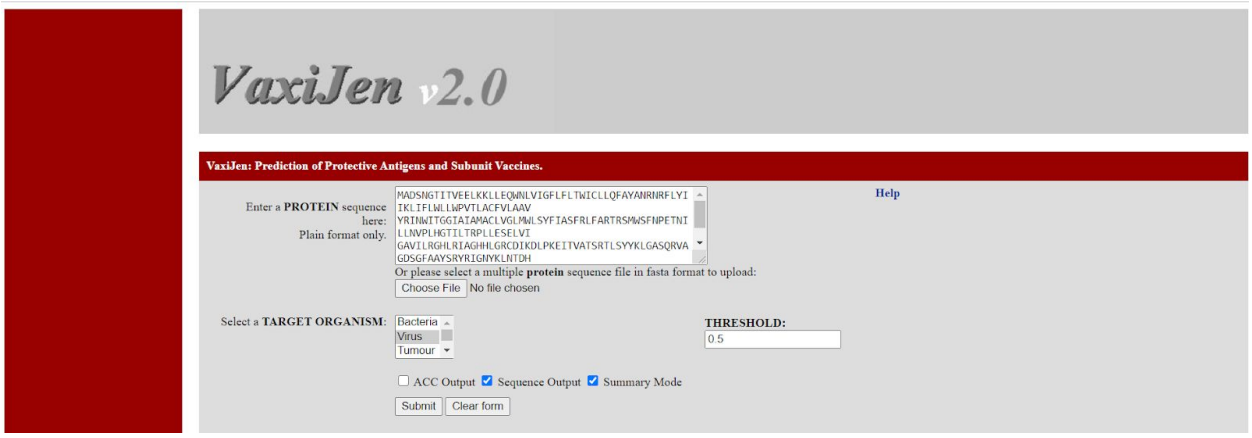


Figure 6: Entry of primary structure of antigen into the *Vaxijen 2.0* server.

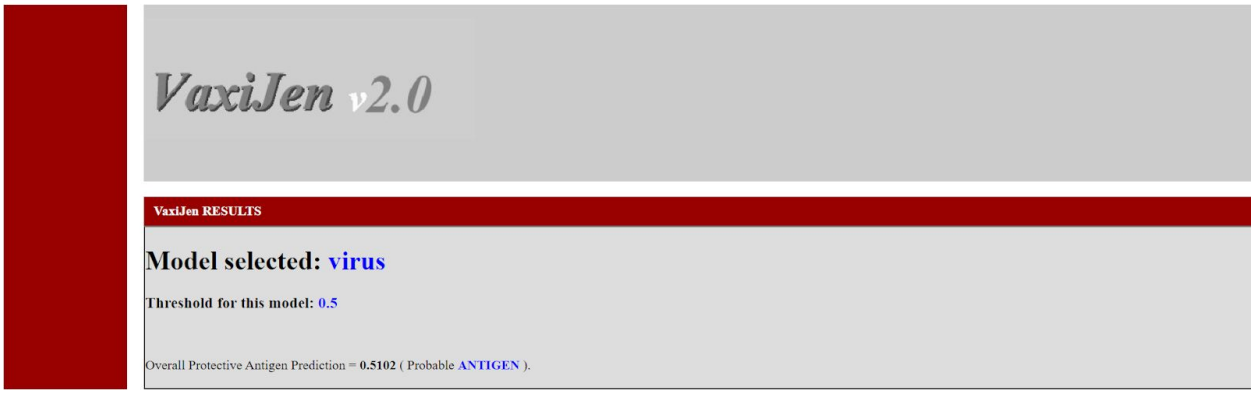


Figure 7: Screenshot of obtained result on *Vaxijen 2.0* server.

4.2:CTL Epitope Identification:

CTL epitopes are essential for generation of enhanced immune response within the host system.They are highly ag-specific and express TCRs that can fight the particular ag. CTL epitopes were identified using the *Net CTL 1.2* server.At a threshold level of 0.75, epitopes were identified of the A1 supertype of MHC I alleles.The underlying algorithm used in this case is built upon the principle of neural networks.The criterion of combined score is the major determinant of epitope selection.This combined score is assigned to the corresponding epitopes based on parameters of C terminal cleavage & TAP transport efficiency.A minimal threshold of 0.15 & 0.05 were respectively set in this regard.The final data are tabulated below:

Table 4: Primary structure of CTL epitopes with corresponding combined scores.

Sequence of CTL epitope	**Combined Score**
SSDNIALLV	2.9325
ATSRTLSYY	2.6146
YSRYRIGNY	1.6623
YANRNRFLY	1.6155
VATSRTLSY	1.4642

Sequence of CTL Epitope	Combined Score
WICLLQFAY	1.4105
LVGLMWLSY	1.3974
AGDSGFAAY	0.8480
SSSDNIALL	0.8308
LLEQWNLVI	0.7745

Table 4: Primary structure of CTL epitopes with corresponding combined scores.(Continued).

Combined scores mentioned above can be used to express sensitivity & specificity of corresponding epitopes as follows:

Score	Sensitivity	Specificity
>1.25	0.54	0.993
>1.00	0.70	0.985
>0.90	0.74	0.980
>0.75	0.80	0.970
>0.50	0.89	0.940

Table 5: Combined score expression system in terms of sensitivity & specificity parameters.

A screenshot of the result obtained in the *NetCTL 1.2* server is given below:

Figure 8: Screenshot of results from *NetCTL 1.2* server with selected epitopes marked in red.

4.3:MHC I allele specificity of CTL Epitope:

Epitopes that were identified in section 4.2 were used as input in this step to obtain MHC I alleles using the IEDB server assigned for this purpose.The version 2009-09-01B was used for this purpose.The parameter used in this regard of epitope selection is percentile rank.A lower percentile rank implies to a higher binding affinity and vice versa.We had chosen a minimal threshold of ≤ 2 for epitope selection in this regard.Allele specific binding of CTL epitope and MHC I alleles are tabulated below with corresponding binding affinity in terms of percentile rank.

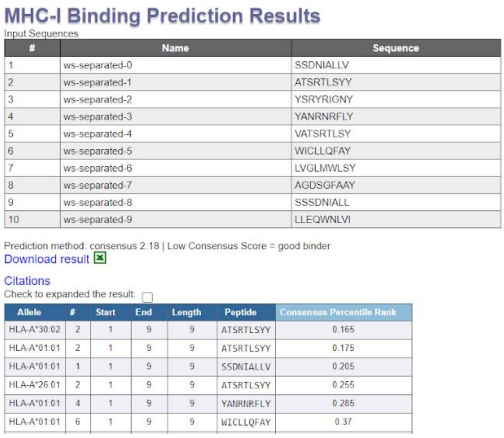
Table 6:MHC I Peptides along with alleles covered with corresponding percentile rank.

Peptide sequence	**Alleles covered**	**Percentile Rank**
SSDNIALLV	HLA-A*01:01	0.205
ATSRTLSYY	HLA-A*30:02,HLA-A*01:01,HLA-A*26:01,HLA-A*03:01,HLA-B*57:01,	0.165,0.175,0.255,1.475,2.0
YSRYRIGNY	HLA-A*26:01,HLA-A*30:02,HLA-A*11:01,HLA-B*15:01,HLA-A*01:01	0.395, 0.49, 0.685 , 0.8,1.95
YANRNRFLY	HLA-A*01:01,HLA-B*35:01,HLA-A*30:02,HLA-A*11:01,HLA-B*01:01,HLA-B*58:01	0.285, 0.4,0.585,1.1
VATSRTLSY	HLA-B*35:01,HLA-A*01:01	0.9,1.05,1.1,1.5,1.65,2.0

	,HLA-B*58:01,HLA-B*53:01,HLA-A*30:02,HLA-B*15:01	
WICLLQFAY	HLA-A*01:01,HLA-B*35:01 ,HLA-A*30:02,	0.37,0.9,1.5
LVGLMWLSY	HLA-A*01:01,HLA-B*15:01 ,HLA-B*35:01,	0.645,1.2,1.8
AGDSGFAAY	HLA-A*30:02 , HLA-A*01:01 ,HLA-B*35:01	0.51,0.75,1.9
LLEQWNLVI	HLA-A*32:01	2.0

Table 6 (Contd.):MHC I Peptides along with alleles covered with corresponding percentile rank.

To wrap up this section , we place a screenshot of the results obtained at this step via the IEDB server.



MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	ws-separated-0	SSDNIALLV
2	ws-separated-1	ATSRTLSYY
3	ws-separated-2	YSRYRIGNY
4	ws-separated-3	YANRNRFLY
5	ws-separated-4	VATSRTLSY
6	ws-separated-5	WICLLQFAY
7	ws-separated-6	LVGLMWLSY
8	ws-separated-7	AGDSGFAAY
9	ws-separated-8	SSSDNIALL
10	ws-separated-9	LLEQWNLVI

Prediction method: consensus 2.18 | Low Consensus Score = good binder

Download result

Citations

Check to expanded the result:

Allele	#	Start	End	Length	Peptide	Consensus Percentile Rank
HLA-A*30:02	2	1	9	9	ATSRTLSYY	0.165
HLA-A*01:01	2	1	9	9	ATSRTLSYY	0.175
HLA-A*01:01	1	1	9	9	SSDNIALLV	0.205
HLA-A*26:01	2	1	9	9	ATSRTLSYY	0.255
HLA-A*01:01	4	1	9	9	YANRNRFLY	0.285
HLA-A*01:01	6	1	9	9	WICLLQFAY	0.37

Figure 9:A screenshot showing a slice of the epitopes analysed for functioning as MHC I peptides.

4.4:MHC II allele identification for HTL epitopes:

MHC II allele are essential players in the enhancement of adaptive immunity of a host, since they do so by binding to HTLs and thus generate humoral & cell mediated immunity.MHC II

alleles are identified using an IEDB server as mentioned earlier.MHC II alleles were identified using the primary antigen as an input.The indicator for MHC II allele identification is also percentile rank; a lower percentile rank indicates a higher binding affinity and vice versa.However, in this case a percentile rank of ≤ 0.5 was used for allele identification.The results are tabulated as follows:

Table 7:HTL epitope with reference to corresponding allele & respective percentile rank.

MHC II peptides	**Alleles covered**	**Percentile Rank**
GLMWLSYFIASFRLF	HLA-DPA1*01:03/DPB1*02:01	0.05
LSYYKLGASQRVAGD	HLA-DRB1*09:01	0.06
IKLIFLWLLWPVTLA	HLA-DQA1*01:01/DQB1*05:01	0.07
NRFLYIIKLIFLWLL	HLA-DRB4*01:01	0.12
LVGLMWLSYFIASFR	HLA-DPA1*01:03/DPB1*02:01	0.25
NLVIGFLFLTWICLL	HLA-DPA1*01:03/DPB1*02:01,HLA-DPA1*03:01/DPB1*04:02,	0.26 , 0.29
SYFIASFRLFARTRS	HLA-DRB1*11:01	0.39
EQWNLVIGFLFLTWI	HLA-DPA1*01:03/DPB1*02:01	0.47

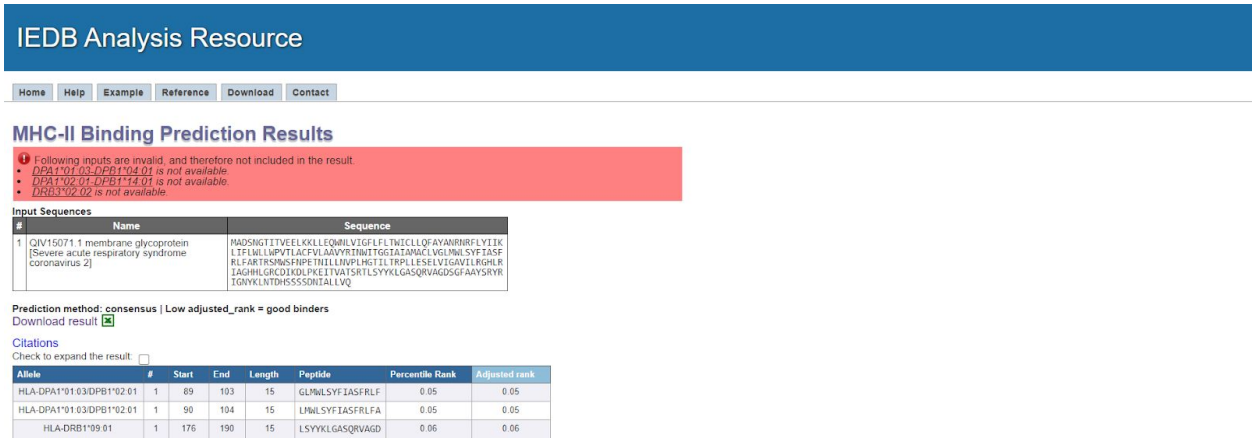
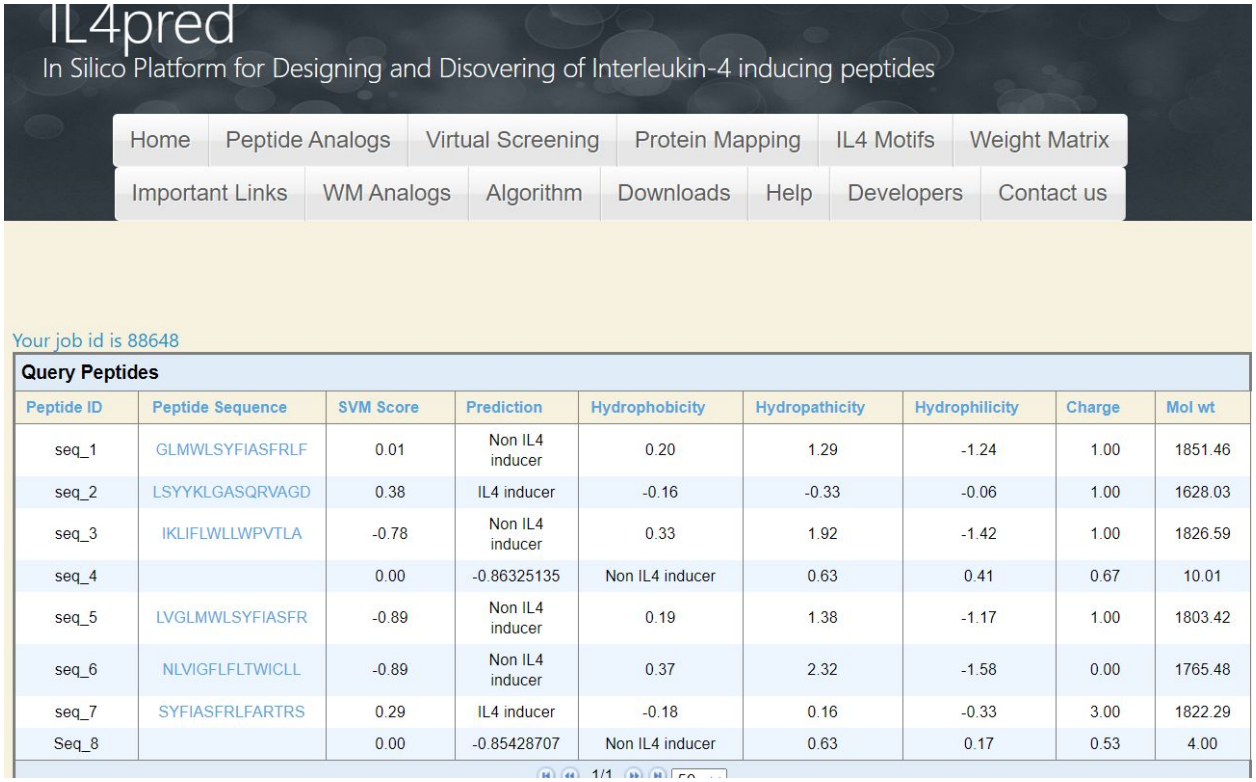


Figure 10: MHC II allele identification specific to HTL epitopes against vaccine antigen.

