In-Silico Based Vaccine Development: Targeting Nucleocapsid Protein of SARS-CoV-2.

By

Yesmine Akter Iva

ID: 17346029

A thesis submitted to the Department of Pharmacy in partial fulfillment of the requirements for the degree of
Bachelor of Pharmacy (Hons)

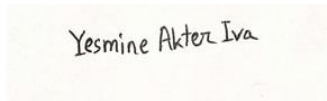Department of Pharmacy

Brac University

Spring 2021

Declaration

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:

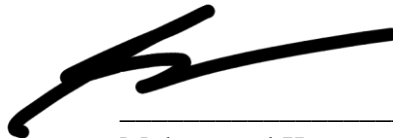*Yesmine Akter Iva*

_____

Yesmine Akter Iva
Student ID: 17346029

Approval

The thesis titled 'SARS COV-2 Vaccine design using nucleocapsid protein' submitted by Yesmine Akter Iva (ID: 17346029) of Spring 2021 has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Bachelor of Pharmacy (Hons) on [Date-of-Defense].

Examining Committee:

Supervisor:
(Member)

_____
Mohammad Kawsar Sharif Siam
Senior Lecturer, Department of Pharmacy
Brac University

Program Coordinator:
(Member)

_____
Hasina Yasmin, PhD
Professor and Deputy Chair, Department of Pharmacy
Brac University

U

Departmental Head:
(Chair)

_____
Eva Rahman Kabir, PhD
Professor and Chairperson, Department of Pharmacy,
Brac University

Ethics Statement

There were no unethical activities engaged in this thesis. No human or animal trials are used in this research. The thesis was conducted maintaining ethical standards at all regard whatsoever.

Abstract

SARS-CoV-2 is a highly contagious and virulent coronavirus that has triggered a pandemic of acute respiratory disease (COVID-19), endangering human health. In this study the goal was to develop a peptide based multi-epitope sub-unit vaccine against SARS-Cov-2 that would utilize the Nucleocapsid protein (N). Antigen was selected based on two primary criteria: (1) Protein antigenicity. (2) Protein Functionality. At first the structural Nucleocapsid protein (N) was selected through proteomic screening for analysis by using Vipr database. Then Vaxijen 2.0 was used to test the antigenicity of our proteins at a minimal threshold of 0.5. This study used immunoinformatics approaches to discover B cell and T cell epitopes for SARS-CoV-2 Nucleocapsid protein (N), then estimated their antigenicity and corresponding relationships with human leukocyte antigen (HLA) alleles. All of the identified epitopes were then linked to the core antigen using the appropriate linkers to develop the vaccine formulation. Antigenicity of this developed vaccine was 0.5135. 11 B cell epitopes, 9 MHC class-I and 2 MHC class-II strong binding of Tcells epitopes were identified. Analyses of allergenicity, toxicity, and physicochemical properties confirmed the epitopes' specificity and selectivity. A PDB model of the final vaccine was created utilizing phyre2 server via homology modeling. From the biochemical analysis of PROTPARAM this research found that the vaccine and crude N protein had an aliphatic index of 48.47 and 52.53 respectively, while the unstability indexes are at 60.89 and 55.81 accordingly. Again, the vaccine construct's and crude N protein's Grand Average of Hydropathicity (GRAVY) values are -1.110 and -0.980 respectively. Then from Patchdock the best complex score of 18094 in the molecular docking algorithm based on shape complementarity principles under the area of 2671.30 square angstrom in the Atomic Contact Energy (ACE) of 464.89 was selected. An in-silico simulation study was conducted utilizing the C-immsim server to further verify the credibility of this vaccine's effectiveness, and as the result was unstable so addition of chaperones can help to stabilize the vaccine.

Yesmine Akter Iva
Id: 17346029
Department of Pharmacy
Brac University

Table of Contents.

List of Tables.

List of Figures.

Figure 11	A score versus position graph for potential B-cell epitopes particular to the SARS CoV-2 N protein
Figure 12	A screenshot of entry of primary structure of vaccine to the Expasy server.
Figure 13	Output obtained from Expasy server based on vaccine construct.
Figure 14	Data entry of vaccine construct in Allergen Online server.
Figure 15	Output obtained corresponding to our vaccine construct from Allergen Online server.
Figure 16	3D model of vaccine developed using Phyre2 server.
Figure 17	Ramachandran plot done using SWISS PDB plotter.
Figure 18	MolProbity results of Ramachandran plotting.
Figure 19	Quality Estimation for Ramachandran plotting.
Figure 20	Residue Quality for Ramachandran plotting.
Figure 21	Z-score versus residue plot.
Figure 22	Local model quality for Knowledge-based energy versus Sequence position.

Figure 23	Molecular Docking Algorithm Based on Shape Complementarity Principles.
Figure 24	3D structure of docked complex between modelled vaccine & TLR8 receptor.
Figure 25	An ag concentration versus days graph also depicting the ab titers (by arbitrary scale).

Figure 26	Graph depicting the concentration of B cells based on subtypes versus days after vaccine administration.
Figure 27	Graph depicting state wise concentration of B cells based on state versus days passed after vaccine administration.
Figure 28	Plasma B cell population growth (by isotype) versus days post vaccine administration.
Figure 29	CD-4$^+$ HTL epitopes counted inclusive and non-inclusive of memory and normal T-cells observed by days post vaccine administration.
Figure 30	CD4$^+$ HTL epitope concentration measured and plotted by state on a daily basis post vaccine administration.
Figure 31	Regulatory T-cell lymphocyte population (CD-4$^+$) concentration versus days.
Figure 32	Growth of memory, non-memory and total CTL concentration growth versus days post vaccination.
Figure 33	CTL growth observation by subtype versus days of vaccine administration.
Figure 34	Total NK cell population growth observed on day to day basis post vaccination.
Figure 35	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 36	Macrophage population growth observed by classification into active, resting, MHC II presenting and internalized groups along with total growth versus days of vaccine administration.
Figure 37	Concentration of total active MHC I presented and virally infected cells versus days post vaccination.
Figure 38	Cytokine population growth versus with danger indicated in the inset.

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 lymphocyte.
HTL	Helper T lymphocyte
Covid-19	Corona Viral Disease 2019
DMV	Di-membrane Vesicle
DNA	Deoxy-ribonucleic acid.
RNA	Ribonucleic acid.
IFN	Interferon
E protein	Envelope Protein
S protein	Spike protein
M protein	Membrane glycoprotein
N protein	Nucleocapsid protein
ACE-2	Angiotensin Converting Enzyme-2

Chapter 1

Introduction & Literature Review

Vaccine provide an active acquired immunity when administered for specific infectious disease or diseases. A vaccine generally made up of a disease-causing microorganisms, its toxins, or one of its surface proteins, all of which are used to prevent or treat disease. At first, it's important to

know what a vaccine is before getting into the process of developing one. The term "vaccine" is derived from the Latin phrase Variolae vaccinae, which Edward Jener utilized in 1798 to eliminate smallpox from people. To boost the immune system, vaccines are provided to the host in a biological or semi-biological form. The innate immunity might be either: 1. Prophylactic. 2. Therapeutic. Therapeutic vaccines help to get rid of the disease that the host has already encountered by making the host's immune cells stronger. Therapeutic vaccines are designed to treat an existing condition whereas prophylactic vaccines are designed to avoid the occurrence of a disease. Vaccines are divided into some major types like live attenuated vaccine, inactivated vaccine, subunit vaccine, toxoid vaccine, conjugate polysaccharide vaccine, nucleic acid vaccine. One important point to remember is that vaccinations are designed to combat infectious illnesses with a pathogenic source. Virus, bacteria, algae, or any other type of pathogen might be the cause of the disease. Vaccines are in the process of being created that target the diseases or their particular morphological organelles (whole cell vaccination). Vaccine development is a lengthy and complex process by itself. As a result, they are divided into numerous groups based on their viewpoints.

1.1: Pathogen vaccines:

The most conventional type of vaccination, that include the complete pathogen but having the antigenicity lowered. Vaccines derived from this source have the potential to induce strong immune responses in the host against the pathogen. Despite the fact that this method is used to produce the majority of vaccines for clinical use and only a small number of vaccines for production. To put it another way, when a full cell pathogen enters the host body, it typically leads in a strong immune reaction within the target host that may end in the loss of such host or serious systemic damage.

1.2: Inactivated Vaccines:

In the nineteenth century, the discovery of inactive or dead illnesses capable of developing immunity within the target host led to the invention of inactivated vaccines by heat or chemical treatment of the causing pathogen. NIAID-developed vaccine Havrix an example of a dormant vaccine against the hepatitis A virus, which was approved in the United States in 1995 with US assistance.

1.3: Live attenuated Vaccines:

Live attenuated vaccines were developed in the 1950s. These vaccinations include a weakened form of the pathogen that has been shown to be less harmful. The MMR vaccine, which protects against the disease of measles, rubella and mumps which protects against chickenpox, are examples of such vaccines. Despite the fact that this method is capable of inducing strong and long-lasting immunity against a specific pathogen when administered at a certain dose, a major limitation is observed that, the development of antibacterial and antiparasitic vaccines is relatively difficult due to the more complex morphological features of these pathogens unlike the pathogen of viral origin.

1.4: Subunit Vaccines:

Rather than using whole cell pathogens, subunit vaccines are constructed extracting specific antigens that elicit a robust the host's immune system responds. The immunological reaction is boosted and the vaccine is stabilized with the addition of adjuvants and linker molecules. Subunit vaccines are associated with less adverse effects than the aforementioned vaccines. While the

whole cell vaccinations were functionally successful, they frequently produced fever and edema at injection sites, contributing in an increase of new rates of infection. A novel pertussis vaccine was produced using acellular B pertussis organisms, which were previously unavailable. It was not composed of entire cells, but rather of pure cellular components. There have been instances of serious reactions including death.

1.5: Nucleic Acid Vaccines:

Nucleic acid vaccine production is a relatively new approach of vaccine development. The antigenic protein is injected into the host system by genetic material that encodes the protein. This approach is appropriate for vaccine creation because of its wide application to a wide variety of diseases, the capacity to provide a higher immune system response and long-term protection to a host, and the simplicity of increasing vaccine development at the industrial scale. Nucleic acid vaccine is mutually variable because they may be constructed using a variety of nucleic acids. To develop a DNA plasmid vaccine for example, the genetic material for a powerful antigenic protein is identified and placed into the plasmid of a bacterium host that is susceptible to the disease. The antigenic protein that acts as the major backbone of the construct will be generated by competent bacterial cells carrying the GOI. Purification, quality control, and testing will continue until the final product, which is the commercial vaccine gets developed. mRNA vaccines are another new class of vaccination that is created using the mRNA template of the causative pathogen's antigenic POI. Non-pathogenic carriers of bacterial or viral source with low antigenicity are often used to transfer nucleic vaccine into the host system, and these carriers have a low antigenicity. HPV vaccine is a form of a nucleic acid vaccination, while options for vaccines against the EBOV and Zika virus are now being investigated under the supervision of the National Institutes of Health (NIAID).

Chapter 2

An Overview of the Genomic and SARS-CoV-2 Pathogenic Origins:

2.1. Virus Morphology and Pathogenesis:

Viridae are the superkingdom of viruses; the phylum is Riboviria; the order is Nidovirales; the sub-order is Cornidovirineae; the family is Coronaviridae; the sub-family is Orthocoronaviridae; the genus is Betacoronavirus; the sub-genus is Sarbecovirus; and the species is Severe Acute Respiratory Syndrome CoronaVirus 2.
First, the virus's taxonomy was evaluated, and the work was finished and stated by the International Commission on Taxonomy of Viruses (ICT) previously in order to have a better understanding of SARS-CoV-2. As previously stated, SARS CoV-2 is a retrovirus with a general sense RNA genome that belongs to the Coronaviridae family. Coronavirinae members are further categorized into four genera based on the phylogenetic tree's rooted or unrooted nature and partial sequence of RNA polymerases that are RNA dependent (Henry & T, 2019; Woo et al., 2009). As hosts, most coronaviruses are found in warm-blooded animals, while members of the coronavirus genus infect primarily birds as hosts (Barreto-Vieira et al., 2021; Cui et al., 2019).

2.2: SARS CoV-2 genomic and morphological organization:

There are 6-11 ORFs in the SARS-Cov-2 genome that code for 9680 polyproteins. SARS-Cov-2 seems to have a positive sense, linear ssRNA genome that is approximately 30 kbp in length. The hemagglutinin esterase gene is missing from the genome, however it does contain two flanking untranslated regions (UTRs) at the 5' end of the 265th nucleotide and the 3' end of the 358th nucleotide. One of the SARS-CoV-2 nucleoside proteases is a pepaine homologue (nsp3), however, the other (nsp4) is the most important protease. There is also a chymotrypsin homologue (nsp5), which has three carbon atoms and is similar to pepaine (Grudlewska-Buda et al., 2021; Klein et al., 2020). Encoding the whole virus' poly proteomic structure falls to the viral genomic sequence. While the non-structural proteins are in charge of keeping the system running smoothly, structural proteins are those that give a virus its morphological structure. The composite proteomic structure is made up of four major structural proteins: spike (S), membrane glycoprotein (M), nucleocapsid (N), and envelope (E) (Mariano et al., 2020; Mittal et al., 2020). The viral RNA is contained within the N protein, which is encapsulated by the S, M, and E proteins. S is a structural protein that interacts with the ACE-2 receptor and is responsible for the spread of viral infection in the body. The S protein, which is a therapeutic target, is cleaved at the S1 (685 amino acids) and S2 (588 amino acids) subunits by proteolytic cleavage. When it comes to function in morphogenesis, assembly, and budding (as opposed to the structural aspects), E and M are responsible. S is a fusion protein that operates in the same manner as E and M (Wang et al., 2020).

2.3: SARS Cov-2 entry, replication, and pathogenesis in the host system:

ACE-2 receptor and the TMPRSS2 serine protease are used by SARS-CoV-2, which is homologous to its parent SARS-CoV, for internalization and priming of the S protein, and eventually for the propagation of the virus (V'kovski et al., 2021). Studies have also shown that the S protein of SARS-CoV-2 has a 10-20 times stronger affinity for ACE-2 receptor than the S protein of SARS-Cov-1. This is consistent with previous findings ((Harrison et al., 2020; Shang et al., 2020). A protein-ligand complex is formed when S binds to the ACE-2 receptor. S protein conformation changes as a result of this complex formation that leads to M Protein-E protein fusions, which allow entrance into the host cell via the endosomal route. The viral RNA is released into the host cell's cytoplasm. Ribosomal frameshifting is necessary for SARS Cov-2 replication. When a virus needs to be assembled, the endoplasmic reticulum (ER) and the Golgi complex work together to accomplish it. Exocytosis is used to release the complexes after they are generated and packed into vesicles (Trougakos et al., 2021).

Chapter 3

Methodology

3.1: Methodologies Overview:

As already stated in the abstract, the study's aim was to create a multi-epitope peptide based subunit vaccine against SARS-CoV-2 using an in-silico method. To discover the most appropriate antigen, researchers used a procedure based on the antigenicity of a protein instead of a whole proteome-driven search strategy. Instead of screening the whole proteome, at first the proteins in

the proteome was divided into two major groups. (1) Proteins that have structural functions. (2) Proteins that aren't structural (nsp). NSPs are virally essential proteins that keep the viral proteome running smoothly (Nain et al., 2019; Shey et al., 2019). The vaccine's primary antigen and structural backbone were chosen to be a viral protein. In contrast to a whole cell vaccine, which targeted the entire viral proteome, a subunit vaccine targets a specific viral antigen. The structural viral protein was designed to be highly antigenic, such that even after vaccine manufacture, However, the host's natural immunity to SARS-CoV-2 is robust enough to prevent the host from succumbing to the disease. After 14 days or two weeks of dormancy, SARS-CoV2 may reactivate and cause COVID-19 symptoms. Overview of applied methodologies are as follows:
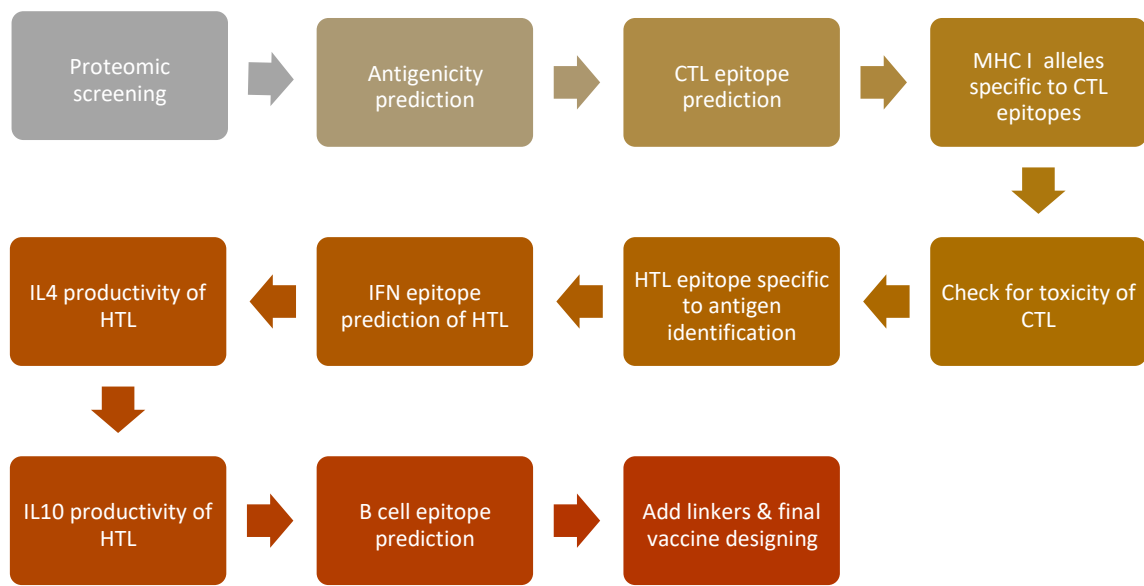

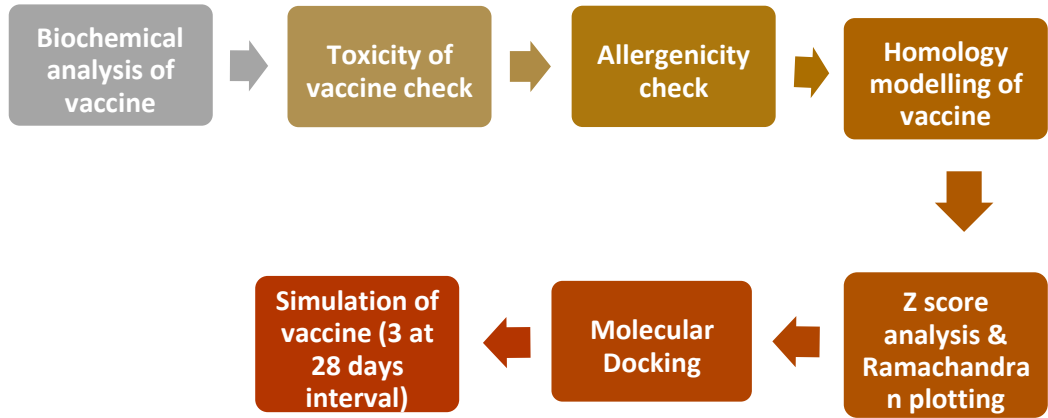Fig 1: Process of construction of final vaccine.

Fig 2: Biochemical analysis of vaccine

The Nucleocapsid Protein (N) of SARS-CoV-2 was chosen as the vaccine's primary antigen after it was proven to have an antigenicity level of at least 0.5 on the Vaxijen 2.0 server and also after screening against other structural proteins. Once the main antigen had been chosen, FASTA formatted sequence was used on the NetCTL 1.2 server to look for lymphocytic epitopes of cytotoxic T-cells. Using the binding affinity of MHC-I, combined scoring, TAP efficiency, and weighted C-terminal cleavage effectiveness criteria CTL epitopes were sorted, with a minimum total combined score of 0.75. Using NetMHC pan 4.0 servers for MHC I and MHC II, the matched peptides were chosen as a metric of binding affinity using the value of percentile rank. The fewer the value of percentile rank, the greater the peptide's affinity for the antigen, and the greater the antigen's effectiveness. NetCTL 1.2 server was used for MHC I peptide, where N protein was used as an input. MHC I epitopes were chosen with a 2% maximum threshold while MHC II epitopes were chosen with a 0.5 percent maximum threshold. MHC II peptides showed a lower percentile rank as compared to MHC I molecules since the peptide sequences were more scattered. Every time, similar or nearly identical epitopes, as well as repetitive peptides were screened out. Tests were performed using pred servers to see whether MHC II epitopes trigger an immunological response and as a consequence, increase immunity or not when administered to mice. The results showed that both IL-4 and IL-10 production was stimulated when administered to mice. Finally, the Biopred Linear Epitope Prediction 2.0 algorithm was used to pick B cell epitopes unique to the selected main antigen in order to complete the epitope identification process for vaccine designing. The 0.5 threshold was used for this method. Afterwards, vaccine's linear main protein structure was catenated along with all of the epitopes that had been identified before, with appropriate linkers placed at regular intervals to aid in the stability and that had been identified before, with appropriate linkers placed at regular intervals to aid in the stability and expression efficiency of the designed vaccine. In order to complete the vaccine construct's biochemical analysis, the PROTPARAM instrument was used in conjunction with the biochemical analysis of the vaccine's main antigen. The N protein of FASTA sequence was utilized as an input for this project. PDB file of the vaccine went through more investigation later on. This file was used to obtain a 3d model for the vaccine, which was built utilizing Phyre2's homology modeling capabilities. The PDB model was then used to construct a Ramachandran plot, which was used to anticipate a mathematically probabilistic vaccine model. Furthermore, the mean Z-score was calculated, and a Z-score vs. residue graph was produced. Then the next step was to bind TLR-8 receptor of the vaccine, which is part of the toll-like receptor family, with the receptor in the PDB. To conclude the study allergenicity was evaluated again and then an insilico simulation was accomplished to determine whether the vaccine can increase host protection against the SARS COV-2 virus or not. The AllergenOnline server and the C-immsim online immuno-simulator both were used to complete the assessment, and both were quite successful in getting the desired results (Nain et al., 2019; Shey et al., 2019).