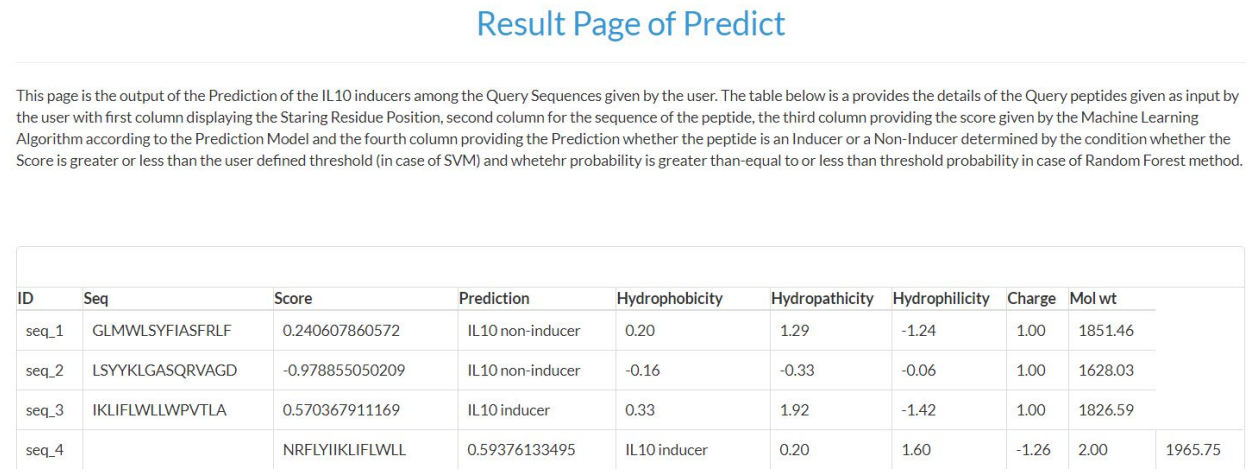
4.5:Cytokine inducing capability of HTL epitopes:

HTL epitopes help induce CTL activation by generating cytokines that also induce inflammatory immune response.Initially we have observed the production of interleukins by HTL epitopes, in particular IL-4 & IL-10.The servers named IL-4pred and IL-10pred were used in this step.A screenshot of the obtained results from the servers are given below followed by a combined tabulation.It should be mentioned that a threshold of 0.2 and -0.3 by default using SVM method was used for IL4 pred and IL10 pred servers respectively.This is done as a process to optimize algorithmic efficacy.Further analysis was carried out at the later steps.



Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq_1	GLMWLSYFIASFRLF	0.01	Non IL4 inducer	0.20	1.29	-1.24	1.00	1851.46
seq_2	LSYYKLGASQRVAGD	0.38	IL4 inducer	-0.16	-0.33	-0.06	1.00	1628.03
seq_3	IKLIFLWLLWPVTLA	-0.78	Non IL4 inducer	0.33	1.92	-1.42	1.00	1826.59
seq_4		0.00	-0.86325135	Non IL4 inducer	0.63	0.41	0.67	10.01
seq_5	LVGLMWLSYFIASFR	-0.89	Non IL4 inducer	0.19	1.38	-1.17	1.00	1803.42
seq_6	NLVIGFLFLTWICLL	-0.89	Non IL4 inducer	0.37	2.32	-1.58	0.00	1765.48
seq_7	SYFIASFRLFARTRS	0.29	IL4 inducer	-0.18	0.16	-0.33	3.00	1822.29
Seq_8		0.00	-0.85428707	Non IL4 inducer	0.63	0.17	0.53	4.00

Figure 11: Screenshot of IL4pred server results page obtained from input HTL epitopes.



ID	Seq	Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt	
seq_1	GLMWLSYFIASFRLF	0.240607860572	IL10 non-inducer	0.20	1.29	-1.24	1.00	1851.46	
seq_2	LSYYKLGASQRVAGD	-0.978855050209	IL10 non-inducer	-0.16	-0.33	-0.06	1.00	1628.03	
seq_3	IKLIFLWLLWPVTLA	0.570367911169	IL10 inducer	0.33	1.92	-1.42	1.00	1826.59	
seq_4		NRFLYIIKLIFLWLL	0.59376133495	IL10 inducer	0.20	1.60	-1.26	2.00	1965.75

Figure 12:Results obtained from IL10 server using HTL epitopes.

Sequence	IL4pred score	IL-4 inducer?	IL-10 score	IL-10 inducer?
GLMWLSYFIASFRLF	0.01	No	0.240607860572	Yes
LSYYKLGASQRVAGD	0.38	Yes	-0.978855050209	No
IKLIFLWLLWPVTLA	-0.78	No	0.570367911169	Yes
NRFLYIIKLIFLWLL	-0.86	No	0.59376133495	Yes
LVGLMWLSYFIASFR	-0.89	No	0.451505905812	Yes
NLVIGFLFLTWICLL	-0.89	No	0.586694517445	Yes
SYFIASFRLFARTRS		Could not be determined .	0.316075950101	Yes
EQWNLVIGFLFLTWI	-0.85	No	0.28289440398	Yes

Table 6 : Table showing IL4 & IL10 production capabilities of MHC II peptides .

4.6:Identification of B-cell epitope:

B-cell epitopes can either be continuous or discontinuous by nature.These epitopes are essential immuno-boosters for the host.This is due to the fact that these cells possess B-cell receptors(BCR) that can bind to specific ag and boost immune response by immunoglobulins of various types.B-cell epitopes can help promote both humoral & cell mediated immunity and hence enhance adaptive immunity as a whole.Here we have used linear B-cell epitopes specific to the viral ag using the Bepipred linear epitope identification 2.0 algorithm was used in this regard.At 0.5 threshold,B-cell epitopes were identified.The starting and end position of the specific B-cell epitopes along with their respective lengths are tabulated below:

Sequence	Start	End	Length
NGTITVEELKKLLEQW	5	20	16

Sequence	Start	End	Length
AN	40	41	2
PLLESE	132	137	6
IKD	161	163	3
KLGASQRVAGDS	180	191	12
YRIGNYKLNTDHS SSSDNIA	199	218	20

Table 7 : Position Wise specification of B cell epitopes .

For further clarification of the predicted B cell epitopes, a graph was obtained from the server plotting residue wise score of the epitopes.The graph is as follows:

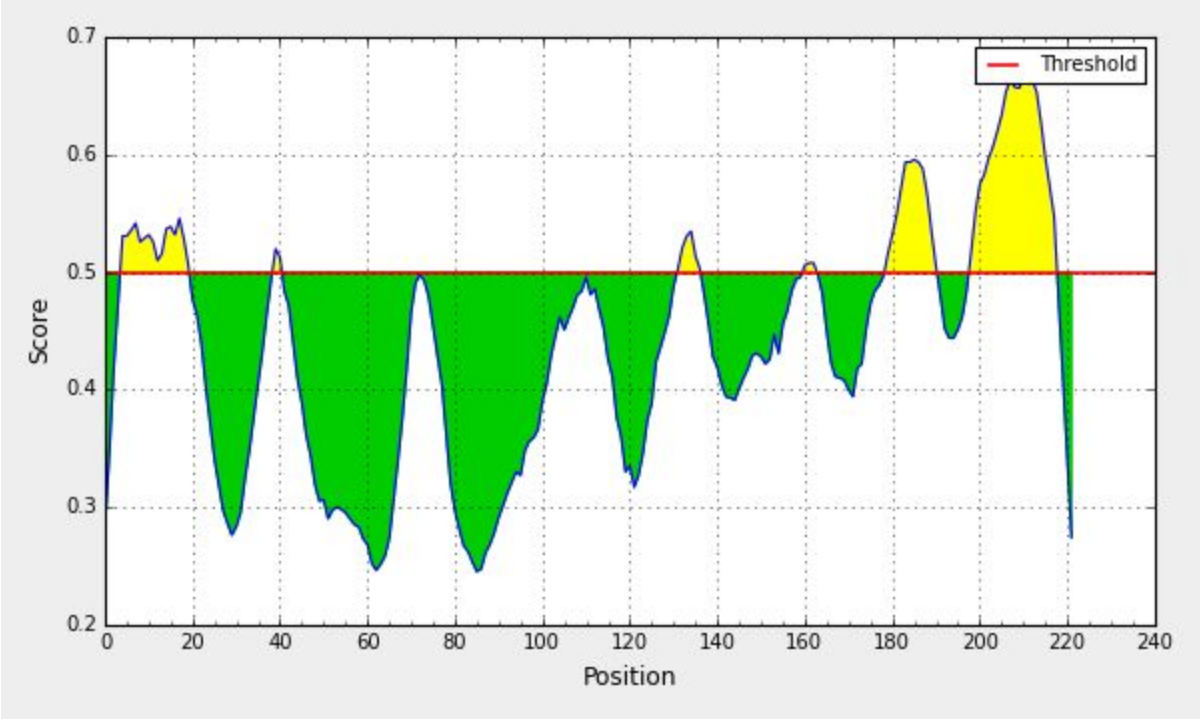


Figure 13:A score versus position graph for potential B-cell epitopes particular to the SARS CoV-2 M protein.

The Bepipred Linear Epitope 2.0 algorithm employed in this step is also capable of keeping track of the maximum,minimum & mean scores achieved by the set of linear B cell epitopes specific to a particular ag.For our particular AOI, we got a maximum score of 0.668, a minimal of 0.245 while on an average the score was around 0.435.

4.7:Construction of primary structure of Final Vaccine Construct:

As mentioned earlier,our primary objective in this work was aimed at the development of a multi-epitope vaccine against SARS CoV-2 utilizing the viral M-protein.The crude protein is also added with B cell epitopes,CTL epitopes and HTL epitopes.The B cell epitopes are essential to the promotion of humoral immunity and production of ig,IL & other specific antibodies;antibodies that act as potential enhancers to the generation of immune response.The CTL epitopes and HTL epitopes identified in the earlier phase of the study specific to MHC I & MHC II alleles respectively.The B cell epitopes are also capable of memory cell production that provides immunity to the host upon future encounter with that particular ag.

A subunit vaccine is developed upon the template of a particular antigenic protein specific to a particular pathogen.The protein is preferred to be highly antigenic.The addition of the aforementioned epitopes meet the end of optimizing the antigenicity of the vaccine to such an extent that it does not prove fatal to the host and yet provide immunity to the host system with substantial efficacy for a satisfactory time duration.Furthermore, the factor of sensitivity,specificity and stability of the final vaccine is also to be considered.This end is met by adding potential linker molecules;potential fusion proteins added to the vaccine template to boost overall efficacy of the vaccine.

In this work, we added EAAAK linkers between the CTL epitopes at 4 epitopes interval.AAY linkers were added between the 8 HTL epitopes at 4 epitopes interval as well.A KK linker was added at the end.A homologous vaccine designing approach was employed by (Shey et al., 2019) and (Nain et al.,2019) to develop vaccines against onchocerciasis and *Elizabethkingia anophelis* respectively.The final vaccine construct is as follows:

MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPV
TLACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILL
NVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLG
ASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQEAAAKSSDNIALLVATSRT
LSYYYSRYRIGNYEAAAKYANRNRFLYVATSRTLSYWICLLQFAYEAAAKLVGLMWLS
YAGDSGFAAYLLEQWNLVIAAYGLMWLSYFIASFRLFLSYYKLGASQRVAGDIKLIFLW
LLWPVTLANRFLYIIKLIFLWLLAAYLVGLMWLSYFIASFRNLVIGFLFLTWICLLSYFIAS
FRLFARTRSEQWNLVIGFLFLTWIGPGPGNGTITVEELKKLLEQWANPLLESEGPGPGIK
DKLGASQRVAGDSYRIGNYKLNTDHSSSSDNIAKK

4.8:Biochemical Analysis of Vaccine Construct:

In this step we used the PROTPARAM tool under the Expasy server to conduct biochemical analysis of the constructed vaccine.The server conducts the test and gives its results based on parameters such as molar mass, molecular formula, instability index (which is a measure of whether or not the protein is stable to work with), theoretical PI as well as aliphatic index and GRAVY and etc features.

Parameter Considered	**Results**
Number of Amino acids	515
Molecular mass	58511.54
Theoretical PI	9.56
Total negatively charged residues (Asp+Glu)	30
Total positively charged residues (Lys+Arg)	49
Molecular Formula	$C_{2745}H_{4205}N_{687}O_{704}S_{13}$
Estimated Half life	30 hrs (*in vitro* mammalian reticulocytes),>10hrs in *E coli* ,*in vivo* >20 hrs in yeast , *in vivo*
Extinction Coefficient.	14975 $M^{-1}Cm^{-1}$(At 280mm water column) if Cys residues form cystines.)149200 if all Cys residues are reduced.
Instability index	32.79(Stable)
Aliphatic index	115.24
Grand average of hydropathicity (GRAVY)	0.392 (Hydrophobic)

Table 8(Above): Biochemical analysis of the designed vaccine construct.

Parameter Considered	**Results**
Number of Amino acids	222
Molecular mass	25146.62
Theoretical PI	9.51

Parameter Considered	Result
Total negatively charged residues (Asp+Glu)	13
Total positively charged residues (Lys+Arg)	21
Molecular Formula	$C_{1165}H_{1823}N_{303}O_{301}S_8$
Extinction coefficient	30 hrs (*in vitro* mammalian reticulocytes),>10hrs in *E coli* ,*in vivo* >20 hrs in yeast , *in vivo*
Estimated Half life	14975 $M^{-1}Cm^{-1}$(At 280mm water column) .
Instability index	39.14
Aliphatic index	120.86
Grand average of hydropathicity (GRAVY)	0.446(More hydrophobic than vaccine.)

Table 9(Contd.): Biochemical analysis of the primary antigen.