3.2: Primary antigen selection:

To properly the methodologies and the web resources used at each stage of the study, the best way will be taking a step back and begin with the fundamentals. The objective of the study in this case was to create a peptide vaccine against N protein of SARS-CoV-2. Unlike complete vaccines, which deliver the pathogen's whole proteome into the host, subunit vaccines utilize just

chosen proteins. To enhance the final vaccine's pathogenicity, certain epitopes were added in precise quantities along with the antigenic Nucleocapsid protein. The incorporation of linkers, which are protein fusions that act as stabilizers and promoters of protein stability as well as expression, was also a part of this process. The final product is often less harmful than the crude protein produced from the pathogen, but possess an optimum capacity to stimulate an adequate immune response inside the host (Dutta et al., 2020). Beginning with the proteome of SARSCoV-2, the Vipr database software was used to screen for suitable protein candidates that would be antigenic enough here to elicit an effective immune response in the host but would not be potentially life-threatening to the host. Eventually, a suitable protein candidate was identified. A protein candidate was found that has the ability to fulfill both these parameters. Here, the proteins available in search of an ideal candidates were evaluated. Structural proteins were emphasized in this regard since it can be utilized as a medium in identification of the pathogen with relative ease. SARS-COV-2 can remain dormant in the target host for up to two weeks, therefore a nonstructural protein might not be a viable target for attaining the goal of targeting (Poran et al., 2020). Initial study focused on the virus's morphological properties, with the objective of discovering proteins that may be utilized as the vaccine's backbone. Following that, protein's possible antigenicity was tested using the Vaxijen 2.0 server, which can be found at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html. The tool was set to target a virus with a threshold of 0.5 (Chen et al., 2020).

3.3: Lymphocytic T-cell Cytotoxic Epitope Identification:

Immune cells that exhibit the T-cell receptor (TCR) on their surfaces are highly potent cytotoxic T-cell lymphocytes that are capable of starting an immune reaction through inflammation and chemokine/cytokine release. As a result, T-cell epitopes may have the potential to function as a robust immune booster in the host, enhancing vaccine-mediated protection. The NetCTL 1.2 server available at http://www.cbs.dtu.dk/services/NetCTL/ was utilized to detect epitopes of CTL. This server identifies CTL epitopes unique to a given antigen using the FASTA file format as input. Here, CTL epitopes are discovered using TAP effectiveness, binding affinity of MHC I, and C-terminal cleavage weight CRITERIAS. Comparative analysis of the epitopes against 12 MHC super types is possible. While artificial neural networks are used to determine the MHC I binding affinity and proteasomal cleavage capacity, a weighted matrix approach is used to calculate the TAP transportation efficacy. C-terminal cleavage and TAP transport were weighed at 0.15 and 0.05, respectively, while maintaining a 0.75 epitope prediction threshold (Khairkhah et al., 2020)(Sohail et al., 2021).

3.4: MHC I Alleles identification:

NetMHC 1 Pan 4.1 server specific to CTL epitopes was used to identify MHC I alleles. In this part of the inquiry, the NetCTL server output was used as input. The level of binding affinity was indicated by the percentile rank. The stronger an epitope's binding affinity for an allele, the lesser the percentile rank. Epitopes with a percentile level of 2 were chosen. The discovered MHC I peptides were homologous to previously identified CTL epitopes. These epitopes were not included into the vaccine formulation further (Reynisson et al., 2021).

3.5: MHC II Alleles identification:

As a whole, adaptive immunity is enhanced by the presence of MHC II alleles in the host's immune system. The HTL epitope was identified by using NetMHC II pan 4.0 server. With a preset peptide length of 15, we used the primary antigen as our input (Hoque et al., 2021). Percentile rank was employed as a possible indication in between epitope and antigen through binding affinity in similar way of the previous stage. This time, though, the highest permissible percentile rank barrier was adjusted to 0.5. This was performed because the peptides included a greater diversity of epitopes than the MHC I epitopes (Dimitrov et al., 2010)(Chakravarty & Vora, 2021).

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

This research focused on predicting IFN stimulating capacity, IL4 productivity and IL10 productivity of HTL for combined analysis of MHC ii peptide induction capability in this stage. For these purposes, the IFN epitope server (,) IL4 pred server (,) and IL10 pred server were used. During the immuno-simulation phase of this research these capabilities were further analyzed (Chukwudozie et al., 2021; Poran et al., 2020).

3.7: B-cell Epitopes identification:

By sequestering immunoglobulins that neutralize antigens that adhere to B-cell receptors, B-cell epitopes can activate humoral immunity (BCR). These epitopes may be linear or continuous in nature, or conformational or discontinuous in nature. Moreover, because of the ease of modifying the main structure of linear epitopes, they are commonly used in vaccine design. The Bepipred Linear Epitope Prediction 2.0 was used to find B cell epitopes corresponding to the main antigenic sequence derived at the very beginning, at a threshold of (Jespersen et al., 2017)(Phan et al., 2021).

3.8: Primary vaccine structure construction:

Based on a main antigen found at https://www.viprbrc.org, the ViPR Database, the multi - epitope vaccine design was constructed. As part of the FASTA format, the protein's structure was enhanced to contain MHC I, MHC II, and B-cell epitopes. For additional biochemical analysis, immunological, and statistical investigation of the final vaccine construction was finished.(Pickett et al., 2012)(Pollett et al., 2021).

3.9: Analysis of the Vaccine's Biochemical Constituents:

The biochemical characteristics of the vaccine were determined at this step using the Protparam tool on the Expasy website. The server at https://web.expasy.org/protparam/ accepts a protein's primary structure as input and analyzes its various biochemical properties, along with its molar mass, theoretical PI, extinction coefficient estimated, half-life prediction, aliphatic index of vaccine, aa composition, vaccine stability index, and grand average of hydropathicity (GRAVY). Furthermore, the vaccine was evaluated in comparison to a protein content control obtained from the ViPR Database (Gao et al., 2021)(Shu et al., 2020).

3.10: Vaccine build homology modeling to generate 3D model:

To generate a 3D structure of the vaccine construct, homology modelling was used as the main structure that was previously created based on entries in the PDB database, which may be found at https://www.rcsb.org/. For this stage, phyre2 server was used, which may be found at http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index. The server generates a 3D representation of a fundamental protein structure and also displays homology of the generated structure in terms of percentage with respect to the utilized templates. The PDB structure acquired is essential for the further study (Jaimes et al., 2020)(Alazmi & Motwalli, 2021).

3.11: Ramachandran Plotting and Analysis of the Tertiary PDB 290 Structure:

Before proceeding with additional study on the protein, it was necessary to investigate the tertiary structure created by homology modeling. In this scenario, two key operations were conducted on the protein's basic model. To begin, a Ramachandran plot was generated using the SWISS PDB plotter, which may be accessed at https://spdbv.vital-it.ch/rama.html. Using a quadrant approach, this map depicts the most likely location of the residues. This graph illustrates a statistical estimation of the distribution of a protein (Yang et al., 2021)(Sarkar & Saha, 2020). A quality indicator for the tertiary model was generated using PROSA web, which displayed z-score against residue (Dash et al., 2021).

3.12: Molecular Docking Analysis of the TLR-8 Receptor and the Vaccine Construct:

To calculate the binding affinity of a receptor with its matching ligand, which results in a binding protein, molecular docking is used, which is an in-silico approach that may be performed. It was necessary to make use of the patchdock server (available at: http://bioinfo3d.cs.tau.edu/PatchDock/index.html). In order to perform the required docking function, the server employs an algorithm which is in the context of image processing, based on object identification and image segmentation.

3.13: In-silico based simulation of the final vaccine for human hosts:

An in silico immunological response in a host body is being modelled as the final step in this research. The C-Immsim server, available at http://kraken.iac.rm.cnr.it/CIMMSIM/index.php?page=1, was used for simulation. Vaccine dosage, simulation phases, and parameters may all be entered into the server using a FASTA protein sequences. During the simulation, there were a total of 300 steps. There was a tri-dosage injection system at steps 1,84 and 168. However, the vaccination was intended to be given three times at 28-day intervals.

1st step=8 hours.

84th step=672 hours=28th day

168th step=1344 hours=56th day

The vaccine parameters (adjuvant and antigen) were established at 100μg & 1000μg respectively. To depict the immune response in a host body following vaccination, a variety of plots had been developed. Each bar graph shows changes in the immune status of the host as measured by the number of immune cells present and the concentrations of 322 immunoglobulin and antigen with

time (Abraham Peele et al., 2020). The immuno-simulation stage is an in-silico process that is used to evaluate the efficiency of a vaccine computationally before going on to a live system study (Castiglione et al., 2021)(Kar et al., 2020).

3.14: Checking for host allergenicity in the final vaccine construction:

The complete design of vaccine was then tested to see if it contains any allergen. Allergen Online server was utilized that is located at http://www.allergenonline.org/. The service receives FASTA protein files as input and returns outputs of homologous sequence data that have the potential to function as allergens, as well as the allergenic potential particles names that are similar to the input sequences. This approach is necessary in order to limit the possibility of unfavorable effects developing later in the research (Akhand et al., 2020)(Kleine-Tebbe & Mailänder, 2020).

3.15: Final Remarks on the Methodology:

The ending note regarding the applied approach of this study will be that, the technique followed in this research was completely based on the use of in-silico technologies More precisely, at every stage of this analysis had relied on online servers. Furthermore, this study makes no claims that the final version will be a highly efficient vaccination capable of entirely eradicating COVID-19 viral infection. It does, however, have the ability to become a vaccine potential, as well as further study in this area is required, specifically at the moist lab and therapeutic levels.

Chapter 4

Analysis and Interpretation of Results:

In this section of the investigation, the results collected at each stage of the experiment will be arranged in chronological order. Additionally, the acquired result will be interpreted in such a way that the reader can easily assess the quality of work performed in this area.