Biochemical analysis of the vaccine construct was done by using its primary structure of the vaccine as input and outputs were obtained accordingly.Placing input into the server, looked somewhat as follows:

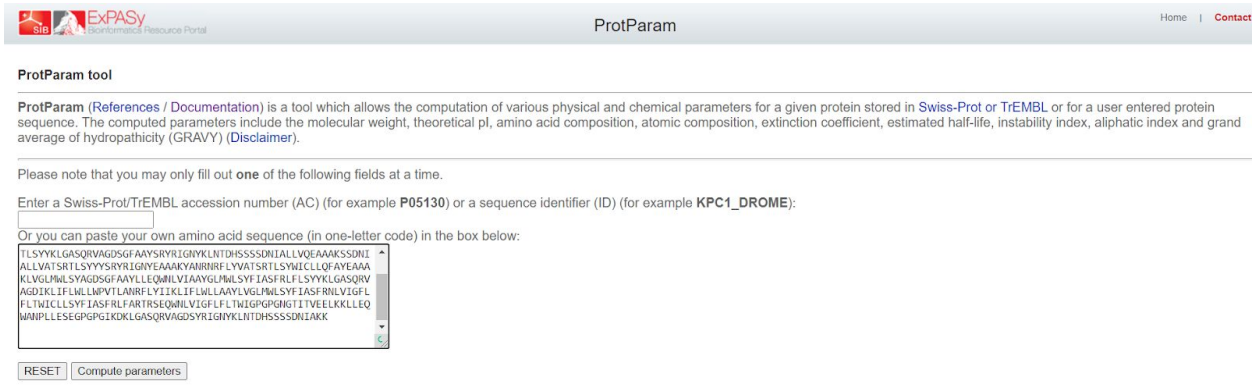


Figure 14:A screenshot of entry of primary structure of vaccine to the Expasy server.

Figure 15: Output obtained from Expasy server based on vaccine construct.

For the sake of further clarification,the primary structure of the antigen (i.e. the M protein) was also undergone biochemical analysis for a comparative analysis with respect to the final vaccine;the data are tabulated in table 8.From the analysis done it was evident that both the vaccine construct (as shown in table 8) and the primary antigens(as shown in table 9) are stable by mass and thermosensitivity.The aliphatic index for the vaccine & crude M protein are 115.24 & 120.86 respectively.While the stability indexes are at 32.79 & 39.14 respectively.Again,the **Grand Average Of Hydropathicity(GRAVY)** values for the vaccine construct & crude M protein are respectively 0.392 & 0.446.

From these results obtained, we can infer that since an instability index of our vaccine is lower than the crude protein and although in both cases it's <40, both are stable by molar mass;however, our vaccine construct is more stable than natural M protein of SARS CoV-2.However,in terms of thermal stability with respect to the aliphatic indexes of the respective proteins,the vaccine is potentially less stable than the crude protein owing to its lower aliphatic index than the crude one.Lastly,in terms of GRAVY values,the GRAVY value of the vaccine is lower than the raw protein.However,it would be preferred that the GRAVY value of vaccine be negative making it hydrophilic.Since hydrophobic,potential risk of contamination and loss of protein functionality particularly during purification & downstream processing.It should be said during purification of hydrophobic proteins,ethanol and such homologous solutions are often used for purification that can often be hazardous to health of the handler as well as the

environment.Proper optimization protocols are required to enhance the protein further in terms of the aforementioned parameters.

4.9:Determination of Allergenicity of Vaccine:

Often a vaccine can cause the induction of autoimmune response within the host system.This may result in the trigger of unintended higher immune response with physiological complexities accompanied with toxic side effects.To estimate the allergenic potency of our vaccine we had used the AllergenOnline server ;for an efficient estimate of the protein's potency we had used a cut-off value of 0.5 based on z-score analysis & the Full FASTA 36 method was used in this respect.Fortunately, no potential allergenic homologue was not discovered.The entry screen and output obtained looked respectively as follows:

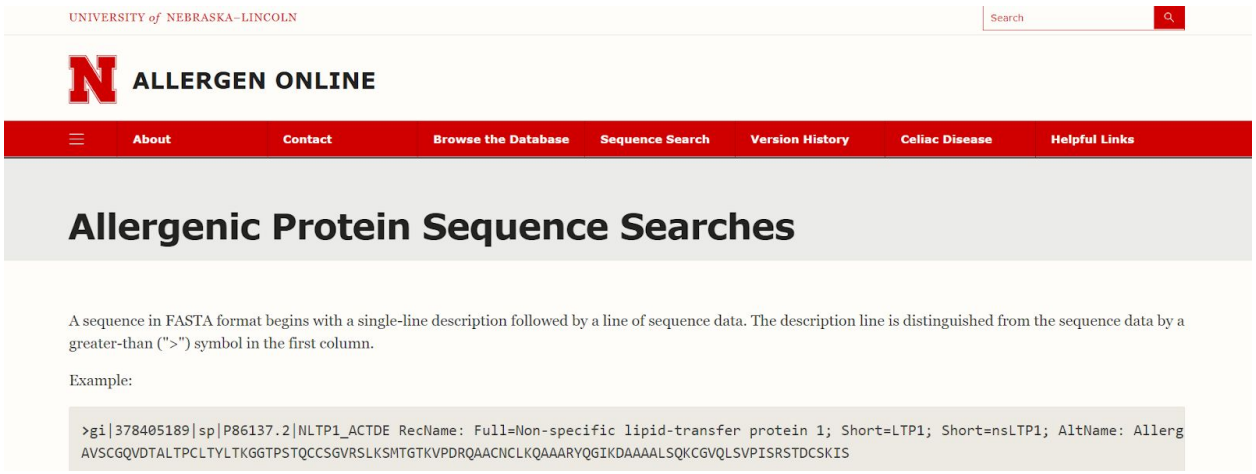


Figure 16: Data entry of vaccine construct done in Allergen Online server.

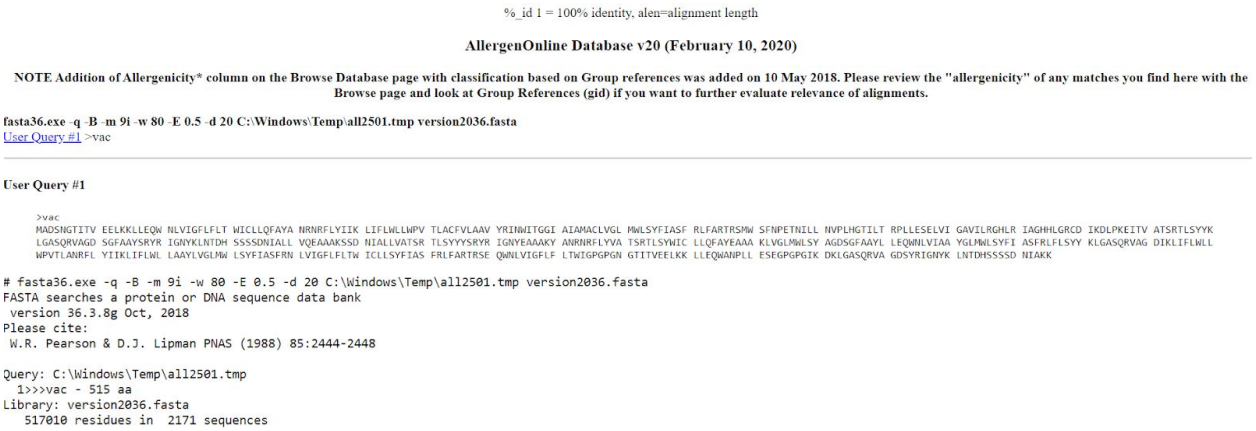


Figure 17: Output obtained corresponding to our vaccine construct from Allergen Online server.

4.10:Homology modelling of vaccine:

Up till now, the vaccine had a primary structure and no physical existence.For further analysis,it's essential that we obtain a 3D structure of our vaccine.This structure is obtained via *in silico* methods using a PDB file of our vaccine.This PDB file is developed using the technique of homology modelling based on past PDB entries used as a template.27 residues (5% of the sequence was modelled with 62.2% confidence using the highest scoring template.The 3D image of the modelled vaccine is given below:

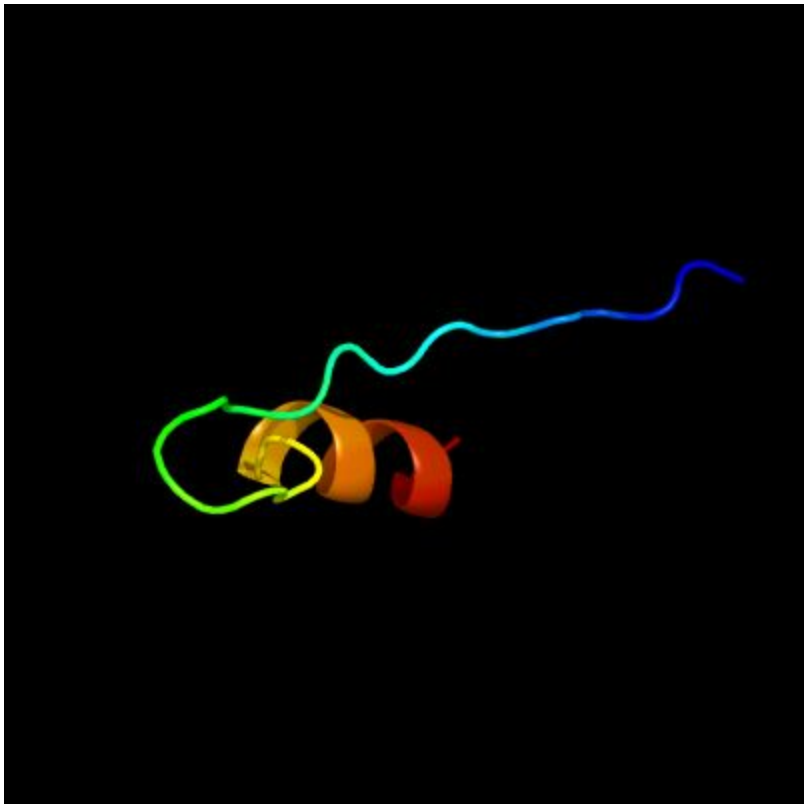


Figure 18:3D model of vaccine developed using Phyre2 server.

In Å units,the dimensions of the model along X,Y & Z axes are 23.78, 33.364 & 26.8 respectively.The highest scoring template is the unliganded α-L-fructosidase from *Lactobacillus casei.*

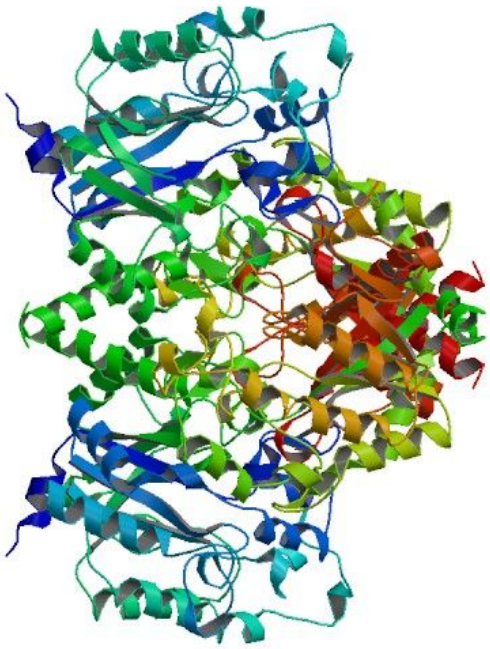


Figure 19: Unliganded α-L-fructosidase from *Lactobacillus casei*,the highest scoring template used to model the vaccine(PDB ID:6O18).

The obtained data from the phyre2 server is briefly tabulated below:

PDB ID of template used.	**Confidence.(%)**	**Identity coverage(Similarity to template).(%)**
6O18	62.2	26
4NI3	58	30
2WVS	56.6	22
3SC0	54.3	24
6SCL	53.8	25
2MMU	53.3	42

Table 10:PDB entries used as template in homology modelling of vaccine.

As seen from table 9, the first entry tabulated is the best model to be used in terms of confidence and coverage.

4.11:Further Analysis of obtained Homologous vaccine model:

The PDB structure of our vaccine obtained from phyre 2 server is further analysed.We had conducted the Ramachandran plot analysis using the RAMPAGE server and a Z-score versus residue analysis curve was constructed using the PROSA server.The RAMPAGE server interpretes the Ramachandrana plot whereas the actual Ramachandran plot is plotted using SAVES version 5.0 server as follows:

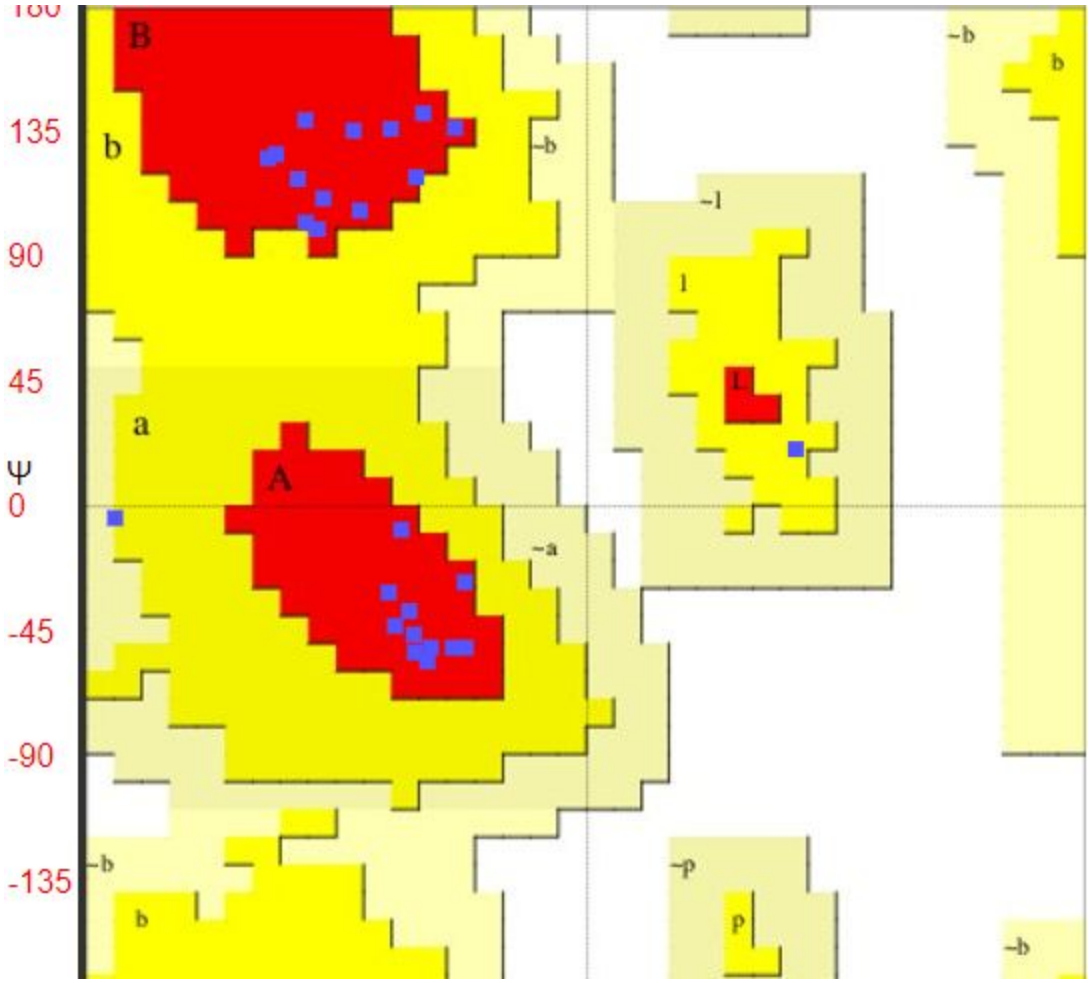


Figure 20: Ramchandran plot done using Saves version 5.0 server.