4.1: Prediction of Antigenicity:

After a thorough screening process, the Nucleocapsid Protein (N) of SARS-CoV-2 was selected as the core antigen. The following is the primary structural sequence of core antigen: Protein Sequence:

MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQ
HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTG
PEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAE
GSRGGSQASSRSSSRSRNSSRNSTPGSSKRTSPARMAGNGGDAALALLLLDRLNQLESK
MSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQ
ELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFK
DQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSK
QLQQSMSSADSTQA

Input and output of Vaxijen 2.0 server:
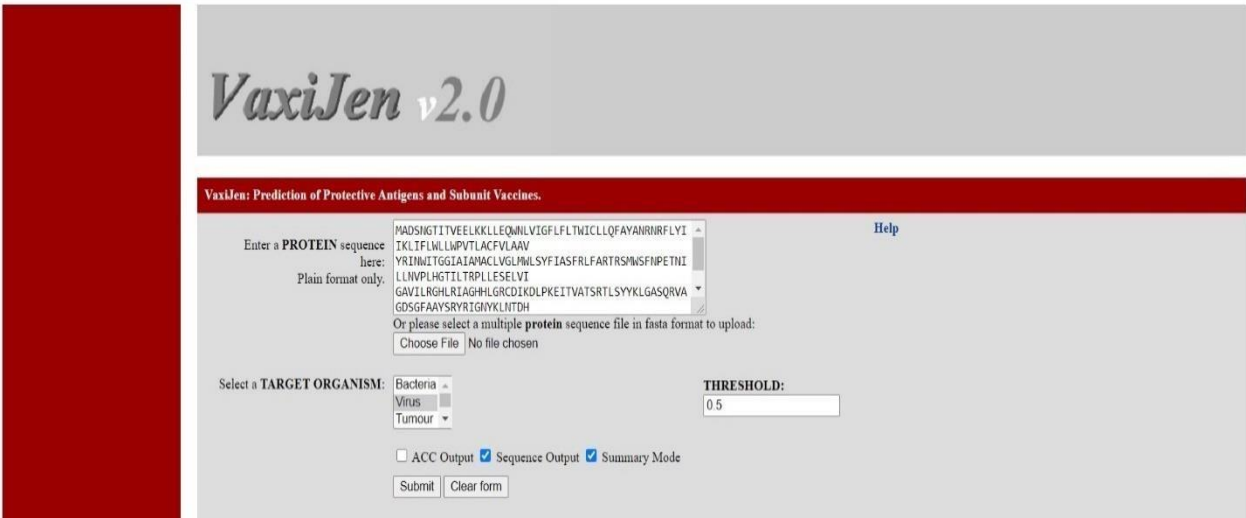

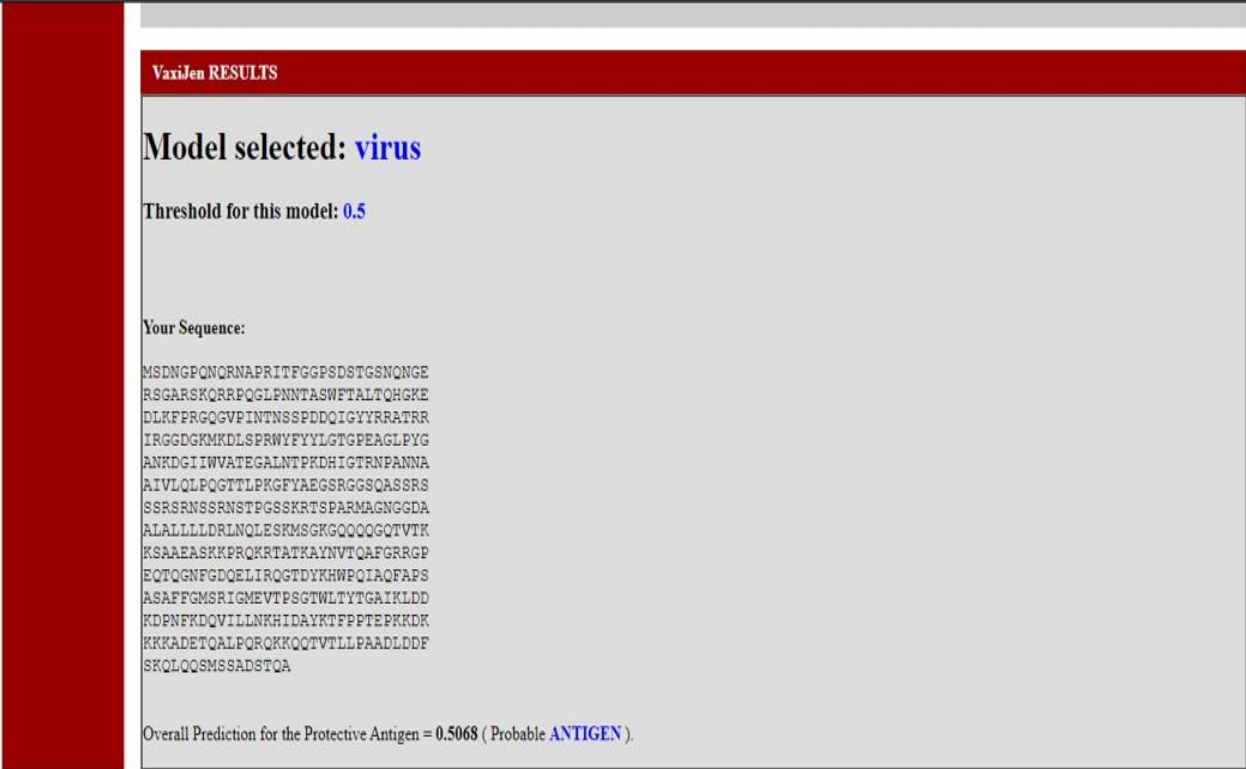Fig 3: Primary structure of the antigen has been entered into Vaxijen 2.0.

Fig 4: The antigenicity of core protein is 0.5068 on the Vaxijen 2.0.

4.2: Identification of CTL Epitope:

CTL epitopes are necessary to increase the immune response of host. Ag-specific TCRs expressed on them combat the specific ag. The NetCTL 1.2 server identified CTL epitopes. The A1 super type of MHC I alleles epitopes were found at 0.75. The neural network-based approach that was applied in this case is described below. When it comes to epitope selection, the combined score is quite important. This combination score is based on the cleavage of the C terminal and the efficiency of TAP transit. The minimum thresholds were established at 0.15 and 0.05. The final data are as follows:

CTL epitope	Combined Score
LSPRWYFYY	2.3408
GTTLPKGFY	1.6848
DLSPRWYFY	1.4994
SSPDDQIGY	1.4541
LLNKHIDAY	1.3867
GTDYKHWPQ	1.2059
SPDDQIGYY	1.1404
NTASWFTAL	0.9521
MKDLSPRWY	0.9120

Table 1: Primary findings of CTL epitopes and their combined scores

To indicate the specificity and sensitivity of matching epitopes, the combined scores can be employed 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 2: Combined sensitivity and specificity score expression system.

The following is an example of the data collected by the NetCTL 1.2.

NetCTL-1.2 Server - prediction results

Technical University of Denmark

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

104 ID vac pep LSPRWYFYY aff   0.4837 aff_rescale  2.0538 cle 0.9746 tap  2.8150 COMB  2.3408 <-E
164 ID vac pep GTTLPKGFY aff   0.3315 aff_rescale  1.4074 cle 0.9691 tap  2.6390 COMB  1.6848 <-E
 78 ID vac pep SSPDDQIGY aff   0.2720 aff_rescale  1.1550 cle 0.9463 tap  3.1440 COMB  1.4541 <-E
352 ID vac pep LLNKHIDAY aff   0.2587 aff_rescale  1.0982 cle 0.9685 tap  2.8630 COMB  1.3867 <-E
295 ID vac pep GTDYKHWPQ aff   0.2889 aff_rescale  1.2267 cle 0.0381 tap -0.5310 COMB  1.2059 <-E
 79 ID vac pep SPDDQIGYY aff   0.2040 aff_rescale  0.8663 cle 0.9769 tap  2.5510 COMB  1.1404 <-E
 48 ID vac pep NTASWFTAL aff   0.1772 aff_rescale  0.7523 cle 0.9557 tap  1.1280 COMB  0.9521 <-E
101 ID vac pep MKDLSPRWY aff   0.1547 aff_rescale  0.6569 cle 0.7184 tap  2.9460 COMB  0.9120 <-E
325 ID vac pep TPSGTWLTY aff   0.1033 aff_rescale  0.4388 cle 0.9754 tap  2.6420 COMB  0.7172
  1 ID vac pep MSDNGPQNQ aff   0.1594 aff_rescale  0.6770 cle 0.1751 tap -0.1960 COMB  0.6935
262 ID vac pep RTATKAYNV aff   0.1425 aff_rescale  0.6052 cle 0.3161 tap  0.6550 COMB  0.6854
290 ID vac pep ELIRQGTDY aff   0.0914 aff_rescale  0.3881 cle 0.9490 tap  2.7600 COMB  0.6684
307 ID vac pep FAPSASAFF aff   0.0883 aff_rescale  0.3747 cle 0.6765 tap  2.4280 COMB  0.5976
115 ID vac pep TGPEAGLPY aff   0.0753 aff_rescale  0.3197 cle 0.9688 tap  2.5920 COMB  0.5946
 53 ID vac pep FTALTQHGK aff   0.1199 aff_rescale  0.5091 cle 0.4246 tap  0.3090 COMB  0.5883
338 ID vac pep KLDDKDPNF aff   0.0813 aff_rescale  0.3450 cle 0.7843 tap  2.4940 COMB  0.5873
331 ID vac pep LTYTGAIKL aff   0.0880 aff_rescale  0.3736 cle 0.9658 tap  1.0680 COMB  0.5719
266 ID vac pep KAYNVTQAF aff   0.0630 aff_rescale  0.2677 cle 0.9736 tap  2.8910 COMB  0.5583
  9 ID vac pep QRNAPRITF aff   0.0579 aff_rescale  0.2456 cle 0.9735 tap  2.8010 COMB  0.5317
216 ID vac pep DAALALLLL aff   0.0853 aff_rescale  0.3623 cle 0.8987 tap  0.6780 COMB  0.5310
260 ID vac pep QKRTATKAY aff   0.0553 aff_rescale  0.2350 cle 0.8870 tap  3.1680 COMB  0.5264
316 ID vac pep GMSRIGMEV aff   0.0876 aff_rescale  0.3719 cle 0.9115 tap  0.1480 COMB  0.5160
229 ID vac pep QLESKMSGK aff   0.0867 aff_rescale  0.3681 cle 0.6815 tap  0.4940 COMB  0.4950
399 ID vac pep DLDDESKOL aff   0.0721 aff_rescale  0.3059 cle 0.9780 tap  0.6360 COMB  0.4844
```

Fig 5: Here is a screenshot of the NetCTL 1.2 server's result

4.3: CTL Epitope specificity for MHC I allele:

NetMHC Pan 4.1 server was utilized to obtain MHC I alleles from epitopes defined in part 4.2. Percentile rank is the parameter used in epitope selection in this caseThe lower the percentile rank, the higher the binding affinity, and vice versa. For epitope selection, a minimum threshold of 2 or less than 2 was applied. CTL epitope and MHC I allele specific binding is listed below, along with the corresponding binding affinity in percentile rank.

MHC 1 alleles specific to CTL epitopes:

For Strong binding peptides, threshold 0.500
For Weak binding peptides, threshold   2.000

		Percentile Rank
LSPRWYFYY	HLA-A*01:01	0.424
GTTLPKGFY	HLA-A*01:01	0.422

Peptide sequence	Alleles covered	
DLSPRWYFY	HLA-A*01:01	0.493
SSPDDQIGY	HLA-A*01:01	0.168
	HLA-A*26:01	0.240
LLNKHIDAY	HLA-A*01:01	0.348
SPDDQIGYY	HLA-A*01:01	0.200
DLSPRWYFY	HLA-A*26:01	0.157
NTASWFTAL	HLA-A*26:01	0.240
LLNKHIDAY	HLA-B*15:01	0.064

Table 3: Alleles covered by MHC I Peptides, as well as their associated percentile rank.

A screenshot of the results was taken that was obtained at this step via the NetMHC pan 4.1 server to put an end to this section.

NetMHCpan Server - prediction results

Technical University of Denmark

```
# NetMHCpan version 4.1b

# Tmpdir made /usr/opt/www/webface/tmp/server/netmhcpan/612FD7BB0000104F49484077/netMHCpanJojaMJ
# Input is in FSA format

# Peptide length 9

# Make EL predictions

HLA-A01:01 : Distance to training data  0.000 (using nearest neighbor HLA-A01:01)