Interpretation of the plot is given below in table 11.

Position of Residue.	Number of residues (in %)
Outlier Region.	1 residue at position 450 (Asn) at (-172.40,-1.83) co-ordinate.(3.8%)
Favoured region.	25 (96.2%) with 98% expectancy.
Allowed region.	0(2% expectancy.)

Table 11: Region-wise interpretation of the Ramachandran plot.

The z-score versus residue was conducted and was found to be -0.94.Since the value is below 0, we can interpret that the raw score is below average.The plot is given in figure 21 below:

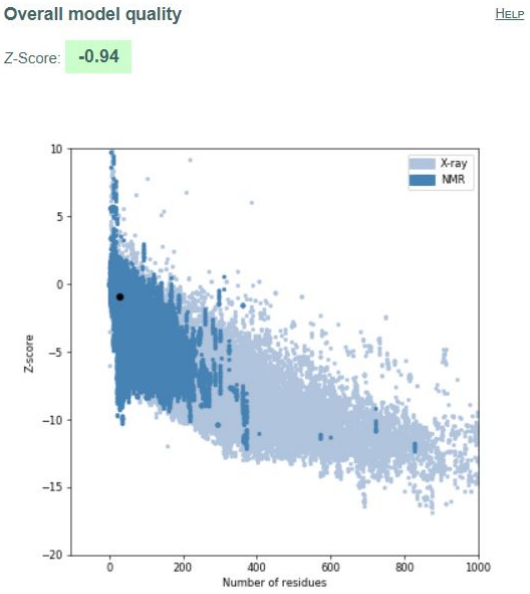


Figure 21:Z-score versus residue plot.

4.12:Result of docking Analysis:

Docking analysis was carried out using the Patchdock server between the TLR8 member of toll-like receptor(TLR) family.This family consists of proteinous receptors that are key to promotion of innate immune response.Constitutively being a single-pass-membrane spanning receptor by structure they are usually found to be expressed in sentinel cells that are essentially responsible for destruction of invasive pathogens within the body.The TLRs interact with epitopes of the mo antigens and stimulate specific immune response against particular ag.TLRs

are distinguished by numbers.They are numbered 1 to 13.Among them TLR 11,12 & 13 are not present in humans.For our experiment we had used the unliganded form of TLR8 (PDB ID:3W3G) as the receptor while the PDB file obtained from phyre2 server was used as the ligand.Docking was carried out using the Patchdock server with clustering RMSD at 4.0 Å .
Patchdock server performs docking and scores the obtained docked complexes at specific transformations.Furthermore,it calculates atomic contact energy along with the area of docking in square angstrom units.Top 10 solutions were initially monitored & are tabulated as follows:

Score	**Area(In square angstrom)**	**Atomic Contact Energy(ACE)**	**Transformation**
15024	2040.50	-44.27	-1.73,0.97,-3.14,14.29,-4.5,30.52
13902	2008.70	158.25	-2.49,0.01,-1.3.14.3,-15.78.54.98
13858	2089.30	-464.68	1.32,1.25,-1.30,29.48,-3.39,40.76
13742	1825.30	-330.42	1.21,1.24,-1.6,32.42,-13.06,37.49,
13710	1763.00	70.95	-.1,0.23,1.63,7.59,24.38,14.21
13650	1864.7	-353.58	-1.61,-1.21,2.05,22.69,39.99,30.28
13648	1767.00	-134.93	-1.00,0.86,-0.87,25.41,-22.20,28.46
13604	2111.90	16.45	-0.94,0.14,2.45,-2.63,54.14,38.27
13596	1618.40	-72.06	1.26,-.98,2.84,13.68,-17.15,26.08
13356	19994.00	-252.92	0.19,-.13,-1.79,19.28,-7.52,29.14

Table 12: Table of docking score with corresponding ACE at specific transformation with area.

From the results obtained,it's evident that among 10 different transformations,at the transformation of **(-1.73,0.97,-3.14,14.29,-4.5,30.52)** we obtained the highest score of 15024 & lowest ACE value of -44.27 $KJmol^{-1}$ covering an area of 2040.5 square angstrom is presumably the best complex obtained between the TLR8 receptor and our designed vaccine model.A PDB structure of the obtained protein-ligand complex viewed using Discovery Studio 2016 64-bit client version.The docked complex will be useful for future research particularly in molecular dynamics *(MD)* simulation studies,which is not under the scope of this work.

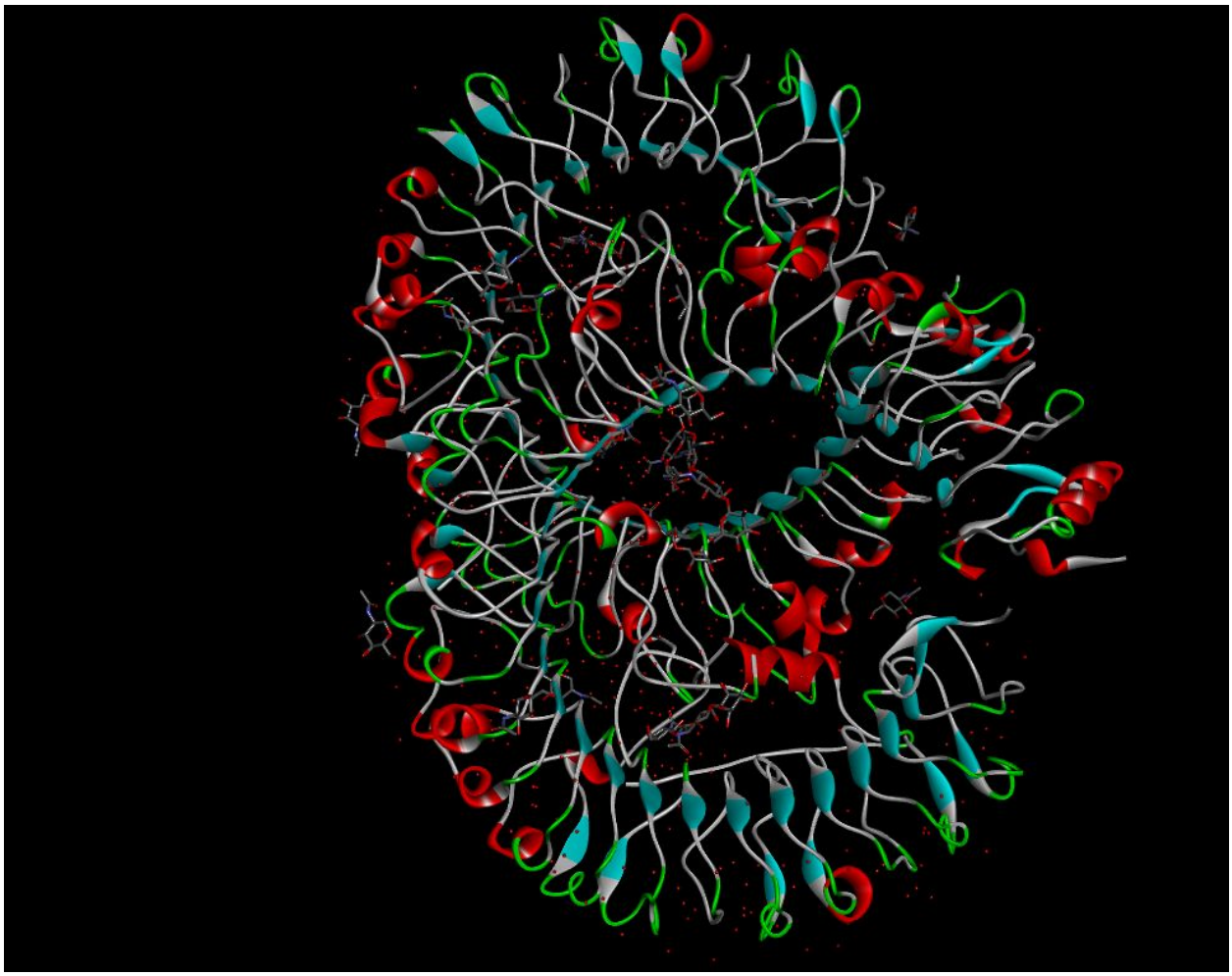


Figure 22:3D structure of docked complex between modelled vaccine & TLR8 receptor.

4.13:*In silico* Immunosimulation Results:

In silico simulation of the designed vaccine is a very important step particularly in terms of quality evaluation and efficacy of vaccine.Immunosimulation was conducted using the c-immsim server using FASTA sequence of the vaccine as the input.The output obtained included various

graphs that depicted the change in concentration of antigen,immunoglobulins as well as immune-cellular epitopes with time.The graphs obtained as well as their interpretations are given below:

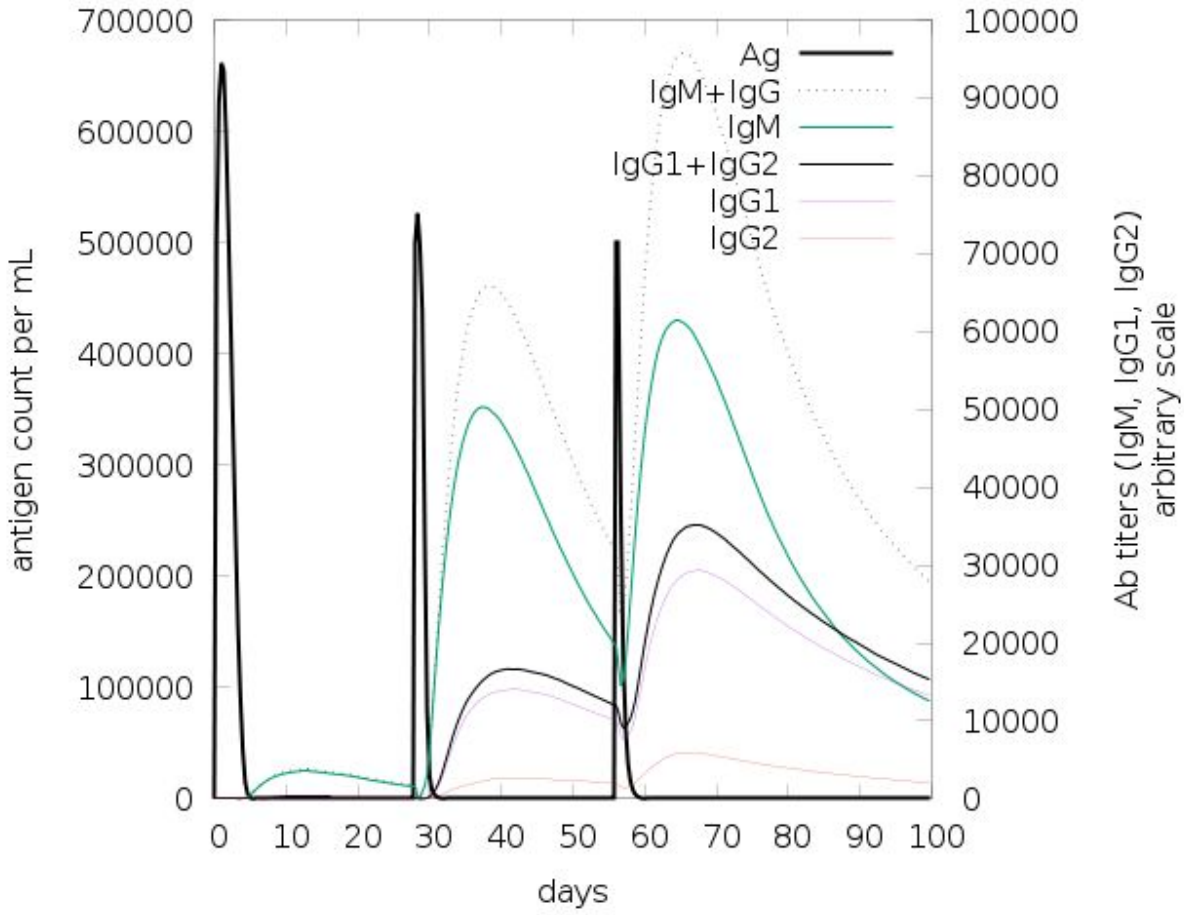


Figure 23: An ag concentration versus days graph also depicting the ab titers (by arbitrary scale).

The C-immsim server also calculates the concentration of B cell lymphocytes upon vaccine administration.B cell epitopes play important roles in exhibition of both humoral & cell mediated immunity.

Furthermore, memory B cells are essential for boosting ag specific immunity that can help protect the host against future attacks by the pathogen.Concentrations of B cells are measured by simultaneously measuring igM and both igG1 & igG2 concentrations respectively.As well as the aforementioned graphs, graphs depicting state-wise B cell population concentrations(ie. active,internalized,MHC II presenting,ag duplicating or even anergic).Lastly,plasma B cells

were also measured by concentration.These cells are potentially important as therapeutic agents.Plasma therapy specific to blood groups are a potential scope of treatment being run on a trial basis worldwide with promising results, despite the probability of causality of secondary infections from the donor.The graphs showing B cell status are given below:

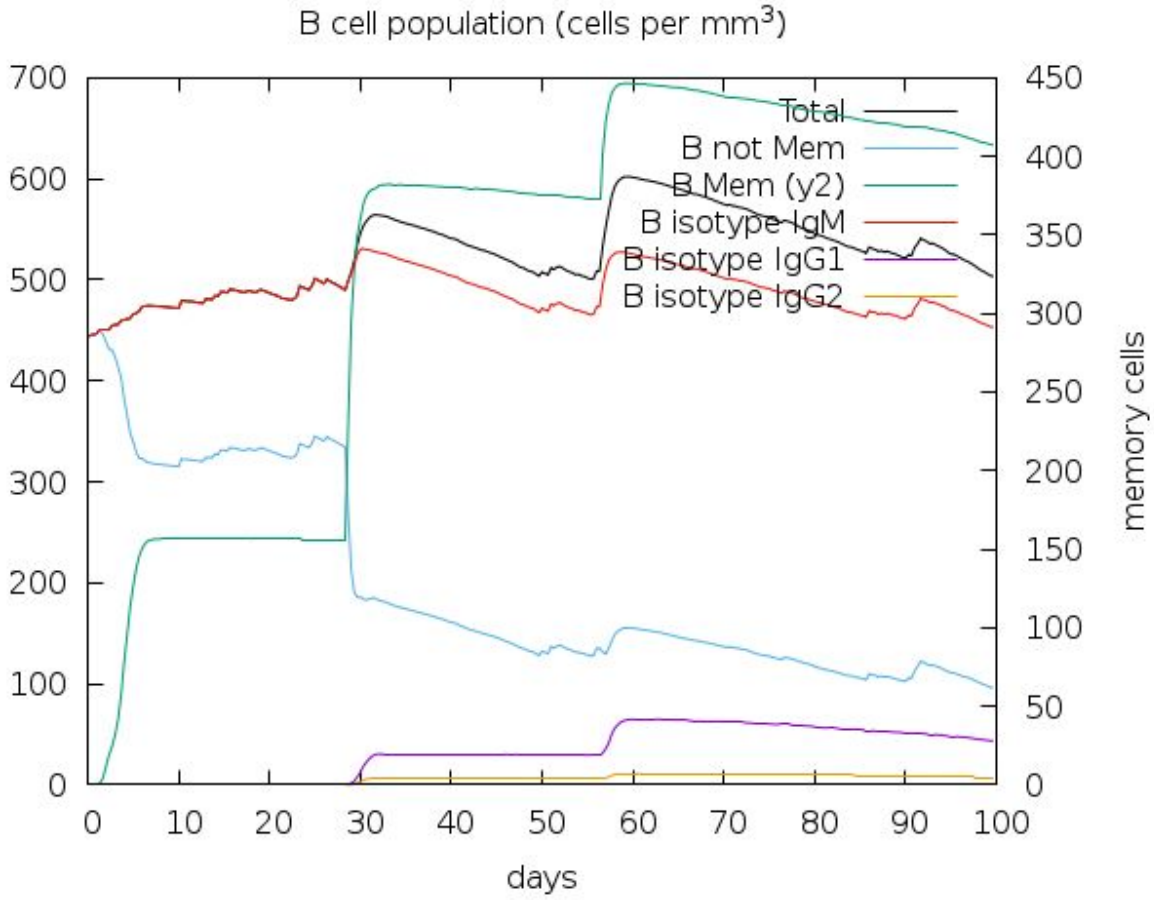


Figure 24: Graph depicting the concentration of B cells based on subtypes versus days after vaccine administration.

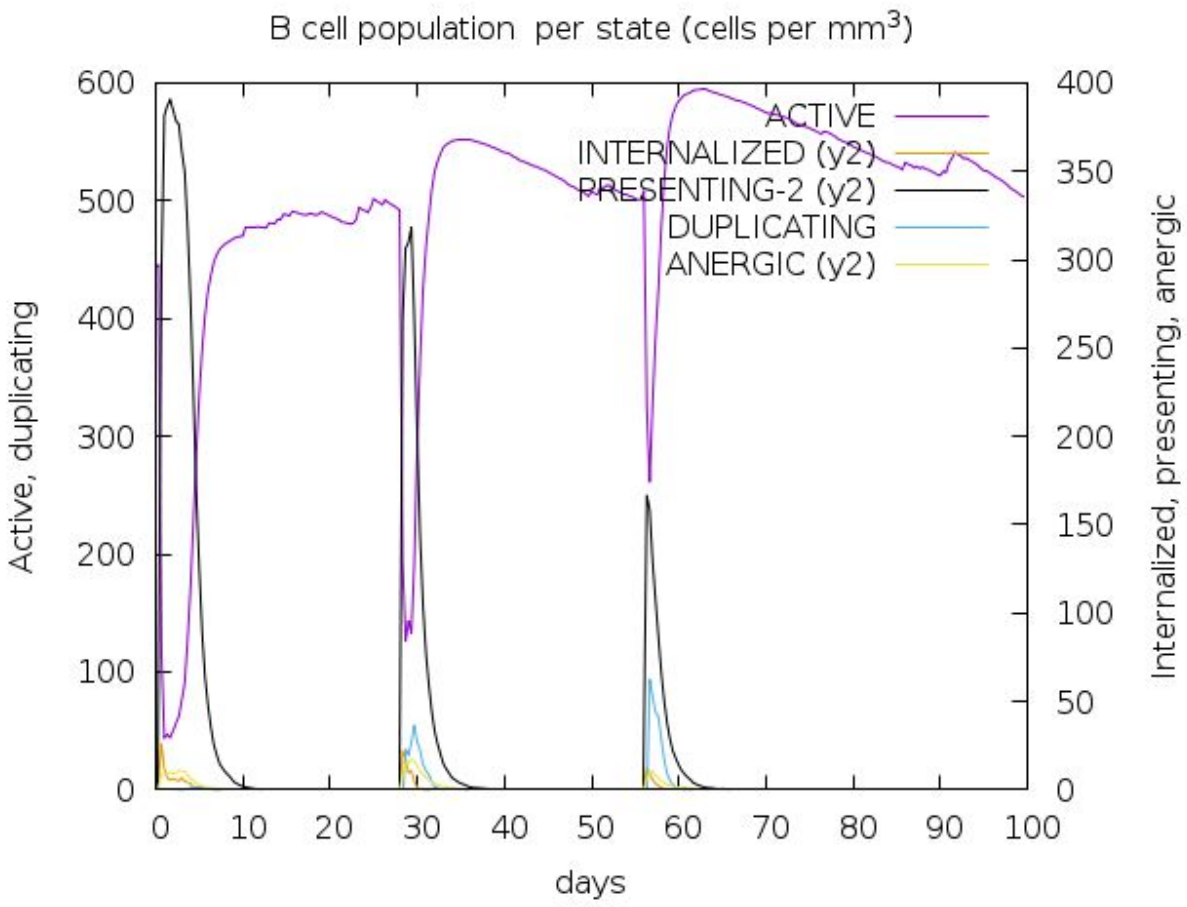


Figure 25: Graph depicting state wise concentration of B cells based on state versus days passed after vaccine administration.

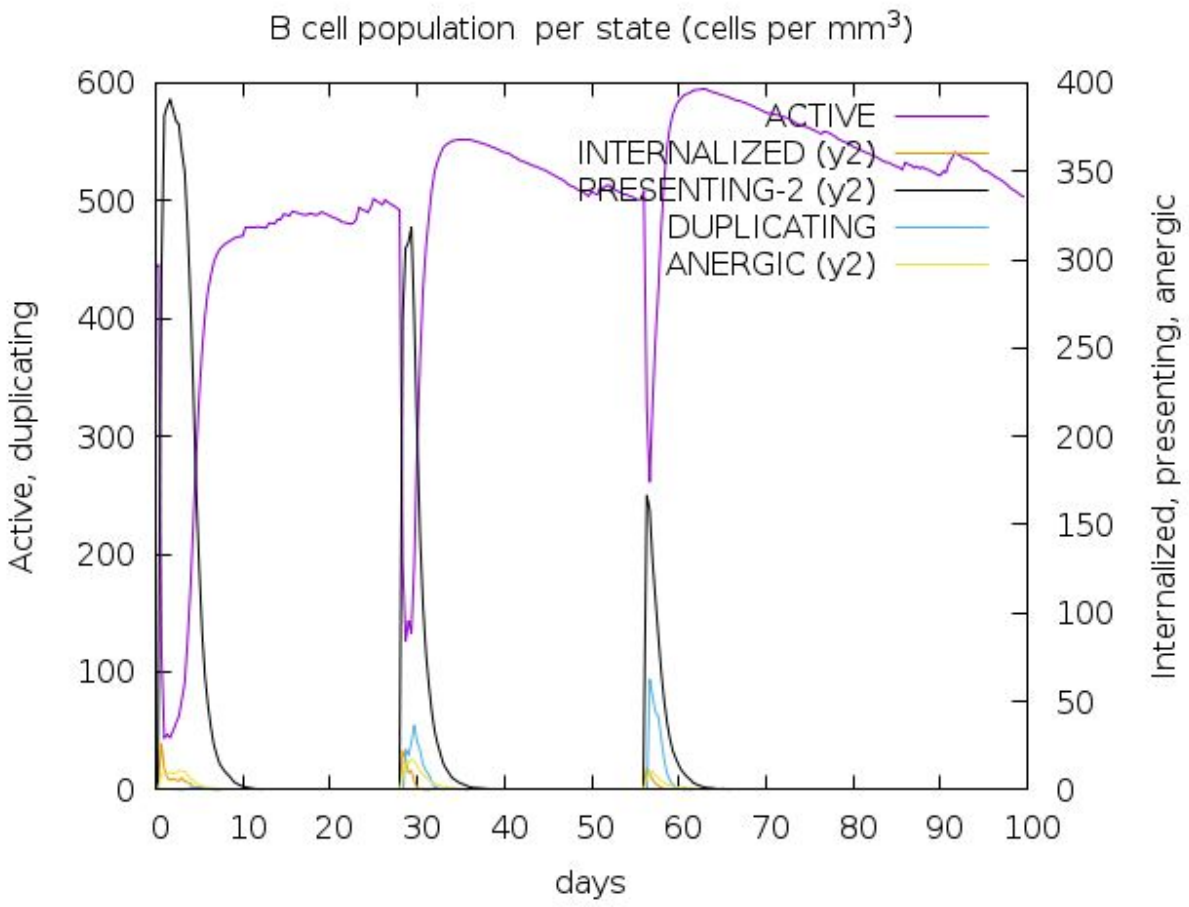


Figure 26:B cell population concentration versus days of vaccine administration with respective states.

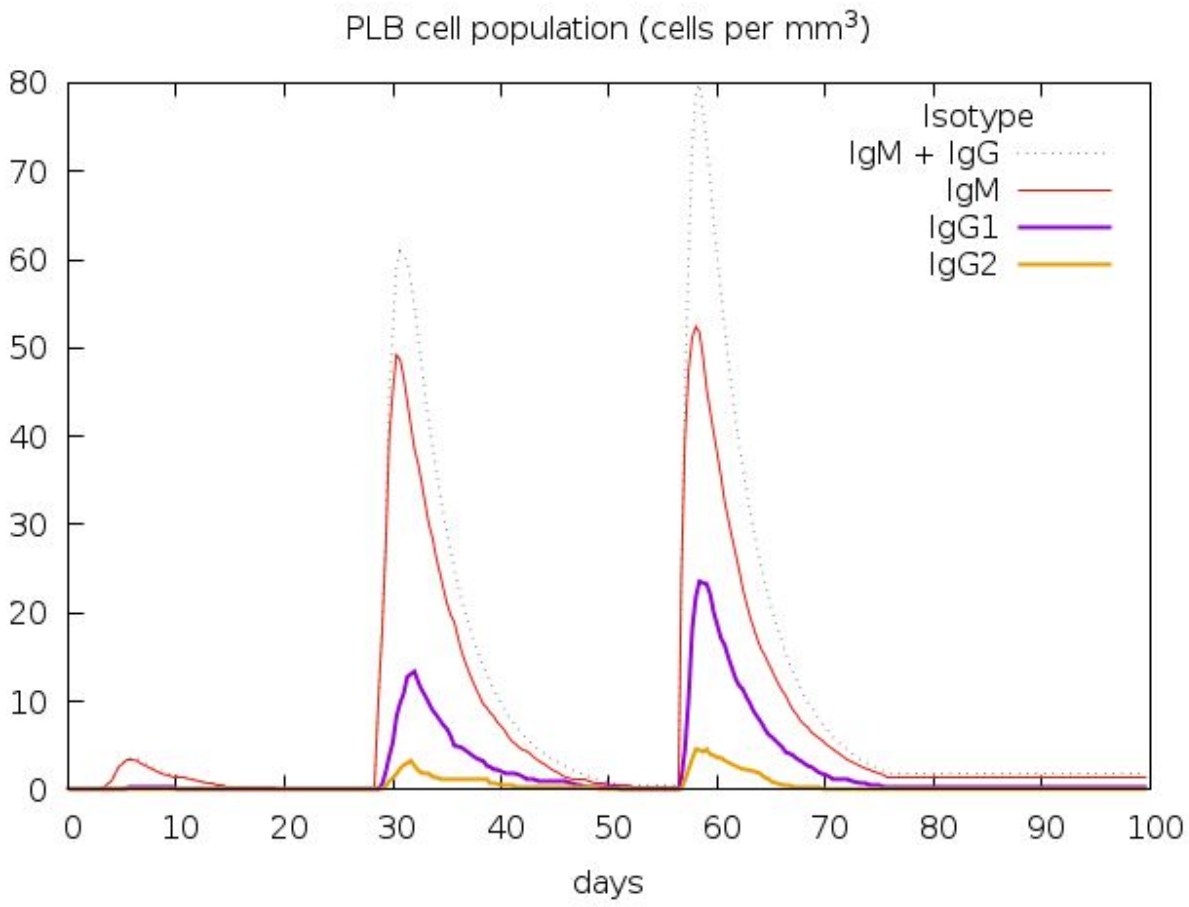


Figure 27: Plasma B cell population growth (by isotype) versus days post vaccine administration.

Similar to the B cell concentration graphs,graphs depicting HTL & CTL epitope concentrations were also obtained from the server on a day to day basis.Graphs depicting CTL and HTL concentration were also plotted by the server.

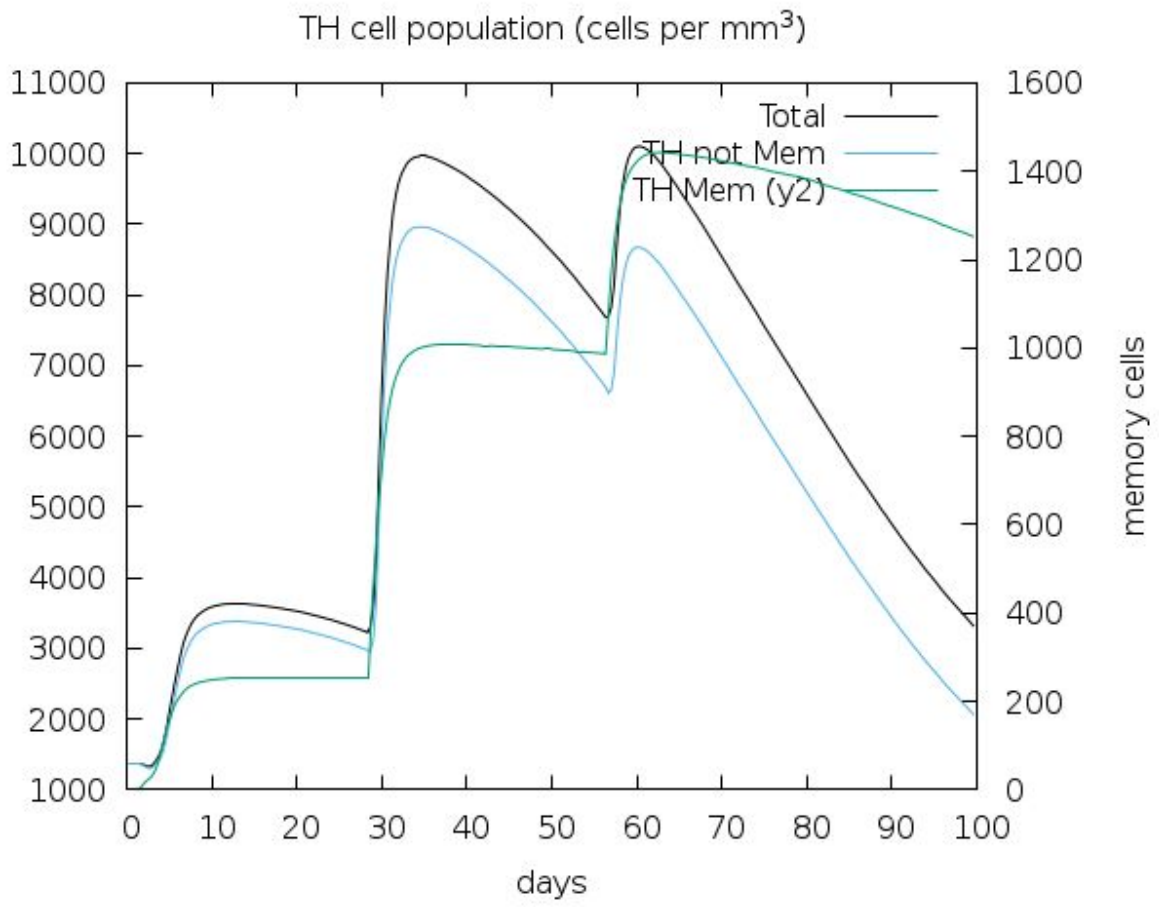


Figure 28: CD-4^{+} HTL epitopes counted inclusive and non-inclusive of memory and normal T cells observed by days post vaccine administration.