# Rank Threshold for Strong binding peptides   0.500
# Rank Threshold for Weak binding peptides   2.000
-----------------------------------------------------------------------------------
 Pos      MHC       Peptide    Core Of Gp Gl Ip Il      Icore      Identity  Score_EL %Rank_EL BindLevel
-----------------------------------------------------------------------------------
   1 HLA-A*01:01    LSPRWYFYY LSPRWYFYY  0  0  0  0  0   LSPRWYFYY           vac1 0.3760470   0.424 <= SB
-----------------------------------------------------------------------------------

Protein vac1. Allele HLA-A*01:01. Number of high binders 1. Number of weak binders 0. Number of peptides 1

Link to Allele Frequencies in Worldwide Populations HLA-A01:01
-----------------------------------------------------------------------------------
# Rank Threshold for Strong binding peptides   0.500
# Rank Threshold for Weak binding peptides   2.000
-----------------------------------------------------------------------------------
 Pos      MHC       Peptide    Core Of Gp Gl Ip Il      Icore      Identity  Score_EL %Rank_EL BindLevel
-----------------------------------------------------------------------------------
   1 HLA-A*01:01    GTTLPKGFY GTTLPKGFY  0  0  0  0  0   GTTLPKGFY           vac2 0.3769020   0.422 <= SB
```

4.4: HTL Epitope specificity for MHC II allele:

Adaptive immunity can be improved by MHC II alleles because they bind to HTLs and generate a host's adaptive immunity that is improved as a result of increased humoral and cell-mediated immunity. As previously indicated, the NetMHCIIpan 4.0 server was used to detect MHC II alleles. MHC II alleles were found using the main antigen as an input. Additionally, we employed percentile rank to detect MHC II alleles; fewer percentile rank indicates greater binding affinity, and vice versa. In this situation, an allele identification percentile rank of 0.5 was used. The following are the findings:

MHC 2 alleles specific to HTL epitopes:
For Strong binding peptides, threshold (%Rank) 1%
For Weak binding peptides, threshold (%Rank) 5%

Peptide sequence	Core	Alleles covered	Percentile Rank	Binding Level	Score-EL
KQQTVTLLPAADLDD	VTLLPAADL	DRB1_0101	0.47	<=SB	0.889674
QQTVTLLPAADLDDF			0.48	<=SB	0.886419
KKQQTVTLLPAADLD			1.63	<=WB	0.645832
QTVTLLPAADLDDFS			1.97	<=WB	0.584190
PKGFYAEGSRGGSQA	FYAEGSRGG	DRB1_0101	0.63	<=SB	0.852024
LPKGFYAEGSRGGSQ			0.96	<=SB	0.782106
KGFYAEGSRGGSQAS			2.03	<=WB	0.574594
TLPKGFYAEGSRGGS			2.10	<=WB	0.563418

Peptide	Core	Allele	Score	Level	Rank
			2.42	<=WB	
		DRB1_010	3.26	<=WB	
PKGFYAEGSRGGSQA		2			0.568253
LPKGFYAEGSRGGSQ			0.27	<=SB	0.481545
			0.54	<=SB	
		DRB1_010	1.13	<=WB	
PKGFYAEGSRGGSQA		3	1.36	<=WB	
LPKGFYAEGSRGGSQ					0.815853
KGFYAEGSRGGSQAS					0.727495
TLPKGFYAEGSRGGS			0.15	<=SB	0.575057
		DRB1_0401			0.536357
PKGFYAEGSRGGSQA					0.900821
GTWLTYTGAIKLDDK	LTYTGAIKL	DRB1_010	1.53	<=WB	0.665807
SGTWLTYTGAIKLDD		1	2.90	<=WB	0.456049
TWLTYTGAIKLDDKD			4.60	<=WB	0.292744
			0.90	<=SB	

Peptide	Core	Allele	Rank	Level	Score
		DRB1_010 2	1.91	<=WB	
			3.09	<=WB	
GTWLTYTGAIKLDDK					0.797802
SGTWLTYTGAIKLDD			0.98	<=SB	0.633115
			2.09	<=WB	
TWLTYTGAIKLDDKD		DRB1_010 3	3.22	<=WB	0.497489
GTWLTYTGAIKLDDK					0.606006
SGTWLTYTGAIKLDD					0.439937
TWLTYTGAIKLDDKD					0.336775
TASWFTALTQHGKED	WFTALTQH	DRB1_010	2.08	<=WB	0.566671 G
			3.26	<=WB	0
NTASWFTALTQHGKE			4.49	<=WB	0.416444
ASWFTALTQHGKEDL					0.300522

GIIWVATEGALNTPK	WVATEGAL	DRB1_010	2.22	<=WB	0.54606
DGIIWVATEGALNTP	N	1	3.81	<=WB	0
					0.36020
					0
LDRLNQLESKMSGK	LNQLESKMS	DRB1_010	2.45	<=WB	0.51003
G		1	4.13	<=WB	2
LLDRLNQLESKMSG			4.39	<=WB	0.33092
K					4
DRLNQLESKMSGKG					0.30872
Q			1.19	<=WB	9
		DRB1_010	2.11	<=WB	
		2	2.43	<=WB	0.74676
LDRLNQLESKMSGK					2
G					0.60775
LLDRLNQLESKMSG					4
K					0.56742
DRLNQLESKMSGKG					2
Q					
TKAYNVTQAFGRRG	YNVTQAFG	DRB1_010	3.10	<=WB	0.43344
P	R	1	4.42	<=WB	5
					0.30647
ATKAYNVTQAFGRR					3
G					

KDGIIWVATEGALNT	IIWVATEGA	DRB1_0101	3.66	<=WB	0.374745
KDGIIWVATEGALNT			2.14	<=WB	0.604313
NKDGIIWVATEGALN		DRB1_0102	2.87	<=WB	0.519761
			0.91	<=SB	
NKDGIIWVATEGALN		DRB1_0401	0.09	<=SB	0.683541
KDGIIWVATEGALNT		DRB1_0404			0.945771
KQQTVTLLPAADLDD	VTLLPAADL	DRB1_0102	0.07	<=SB	0.968796
QQTVTLLPAADLDDF			0.08	<=SB	0.968712
KKQQTVTLLPAADLD			0.34	<=SB	0.909021
QTVTLLPAADLDDFS			0.45	<=SB	0.885454
NAAIVLQLPQGTTLP	IVLQLPQGT	DRB1_0102	3.40	<=WB	0.468621
QIGYYRRATRRIRGG	YYRRATRRI	DRB1_0103	0.33	<=SB	0.792207
DQIGYYRRATRRIRG			0.58	<=SB	

Peptide	Core	Allele	Value	Level	Score
					0.711627
GQTVTKKSAAEASKK	VTKKSAAEA	DRB1_0103	1.09	<=WB	0.582733
TGAIKLDDKDPNFKD	IKLDDKDPN	DRB1_0301	0.11	<=SB	0.943920
YTGAIKLDDKDPNFK			0.17	<=SB	0.925601
TGAIKLDDKDPNFKD			0.21	<=SB	
TYTGAIKLDDKDPNF		DRB1_0305	0.86	<=SB	
					0.641671
					0.408465
LALLLLDRLNQLESK	LLLDRLNQL	DRB1_0301	3.89	<=WB	0.292654
TASWFTALTQHGKED	WFTALTQHG	DRB1_0401	0.11	<=SB	0.919536
PRWYFYYLGTGPEAG	FYYLGTGPE	DRB1_0405	0.35	<=SB	0.843138
KDQVILLNKHIDAYK	VILLNKHID	DRB1_0808	0.02	<=SB	0.966627
		DRB1_1303			
		DRB1_1401			

Peptide	Core	Allele			Score
LTYTGAIKLDDKDPN 1	YTGAIKLDD 7	DRB1_080	0.37	<=SB	0.85149
KDQVILLNKHIDAYK 2	ILLNKHIDA 0.16 <=SB 2	DRB1_130	0.08	<=SB	0.73334
2	0	DRB1_140			0.48299
TKAYNVTQAFGRRG P 1	YNVTQAFG R 9	DRB1_130 2	1.43 0.39	<=WB <=SB	0.46157 0
		DRB1_160			0.81454
QIGYYRRATRRIRGG 2	YRRATRRIR 5	DRB1_130	0.80	<=SB	0.52607
NKHIDAYKTFPPTEP	IDAYKTFPP	DRB1_150 1	0.03 0.04	<=SB <=SB	0.97632 8
		DRB1_150 3	0.61	<=SB	0.89209 0
		DRB5_020 2			0.62356 3
GLPYGANKDGIIWV	YGANKDGII	DRB3_020	0.68	<=SB	0.48375

A		2			4
DQELIRQGTDYKHWP	LIRQGTDYK	DRB4_0101	0.63	<=SB	0.404927

Table 4: HTL epitope in relation to the associated allele, percentile rank, and binding level.

Fig 7: The NetMHCIIpan 4.0 server was used to identify MHC II alleles specific to HTL epitopes.

4.5: The capability of HTL epitopes to induce cytokine production:

The cytokines produced by HTL epitopes, which induce an inflammatory immune response, help to activate CTLs. When HTL epitopes were first identified as a source of interleukin production, they included IFN epitope prediction by HTL, IL-4 productivity by HTL, and IL-10 productivity by HTL IFN epitope, IL-4pred, and IL-10pred servers were employed in this step. In the picture below, you can see a consolidated tabulation of the results gathered from the servers. Using the SVM approach, the default thresholds of 0.2 and -0.3 were used for the prediction servers for the IL4 and IL10 preds. This is done as part of the process of optimizing algorithmic efficacy. At a later stage, further analysis has been performed.

Fig 8: IFN epitope prediction for HTL-positive is accepted.

Fig 9: IL4pred-inducer for HTL epitope.

IL10pred for productivity of HTL:
QIGYYRRATRRIRGG
DQIGYYRRATRRIRG
DQELIRQGTDYKHWP

	A	B	C	D
1	**Peptide**	**IFN**	**IL4**	**IL10**
2	KQQTVTLLPAADLDD	Negative	IL4 non-inducer	IL10 non-inducer
3	QQTVTLLPAADLDDF	Negative	IL4 non-inducer	IL10 non-inducer
4	PKGFYAEGSRGGSQA	Negative	IL4 non-inducer	IL10 non-inducer
5	LPKGFYAEGSRGGSQ	Negative	IL4 non-inducer	IL10 non-inducer
6	GTWLTYTGAIKLDDK	Negative	IL4 non-inducer	IL10 non-inducer
7	NKDGIIWVATEGALN	Negative	**IL4 inducer**	IL10 non-inducer
8	KDGIIWVATEGALNT	Negative	IL4 non-inducer	IL10 non-inducer
9	KKQQTVTLLPAADLD	Negative	IL4 non-inducer	IL10 non-inducer
10	QTVTLLPAADLDDFS	**Positive**	IL4 non-inducer	IL10 non-inducer
11	QIGYYRRATRRIRGG	**Positive**	**IL4 inducer**	**IL10 inducer**
12	DQIGYYRRATRRIRG	**Positive**	**IL4 inducer**	**IL10 inducer**
13	TGAIKLDDKDPNFKD	Negative	**IL4 inducer**	IL10 non-inducer
14	YTGAIKLDDKDPNFK	Negative	**IL4 inducer**	IL10 non-inducer
15	TYTGAIKLDDKDPNF	Negative	IL4 non-inducer	IL10 non-inducer
16	TASWFTALTQHGKED	Negative	IL4 non-inducer	IL10 non-inducer
17	PRWYFYYLGTGPEAG	Negative	**IL4 inducer**	IL10 non-inducer
18	KDQVILLNKHIDAYK	Negative	IL4 non-inducer	IL10 non-inducer
19	LTYTGAIKLDDKDPN	Negative	IL4 non-inducer	IL10 non-inducer
20	TKAYNVTQAFGRRGP	Negative	IL4 non-inducer	IL10 non-inducer
21	NKHIDAYKTFPPTEP	Negative	IL4 non-inducer	IL10 non-inducer
22	GLPYGANKDGIIWVA	Negative	**IL4 inducer**	IL10 non-inducer
23	DQELIRQGTDYKHWP	Negative	IL4 non-inducer	**IL10 inducer**

Fig 10: Prediction & evaluation of HTL epitopes for MHC II alleles

4.6: B-cell epitope identification:

Continuous and discontinuous B-cell epitopes are the two forms of B-cell epitopes. These epitopes serve as immuno-boosters for the host. As a result of the presence of B-cell receptors (BCRs) on these cells, the immune response can be boosted by a variety of immunoglobulins. Bcell epitopes may stimulate humoral and cell-mediated immunity, which can help improve overall adaptive immunity. Linear B-cell epitopes were identified using the Bepipred linear epitope identification 2.0 approach. Threshold of 0.5 was used to identify B-cell epitopes. Starting

and ending positions of the B-cell epitopes, as well as their lengths are shown in the following table:

Sequences	Starting	Ending	Length
NGPQNQRNAPRI	4	15	12
FGGPSDSTGSNQNGERSGARSKQRRPQGLPNN	17	48	32
QHGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS	58	105	48
AGLPYGAN	119	126	8
GALNTPKDHIGTRNPANNAAIVL	137	159	23
TLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSKRTSPARMAGNGGDA	166	217	52
LNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA	227	267	41
RRGPEQTQGNFGDQELIRQGTDYK	276	299	24
DPNFKD	343	348	6
DAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDD	358	402	45

	404	416	13
SKQLQQSMSSADS			

Table 5: Position Wise specification of B cell epitopes.

The predicted B cell epitopes were then clarified through the use of a graph obtained from the server, which plotted the epitopes' residue wise scores. This is the graph:
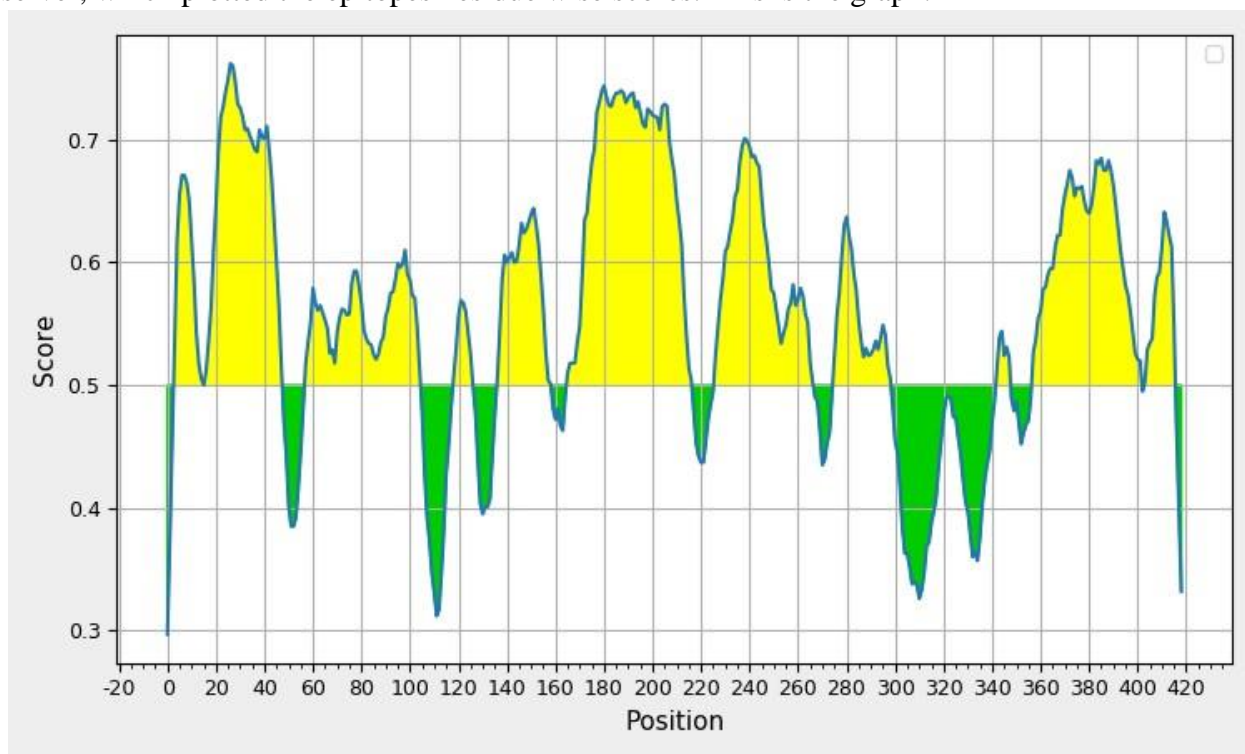
Fig 11: A graph of possible B-cell epitopes specific to the SARS-CoV-2 Nucleocapsid protein.

It is possible to track the highest, lowest, and average B cell epitope scores using the Bepipred Linear Epitope 2.0 algorithm utilized in this part. When it came to the specific area of interest (AOI), the highest score was 0.762, the lowest was 0.2797, and the average score was 0.5606.

4.7: Primary Vaccine Structure Construction:

The study's primary objective was to develop an in-silico based multiepitope subunit peptide vaccine against SARS-CoV-2 using the viral N-protein. All three types of antigens were introduced to the protein, which included B cell, CD8 T-cell, and HTL (Helper T Lymphocytes) epitopes. B cell epitopes are necessary for the increase of humoral immunity and the creation of IL, ig and other specific antibodies as possible enhancers of the immune response induction. There were previously described CTL and HTL epitopes specific for MHC I and MHC II alleles in this research. Additionally, the B cell epitopes can produce memory cells, which protect the

host in the future if the body is exposed to the same antigen. Antigenic protein templates from particular pathogens were used to produce an in-silico-based multiepitope peptide vaccine. Antigenicity can be optimized to the point where the vaccine is not lethal to the host while yet delivering considerable host system efficacy over an appropriate length of time by the integration of the aforementioned epitopes. As a result, the final vaccination must be tested in terms of sensitivity, specificity, and stability. When possible linker molecules are included in the equation, this goal is met. The vaccine template was modified to include fusion proteins in order to boost the vaccine's overall efficacy. Between the main protein and CTL epitope, an EAAAK linker was inserted; between the CTL epitopes, linkers of AAY were added; between the HTL epitopes, GPGPG linkers were added; and between the B-cell epitopes, KK linker was added 488 (He et al., 2021; Sayed et al., 2019; Usmani et al., 2018). Final vaccine construct is as follows:

MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQ
HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTG
PEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEG
SRGGSQASSRSSSRSRNSSRNSTPGSSKRTSPARMAGNGGDAALALLLLDRLNQLESKM
SGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQE
LIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKD
QVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQ
LQQSMSSADSTQAEAAAKLSPRWYFYYAAYGTTLPKGFYAAYDLSPRWYFYAAYSSP
DDQIGYAAYLLNKHIDAYAAYSPDDQIGYYAAYDLSPRWYFYAAYNTASWFTALAAY
LLNKHIDAYGPGPGQIGYYRRATRRIRGGGPGPGDQIGYYRRATRRIRGKKNGPQNQR
NAPRIKKFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNKKQHGKEDLKFPRGQGVPI
NTNSSPDDQIGYYRRATRRIRGGDGKMKDLSKKAGLPYGANKKGALNTPKDHIGTRNP
ANNAAIVLKKTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSKRTSPARMAGNG
GDAKKLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAKKRRGPEQTQ
GNFGDQELIRQGTDYKKKDPNFKDKKDAYKTFPPTEPKKDKKKKADETQALPQRQKK
QQTVTLLPAADLDDKKSKQLQQSMSSADS

4.8: Biochemical Analysis of Vaccine Construct:

Expasy's Protparam tool was used to conduct biochemical analysis at this phase of vaccine development. Analytical results include molar mass, molecular formula, stability index (which shows the stability of vaccine), theoretical PI, GRAVY values and aliphatic index. The server performs the test and delivers the analytics.

Parameter	Result
Amino acids	895

Molecular weight of vaccine	98489.84
Theoretical PI of N vaccine	10.19
(-) charged residues (Asp+Glu) value	74
(+) charged residues (Lys+Arg) value	154
Molecular Formula for vaccine	$C_{4287}H_{6796}N_{1322}O_{1330}S_{11}$
Estimated Half life	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).
Extinction Coefficient.	Ext. coefficient: 112080 Abs 0.1% (=1 g/l): 1.138
Instability index	60.89 (Unstable)
Aliphatic index	48.47
Grand average of hydropathicity (GRAVY) value	-1.110

Table 6: Biochemical testing of the vaccine design.

Parameter	Result
Amino acids	419
Molecular weight of N protein	45696.82
Theoretical PI of N protein	10.09
(-) charged residues (Asp+Glu) value	36

(+) charged residues (Lys+Arg) value	61
Molecular Formula for N protein	$C_{1975}H_{3146}N_{608}O_{629}S_7$
Extinction coefficient	Ext. coefficient 43890 Abs 0.1% (=1 g/l) 0.960
Estimated Half life	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).
Instability index	55.81 (Unstable)
Aliphatic index	52.53
Grand average of hydropathicity (GRAVY) value	-0.980

Table 7: Biochemical testing of the primary antigen N protein.

Biochemical analysis of the vaccine development was performed utilizing the vaccine's primary structure as an input and with that the corresponding outputs were produced. Placing data into the server looked like the following:

Fig 12: A screenshot of the vaccine's primary structure being entered into Expasy.

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MSDNGPQNQR NAPRITFGGP SDSTGSNQNG ERSGARSKQR RPQGLPNNTA SWFTALTQHG

        70         80         90        100        110        120
KEDLKFPRGQ GVPINTNSSP DDQIGYYRRA TRRIRGGDGK MKDLSPRWYF YYLGTGPEAG

       130        140        150        160        170        180
LPYGANKDGI IWVATEGALN TPKDHIGTRN PANNAAIVLQ LPQGTTLPKG FYAEGSRGGS

       190        200        210        220        230        240
QASSRSSSRS RNSSRNSTPG SSKRTSPARM AGNGGDAALA LLLLDRLNQL ESKMSGKGQQ

       250        260        270        280        290        300
QQGQTVTKKS AAEASKKPRQ KRTATKAYNV TQAFGRRGPE QTQGNFGDQE LIRQGTDYKH

       310        320        330        340        350        360
WPQIAQFAPS ASAFFGMSRI GMEVTPSGTW LTYTGAIKLD DKDPNFKDQV ILLNKHIDAY

       370        380        390        400        410        420
KTFPPTEPKK DKKKKADETQ ALPQRQKKQQ TVTLLPAADL DDFSKQLQQS MSSADSTQAE

       430        440        450        460        470        480
AAAKLSPRWY FYYAAYGTTL PKGFYAAYDL SPRWYFYAAY SSPDDQIGYA AYLLNKHIDA

       490        500        510        520        530        540
YAAYSPDDQI GYYAAYDLSP RWYFYAAYNT ASWFTALAAY LLNKHIDAYG PGPGQIGYYR

       550        560        570        580        590        600
RATRRIRGGG PGPGDQIGYY RRATRRIRGK KNGPQNQRNA PRIKKFGGPS DSTGSNQNGE
```

Fig 13: Expasy server output based on a vaccine design.