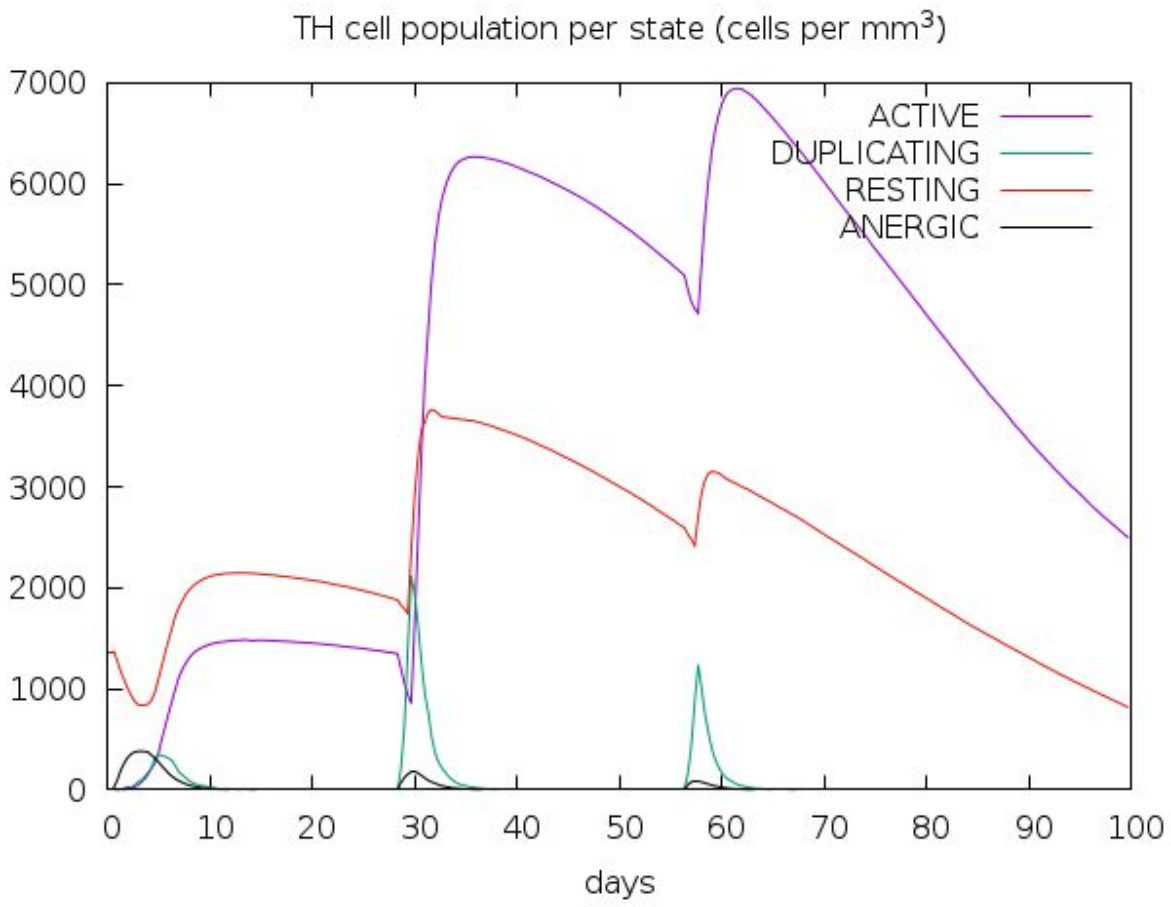


Figure 29:CD4^{+} HTL epitope concentration measured and plotted by state on a daily basis post vaccine administration.

Lastly among the CD-4^{+} T-cell population,a graph similar to Figure-27 was plotted by state for regulatory T-cell lymphocyte population as shown in Figure-28 as below:

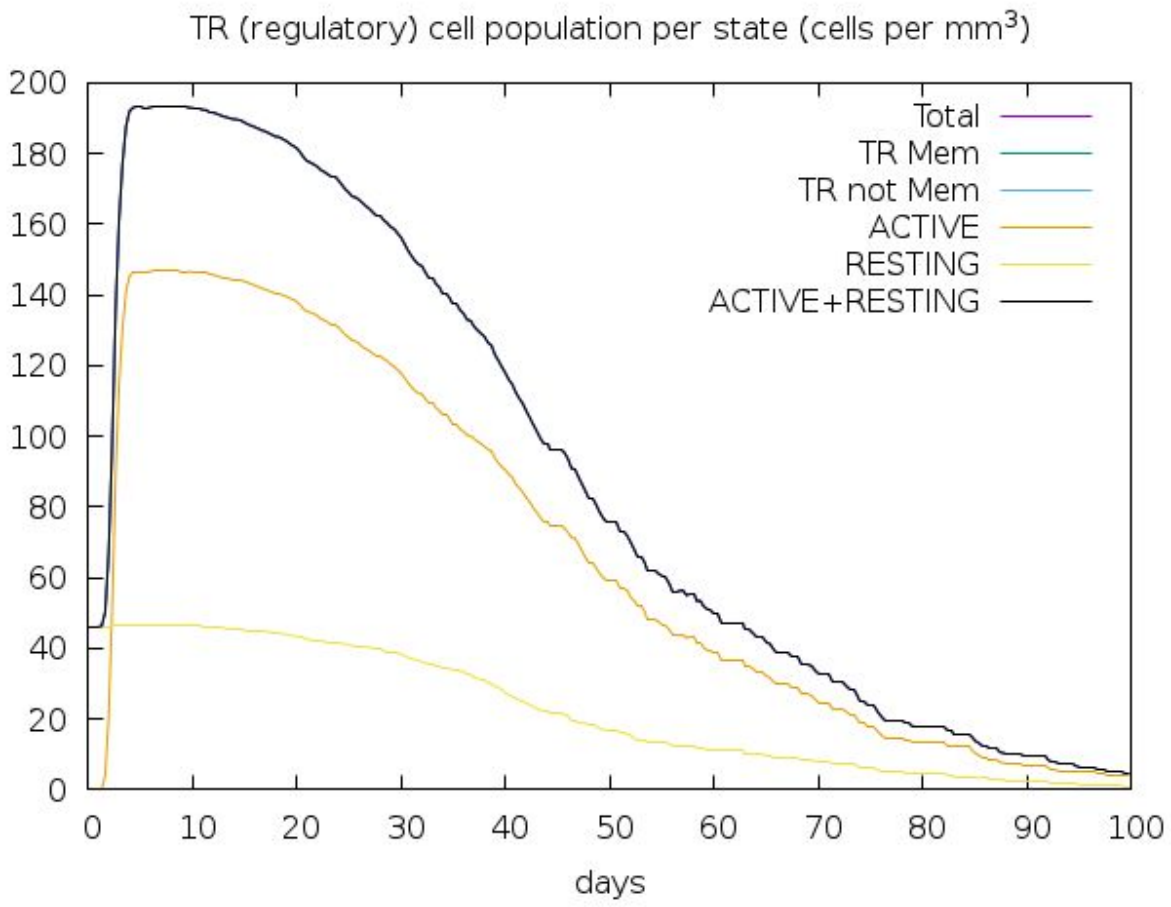


Figure 30: Regulatory T-cell lymphocyte population (CD-4^{+}) concentration versus days post vaccination.(States of Regulatory T cells shown.)

Now,let us focus on the growth of potential CTL epitopes upon vaccination.CD-8^{+} CTL epitope concentration of both memory inducing and non-memory inducing CTL concentration was observed.Also,states of CTL epitope along with concentration post vaccination was plotted as follows:

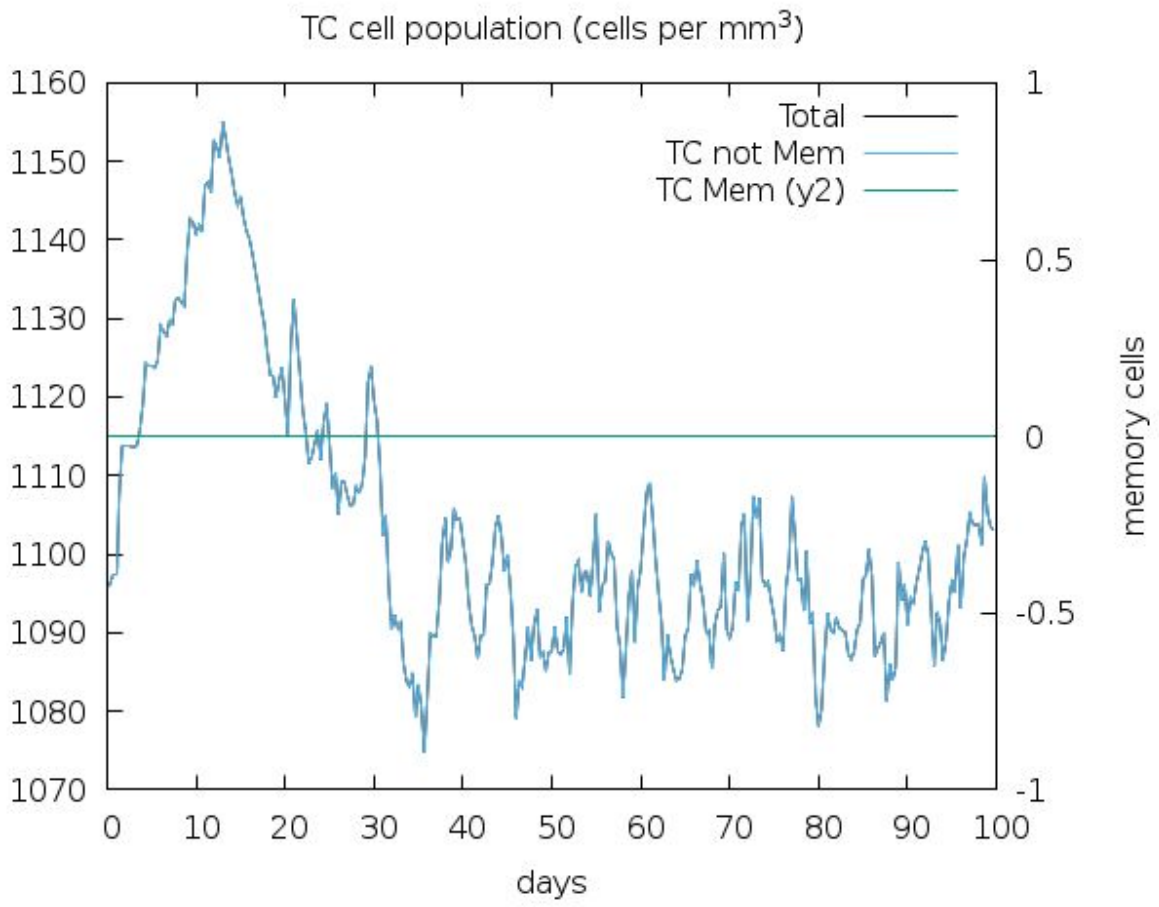


Figure 31: Growth of memory,non-memory and total CTL concentration growth versus days post vaccination.

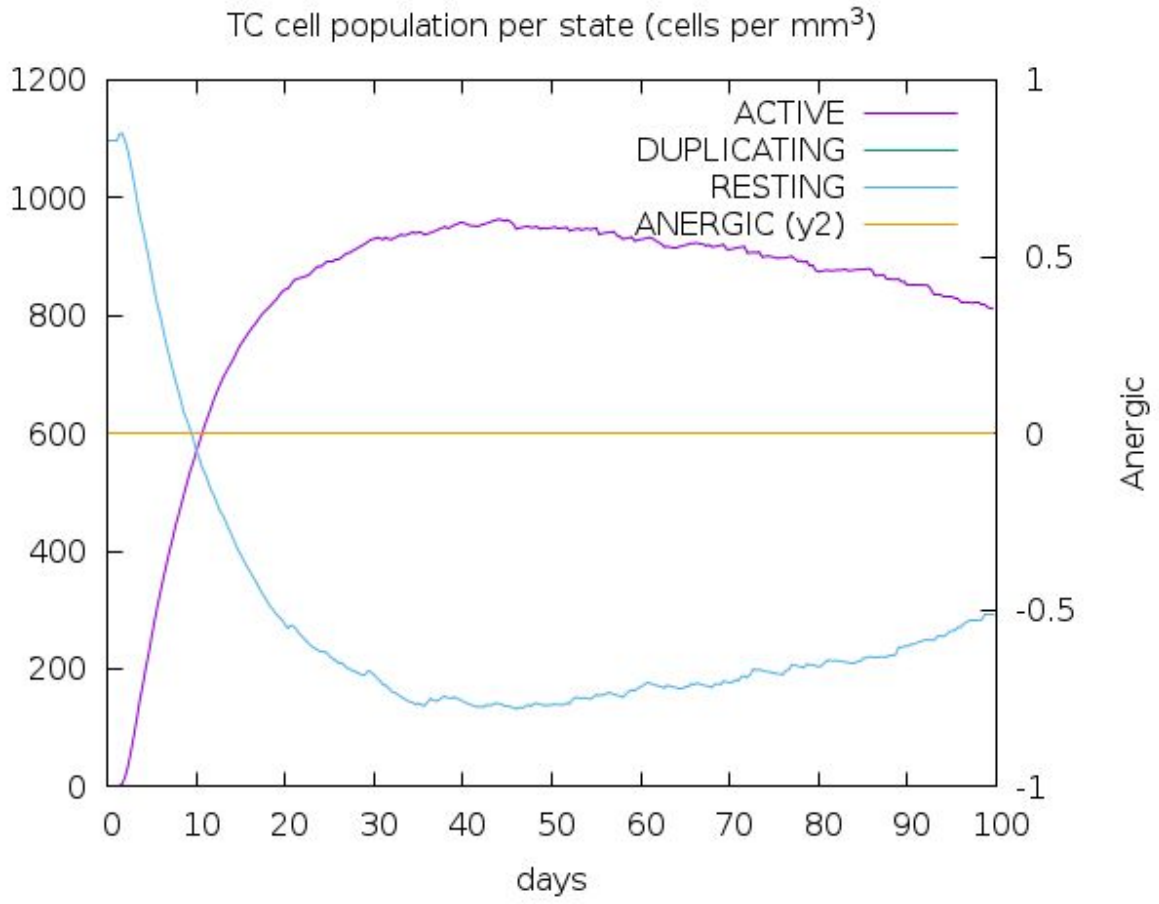


Figure 32:CTL growth observation by subtype versus days of vaccine administration.

Lastly, the C-Immsim server has also monitored the probable increment of WBCs in the host's system post vaccination.Cells included NK cells,DC population & macrophages as well as epithelial cells.The graphs are as follows:

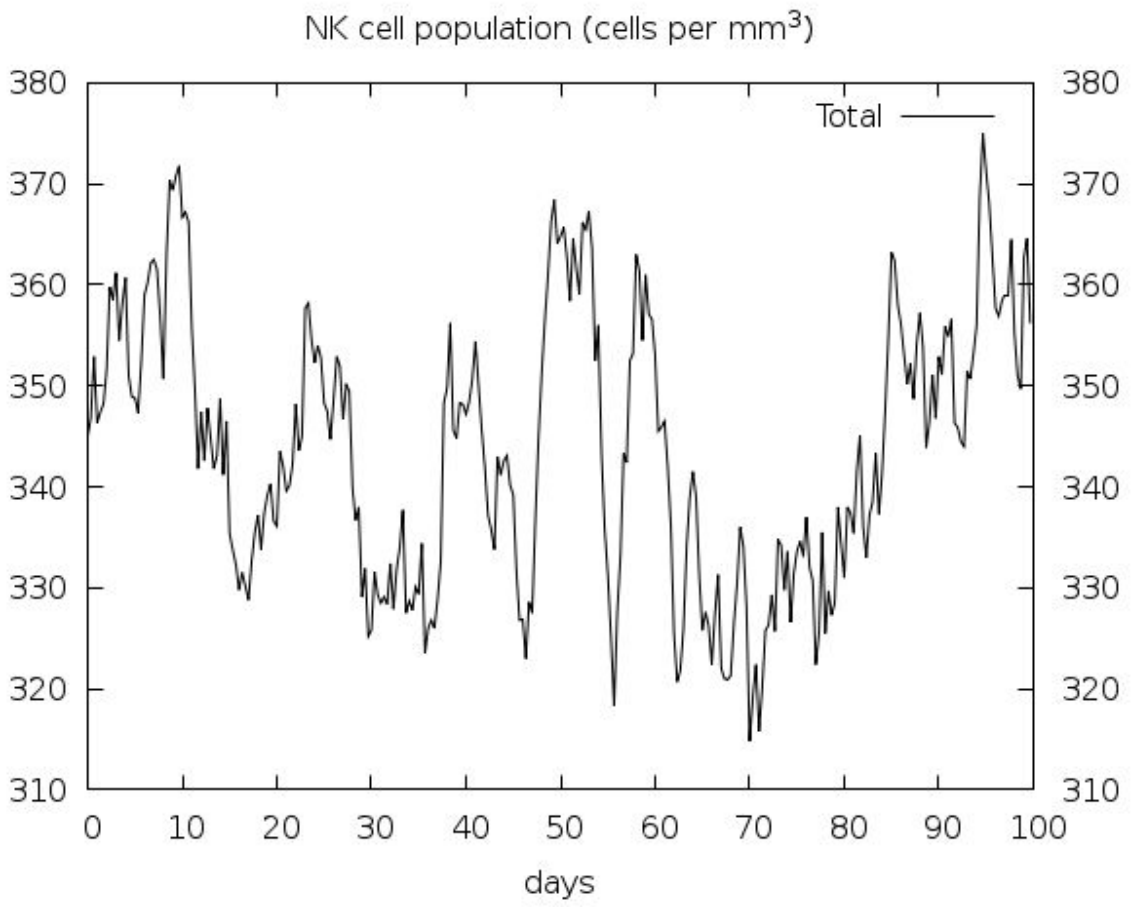


Figure 33: Total NK cell population growth observed on day to day basis post vaccination.

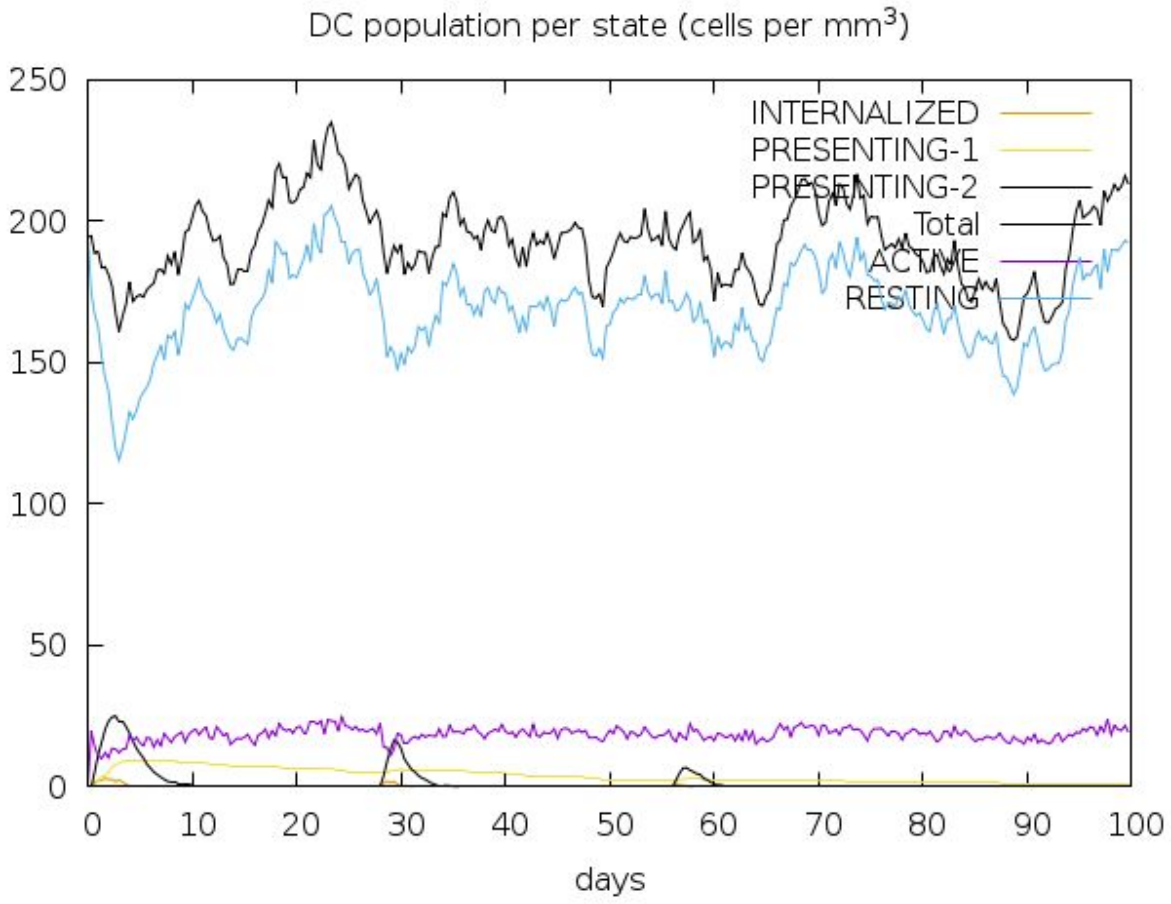


Figure 34:DC population can present ag presented both by MHC I & MHC II molecules.Cells were plotted further based on active,resting & internalized states respectively.

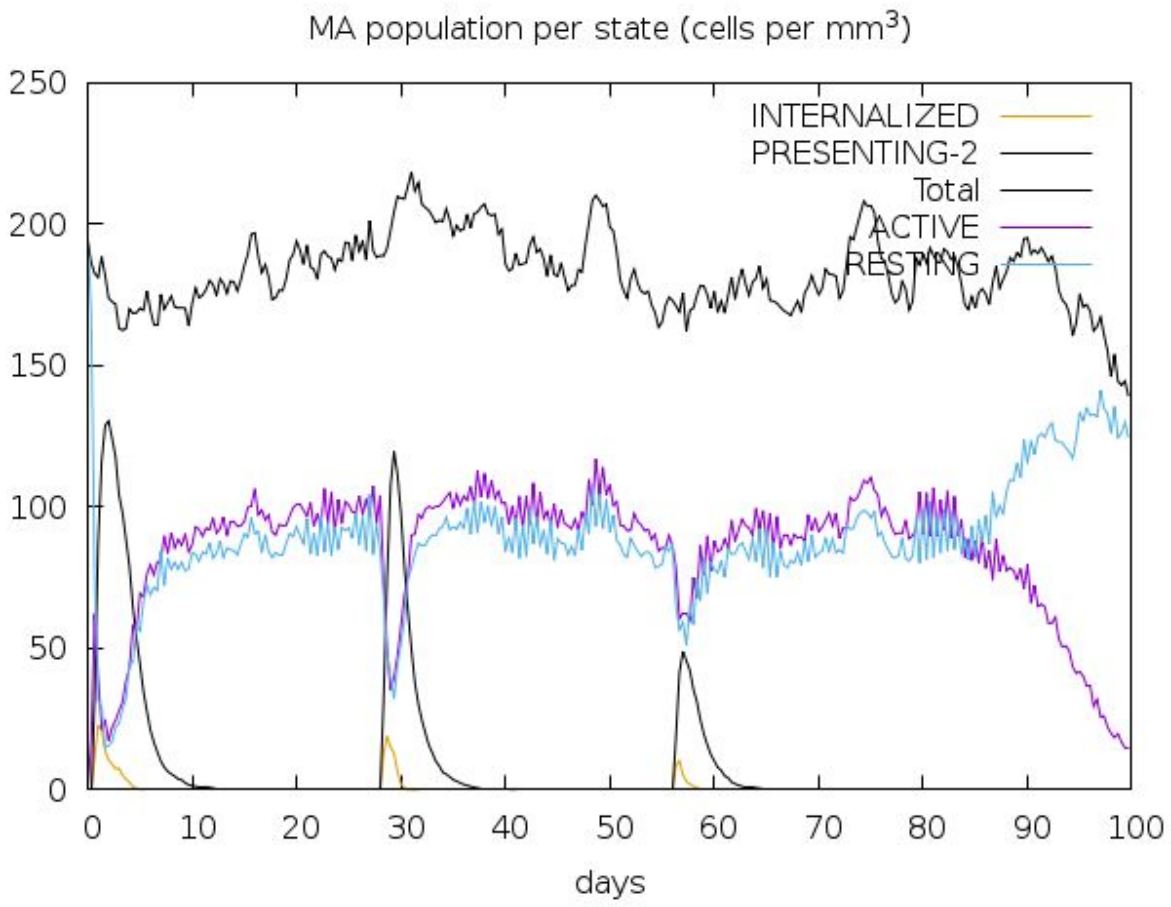


Figure 35: Macrophage population growth observed by classification into active, resting,MHC II presenting and internalized groups along with total growth versus days of vaccine administration.

Lastly, graphs were plotted to observe epithelial cell growth and cytokine growth post vaccination.Epithelial cells are suitable sites of viral infection spread and cytokines include interferon and interleukin populations along with other agents that promote inflammation within the host.Promotion of inflammatory response help boost immunological tolerance.Graphs obtained are given below:

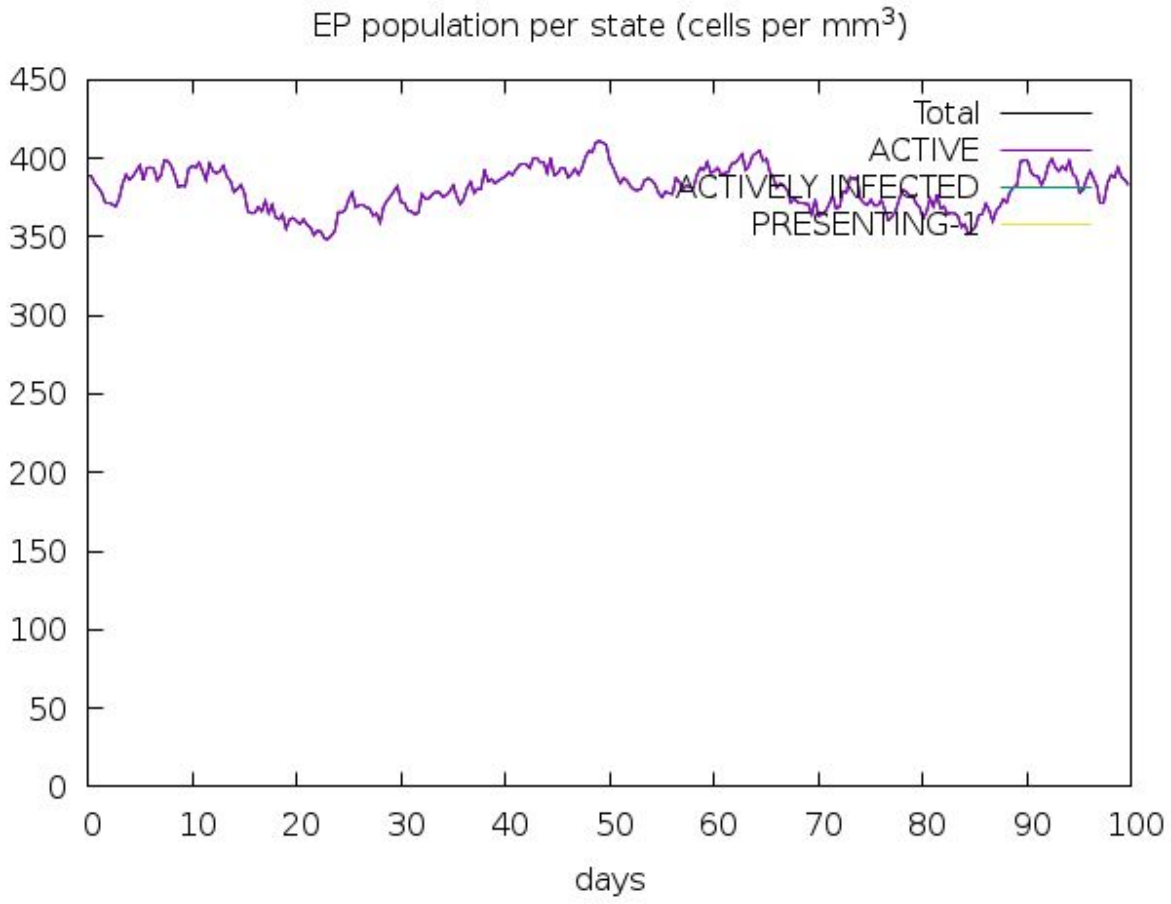


Figure 36: Concentration of total,active,MHC I presented and virally infected cells versus days post vaccination.