To further illustrate, primary structure of antigen (N protein) was also undergone biochemical analysis in order to conduct a comparison with the final vaccine design; these findings are summarized above. This research indicates that both the vaccine construct and the primary antigens are unstable. The aliphatic index of the vaccine and crude N protein were 48.47 and 52.53, respectively. The measures of instability were 60.89 and 55.81, accordingly. The vaccine design and crude N protein have Grand Average of Hydropathicity (GRAVY) values of -1.110 and -0.980, respectively. Based on these data, it may be inferred that the vaccine's instability index is higher than that of crude proteins, and that, while both have an instability index of >40, both are unstable by molar mass. Finally, when comparing to the core protein, the vaccination has a lower GRAVY value. However, a vaccination with a negative GRAVY value that renders it hydrophilic, is preferred. During purification and downstream processing, proteins that are hydrophobic face an increased risk of contamination and functional loss. The use of ethanol and similar homologous solutions for purification of hydrophobic proteins should be noted since they can be dangerous to the health and the environment when handled improperly.

4.9: Vaccine Allergenicity Determination:

An autoimmune reaction in the target host is frequently induced by vaccines. This might lead to an unexpected increase in immunological response, which could lead to physiological complications and hazardous side effects. The Allergen Online server was utilized to evaluate the allergenic potency of the vaccine; based on z-score analysis, a cut-off value of 0.5 was chosen for an efficient assessment of the protein's efficacy, and the Full FASTA 36 approach was used in this

regard. Fortunately, no allergic homologues were discovered. The following was the input screen and output:
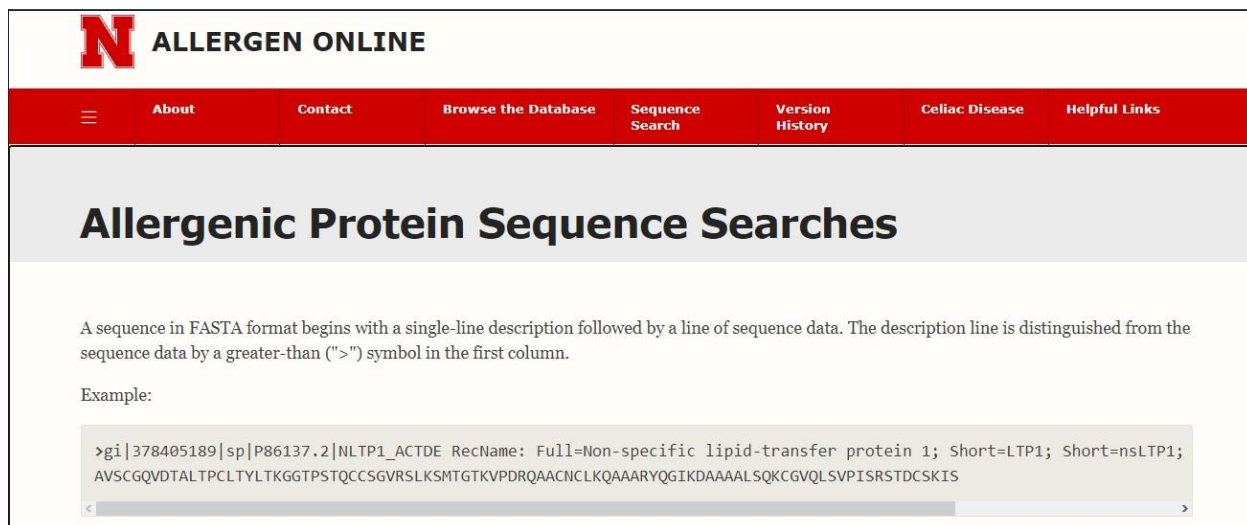
Fig 14: The vaccine construct's data is entered into the Allergen Online server.
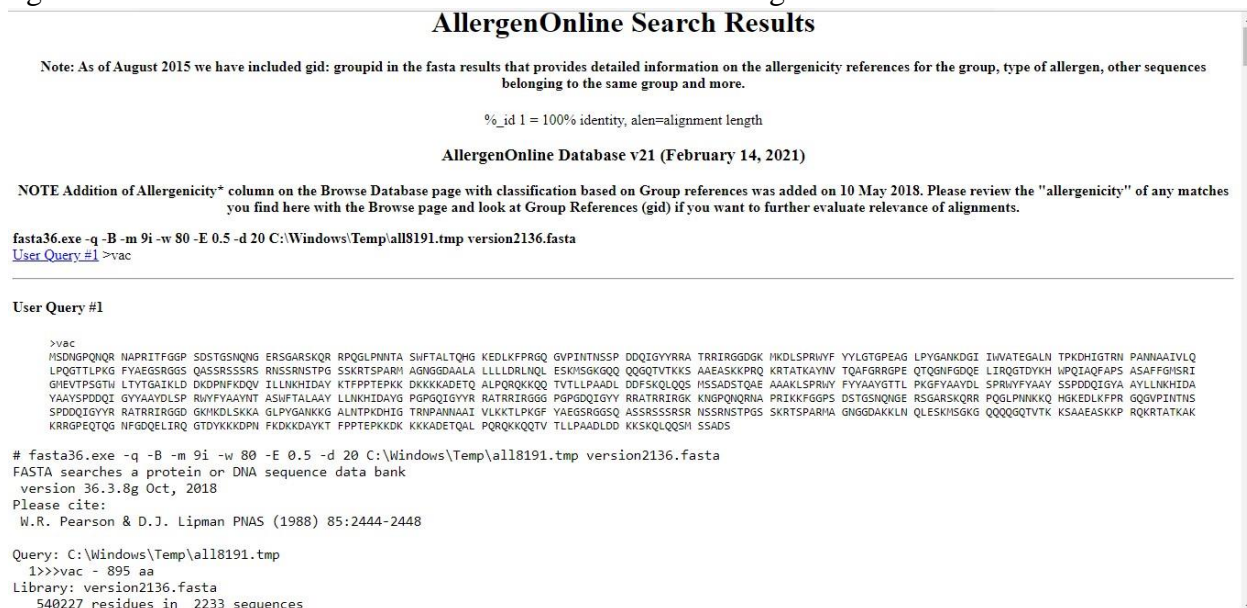
Fig 15: Allergen Online server output related to our vaccine design.

4.10: Vaccine's homology modelling:

Until yet, the vaccine possessed just a little structural integrity and lacked physical presence. It is critical to get a three-dimensional configuration of vaccine for subsequent investigation. This structure was derived in-silico using the vaccine's PDB file. This PDB file was generated utilizing homology modeling technique, with previously published PDB entries serving as a template. With 100 percent confidence, 157 residues (18% of the sequences) were simulated using the best scoring templates. The following is a three-dimensional picture of the vaccination that was modelled:
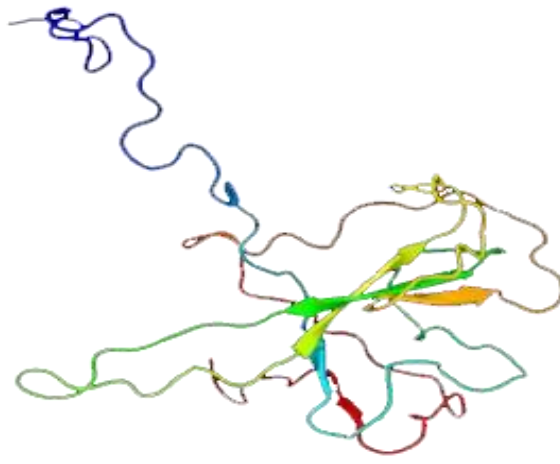
Image coloured by rainbow N → C terminus
Model dimensions (Å): **X**:58.748 **Y**:56.620 **Z**:54.587

Fig 16: Phyre2 server was used to create a 3D model of the vaccine.

In Å units, the model's X, Y, and Z dimensions are 58.748, 56.620, and 54.587, respectively.

4.11: Analysis of the Homologous vaccine model that was acquired:

The vaccine's PDB structure was examined further after receiving this from the phyre 2 server. ProSA web server was used to create a Z-score versus residue analysis graph for the Ramachandran plot analysis, using the SWISS PDB plotter.
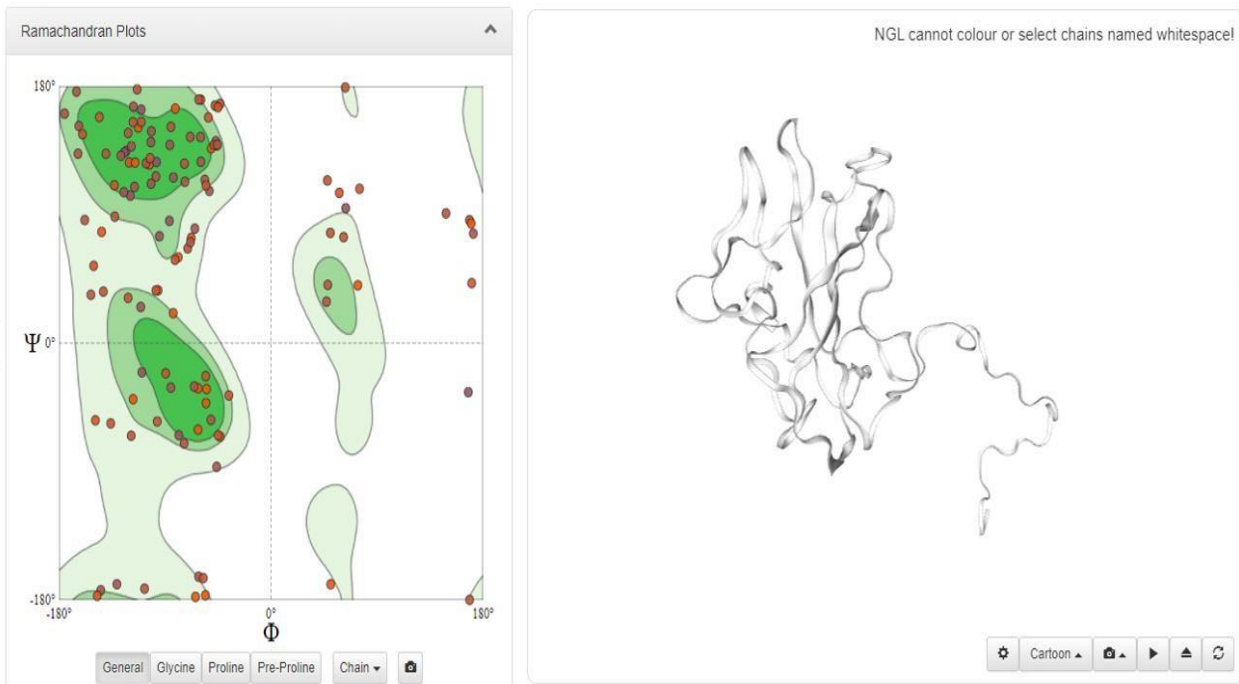
Figure 17: Ramachandran plot done using SWISS PDB plotter.