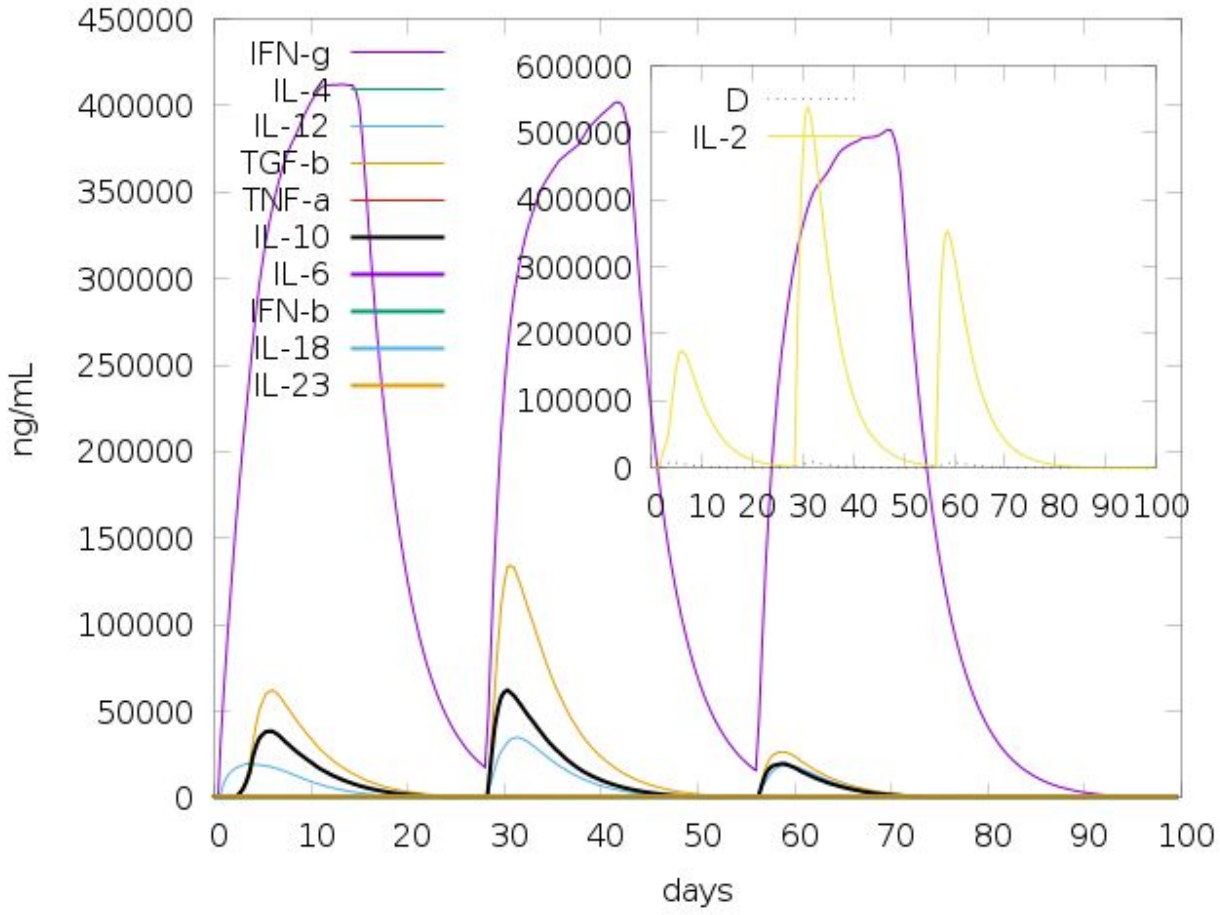


Figure 37: Cytokine population growth versus with danger indicated in the inset.

4.14:Interpretation of *in silico* simulation results:

In this section we will attempt to interpret the results obtained by the C-Immsim server.*In silico* simulation analysis is a fundamental step to immunoiformatic vaccinomics,particularly since it provides a feasibly validational basis to the conducted research.From the graphs showed previously, it's evident that the vaccine promoted enhanced ab-mediated immunity within the host in the form of igM,igG within few days of vaccine administration.The server considers the vaccine as a typical ag.The host immunity is triggered by antibodies generated by immunostimulation caused by the vaccine.As seen in Figure(), vaccine administration to the host initially results in a spike in concentration of SARS-CoV-2 specific ag, in this case, the M protein along with the potential epitopes used in its development.Gradually with time the ag concentration is reduced due to development of ab specific to SARS CoV-2, which is the

primary objective of vaccine development.Ab particular to a specific ag is developed in a host via a potential trade-off with ag. Ag is neutralized by ab upon every administration of vaccine.As mentioned earlier the vaccine was developed on a tri-dosage system.Thus, upon administration of every dose an antigenic spike is initially observed in the host that will be gradually reduced gradually by ab developed against the SARS CoV-2 M protein.

Again,the C-Immsim server also plotted B cell concentration upon vaccine delivery;the plots() clearly showed a visible increment in B cell populations;particularly among memory B cells,antibodies (in the form of immunoglobulins such as igA,igG and igM to name a few) as well as plasma B cells.Memory B cells specific to a particular ag are capable of host protection against it upon future encounters.Plasma B cells specific to the particular blood group of the individual is now considered a potential therapeutic agent against the virus.Thus, from the graphs it was evident that the vaccine was particularly effective in B-cell mediated host immunity.

Furthermore,plots obtained in figures() also implied enhanced growth of $CD4^+$ HTL & regulatory T cell lymphocytic epitopes that are key to the promotion of immunological tolerance in the host.This concentration is lowered with passage of time but doesn't potentially reduce the immunity level.A similar pattern was observed among $CD8^+$ cell populations,particularly among the CTL epitopes.

Lastly,graphs representing cellular concentrations of WBCs of varied types showed highly effective activity upon vaccine administration.I.e. To say,DC cells,macrophages etc besides showing notable reduction in number of viral infections.

Lastly,cytokine growth was also seen to increase.The plot also highlighted critical levels of cytokine growth that needs to be monitored since excessive inflammation can be potentially harmful and may result in internal organ failure.However, the cytokine curve has also showed growing populations of interleukins & interfererons,thus rendering our vaccine effective against SARS CoV-2.

Chapter 5

Discussion & Future Prospects Of Conducted Research

Here we have made an attempt to develop a subunit vaccine against the SARS CoV-2 virus targeting its M protein(i.e.its membrane glycoprotein as the primary antigen in our construct).SARS CoV-2 is a retrovirus belonging to the *Cooronaviridae* family.Declared a pandemic by WHO as mentioned earlier, no candidate has yet been discovered that in turn has proven to be a dominant candidate to cure this disease.As such, both physicians,pharmacists & biotechnologists tend to prefer a combination of medications to encounter this notorious followed by optimization of dosage based on physiological conditions.Currently, the approach being made is majorly based on the procedure of trial and error.

At this very stage of the pandemic where the SARS CoV-2 virus is showing a constantly exponential growth rate in terms of virulence exhibit and host attack,a preventive approach is the best path to follow in this regard.Wearing a mask particularly of the N-95 category,maintaining optimal social & physical distancing during socializing as well are notable 'preventive' approaches that have been proven globally to prevent SARS CoV-2 infection greatly.

Good hygiene practices(GHP) including usage of soaps,hand-sanitizers,bleach or H_2O_2(Hydrogen Peroxide) are highly recommended as means of killing the virus in study,although their mode of action does vary significantly.

In silico methods are a very effective and efficient form of approach in the field of vaccinomics with potential advantages over traditional methods.For example, this approach can actually give a feasible idea of a well-developed vaccine as well as predict its efficacy and infer potential side-effects of the end-product even before going in to human trials or *in vitro* experiments.But that being said, we don't claim that *in silico* approaches alone are sufficient enough in vaccine development.Rather,it is a means of helping develop efficient vaccines alongside the existing methodologies.

Furthermore, we humbly state that the subunit vaccine developed by us is a credible attempt to discover means to combat the SARS CoV-2 from a preventive perspective.That is to be said, although this vaccine has not been physically developed, we primarily recommend this vaccine to be administered to healthy individuals and if necessary to patients exhibiting mild degree of

symptoms.We don't however,prefer the vaccine to be given to patients affected by COVID-19 at an acute or chronic level.This is because such patients at that point are already exhibit an unusually high degree of immune-response within their system.Vaccination then will boost the immune response and thus worsen the physiological condition of the patient further.

Laborious efforts are being constantly made to get a potential vaccine candidate against SARS CoV-2 developed.Countrie,nations,governments and private sectors are all acting actively in this regard.It's worth mentioning that UK is potentially the frontrunner to be the first country to develop and commercially mass produce a feasibly efficient vaccine against SARS CoV-2.Oxford University UK , is the one responsible behind this research work.Their developed vaccine candidate named **ChAdOx-1 nCoV-19** tends to exhibit a 90% curative rate in terms of treating the disease.The vaccine showed effective immune response and subjects who underwent the trial initially showed symptoms of mild fever,uneasiness and slight breathing difficulties but were shortly cured after being administered with paracetamol.

Even Globe Biotech,Bangladesh is also trying to develop a vaccine in this context to help our countrymen get a vaccine against the disease that is easily accessible to everyone.WHO decrees that in case of a global pandemic,the vaccine is likely to be freely distributed worldwide,particularly among third world nations.

Despite these facts,though they are highly appreciable, we have tried to make a minor attempt to develop a subunit vaccine against SARS CoV-2 using computational biology and *in silico* approaches.Although,our efforts are likely to be error prone despite our utmost laborious efforts,we believe that we have been quite successful in developing a potentially effective candidate capable of combating SARS CoV-2.

We thus believe that in these critical times when SARS CoV-2 is at a peak-pandemic level,a combined approach by both chemotherapeutic and biotherapeutic candidates is preferred at this point before a perfect candidate against SARS CoV-2 is developed.

We hence,conclusively remark that the vaccine we developed here is still at a primary stage.It requires further *in vitro*,*in vivo* analysis as well as clinical trials at multiple phases before it's deemed safe for usage.We believe that further studies continued based on our work will help develop an efficient vaccine against the COVID-19 disease.

References:

Kaur, A., & Saxena, S. K. (2020). COVID-19: An Ophthalmological Update. In *Coronavirus Disease 2019 (COVID-19)* (pp. 81-93). Springer, Singapore.

Woo, P. C., Lau, S. K., Huang, Y., & Yuen, K. Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine*, *234*(10), 1117-1127.

Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews Microbiology*, *17*(3), 181-192.

Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., ... & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. *Military Medical Research*, *7*(1), 1-10.

Zhang, H., Penninger, J. M., Li, Y., Zhong, N., & Slutsky, A. S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive care medicine*, 1-5.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Müller, M. A. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260-1263.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research*, *176*, 104742.

Matsuyama, S., & Taguchi, F. (2009). Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. *Journal of virology*, *83*(21), 11133-11141.

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, *15*(8), 502.

Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., ... & Yuen, K. Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences*, *102*(39), 14040-14045.

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., ... & Choe, H. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, *426*(6965), 450-454.

Lai, M. M., & Cavanagh, D. (1997). The molecular biology of coronaviruses. In *Advances in virus research* (Vol. 48, pp. 1-100). Academic Press.

Banerjee, A., Kulcsar, K., Misra, V., Frieman, M., & Mossman, K. (2019). Bats and coronaviruses. *Viruses*, *11*(1), 41.

Perlman, S. (2020). Another decade, another coronavirus.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, *579*(7798), 270-273.

Ji, W., Wang, W., Zhao, X., Zai, J., & Li, X. (2020). Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J. Med. Virol*.

Lam, T. T. Y., Shum, M. H. H., Zhu, H. C., Tong, Y. G., Ni, X. B., Liao, Y. S., ... & Leung, G. M. (2020). Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *BioRxiv*.

Gu, J., Han, B., & Wang, J. (2020). COVID-19: gastrointestinal manifestations and potential fecal–oral transmission. *Gastroenterology*, *158*(6), 1518-1519.

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., ... & Du, B. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine*, *382*(18), 1708-1720.

Jaimes, J. A., Millet, J. K., Stout, A. E., André, N. M., & Whittaker, G. R. (2020). A Tale of Two Viruses: The Distinct Spike Glycoproteins of Feline Coronaviruses. *Viruses*, *12*(1), 83.

Wan, Y., Graham, R., Baric, R. S., & Li, F. (2020). An analysis based on decade-long structural studies of SARS 3, JVI Manuscript Posted Online 29 January 2020. *J. Virol.* Accepted

Li, H., Liu, L., Zhang, D., Xu, J., Dai, H., Tang, N., ... & Cao, B. (2020). SARS-CoV-2 and viral sepsis: observations and hypotheses. *The Lancet*.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., ... & Hotchkiss, R. S. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*, *315*(8), 801-810.

Dienz, O., Rud, J. G., Eaton, S. M., Lanthier, P. A., Burg, E., Drew, A., ... & Rincon, M. (2012). Essential role of IL-6 in protection against H1N1 influenza virus by promoting neutrophil survival in the lung. *Mucosal immunology*, *5*(3), 258-266.

Russell, C. D., Millar, J. E., & Baillie, J. K. (2020). Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *The Lancet*, *395*(10223), 473-475.

Shang, L., Zhao, J., Hu, Y., Du, R., & Cao, B. (2020). On the use of corticosteroids for 2019-nCoV pneumonia. *Lancet (London, England)*, *395*(10225), 683.

Liu, J., Li, S., Liu, J., Liang, B., Wang, X., Wang, H., ... & Xiong, L. (2020). Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine*, 102763.

Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., ... & Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific reports*, *9*(1), 1-18.

Nain, Z., Abdulla, F., Rahman, M. M., Karim, M. M., Khan, M. S. A., Sayed, S. B., ... & Adhikari, U. K. (2019). Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis using immunoinformatic approaches. *Journal of Biomolecular Structure and Dynamics*, 1-18.

Weiskopf, D., Angelo, M. A., de Azeredo, E. L., Sidney, J., Greenbaum, J. A., Fernando, A. N., ... & Mattia, K. A. (2013). Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proceedings of the National Academy of Sciences*, *110*(22), E2046-E2053.

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., & Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics*, *63*(6), 325-335.

Chloroquine's use to treat COVID-19 is backed by US government, but many questions remain. (2020).Retrieved 22 May 2020, from https://cen.acs.org/pharmaceuticals/drug-development/Coronavirus-puts-spotlight-chloroquine-questions/98/i12

Oxford COVID-19 vaccine to begin phase II/III human trials | University of Oxford. (2020). Retrieved 14th June 2020, from http://www.ox.ac.uk/news/2020-05-22-oxford-covid-19-vaccine-begin-phase-iiiii-human-trials

Peele, K. A., Srihansa, T., Krupanidhi, S., Sai, A. V., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. *Journal of Biomolecular Structure & Dynamics*, 1.

(Abraham Peele, Srihansa, Krupanidhi, Vijaya Sai & Venkateswarulu, 2020)