Interpretation of the plot is given below:

Fig 18: MolProbity results of Ramachandran plotting.

MolProbity score was 3.31; Clash score was 76.73; Ramachandran Favoured 53.21% and Ramachandran Outliers were 17.95%.
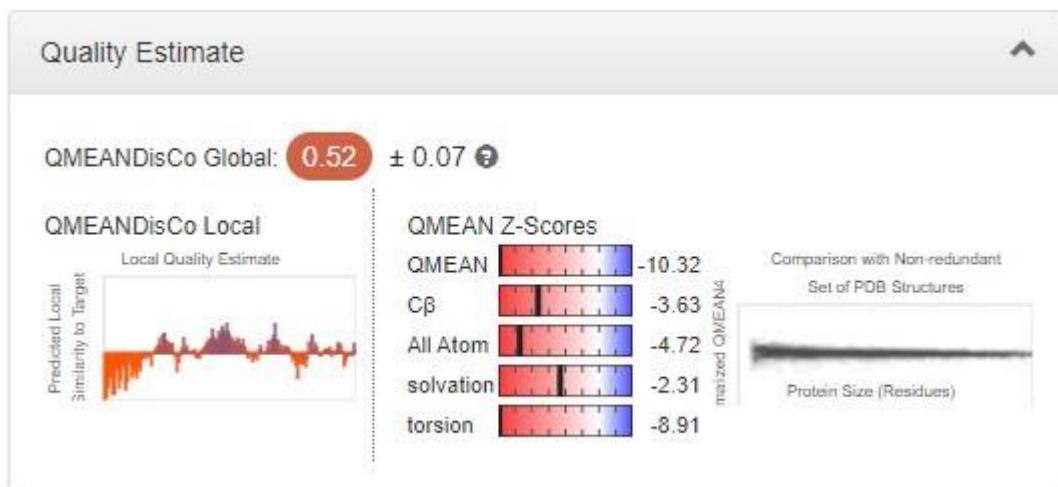
Fig 19: Quality Estimation for Ramachandran plotting.
Here, QMEAN is Co Global score which indicate the average per-residue score is 0.52.
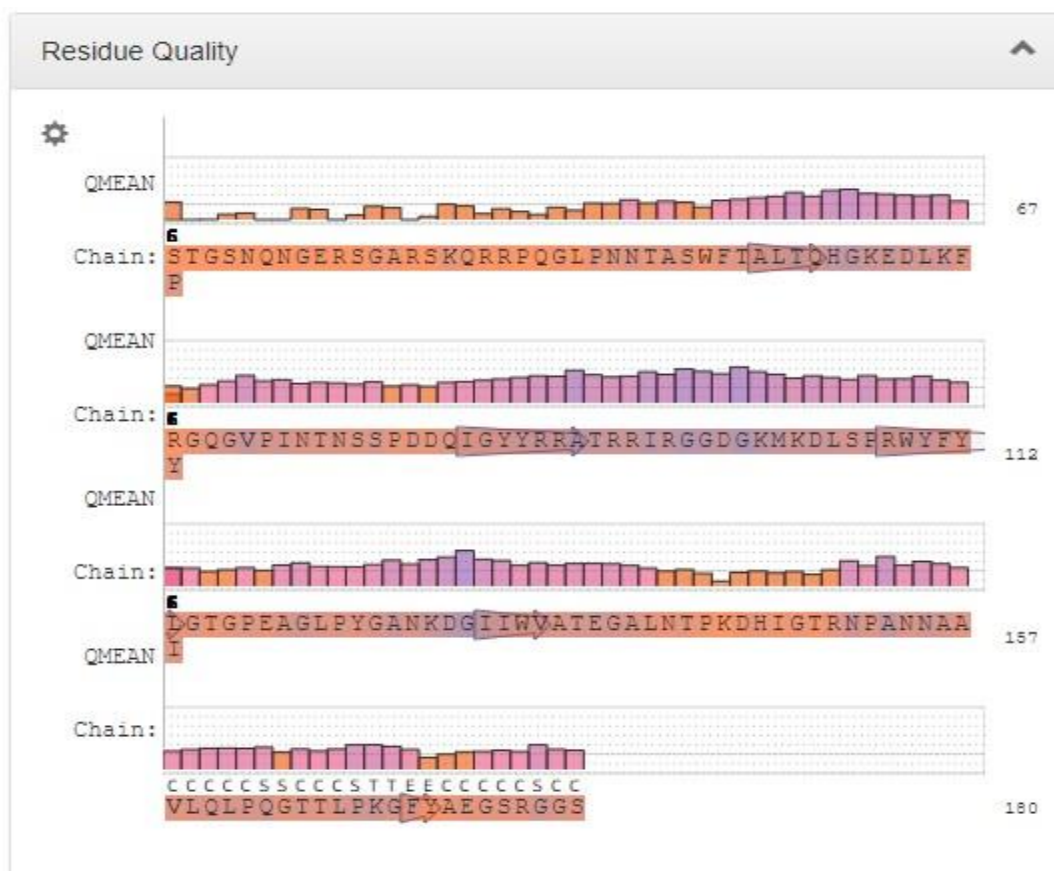
Fig 20: Residue Quality for Ramachandran plotting.

The z-score was found to be -5.11 in respect to the residue. As the value is less than 0, the raw score can be extrapolated to be below average. The plot is like this:
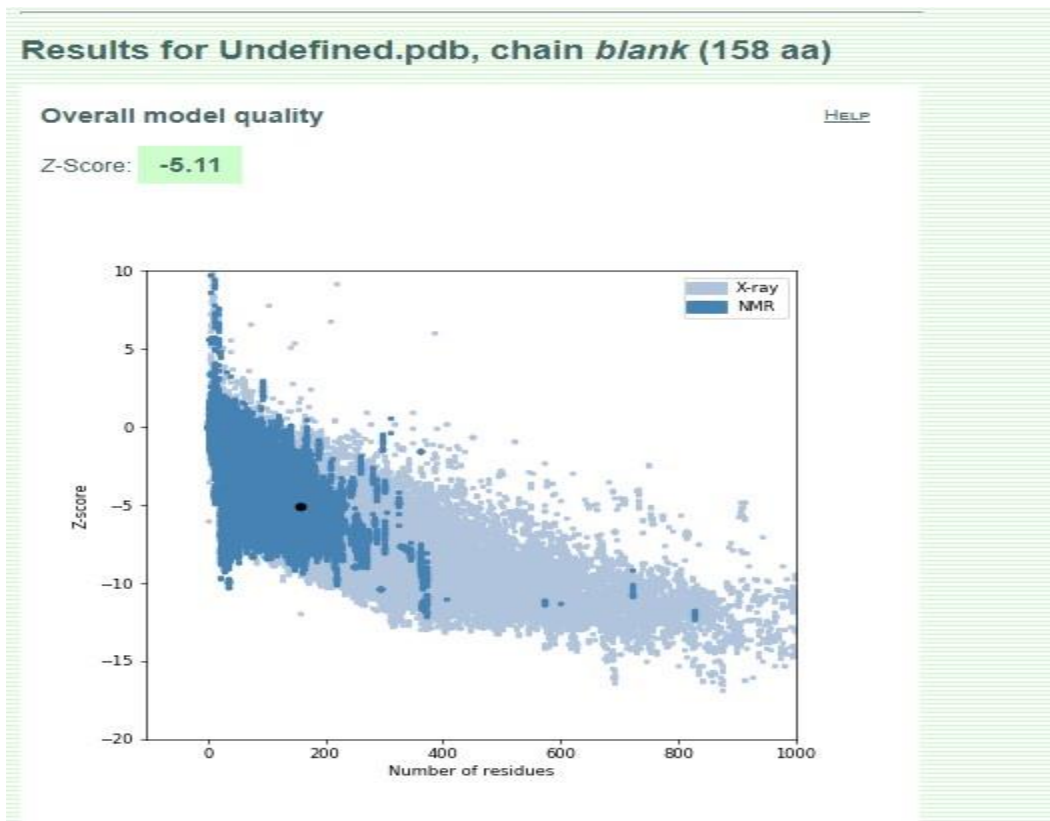
Fig 21: Figure showing the relationship between the Z-score and the residue.
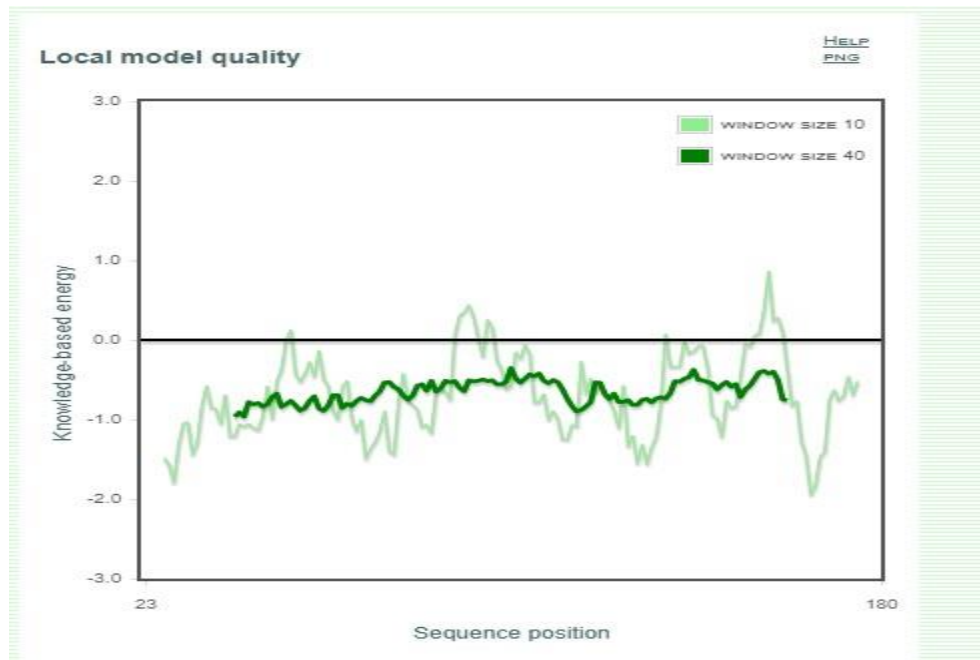
Fig 22: Local model quality for Knowledge-based energy versus Sequence position.

4.12: Analysis of docking results:

The Patchdock software was used to perform docking research between TLR8, a member of the TLR family, and another receptor. This receptor family is made up of protein receptors that are involved in the activation of the innate immune responses. Sentinel cells, which play a major role in the body's defense against invading pathogens, tend to overexpress these receptors as they are single-pass membrane-spanning receptor. TLRs are differentiated by the numbers on their labels. They are labeled with the numbers 1 to 13. TLR 11, 12, and 13 are among the receptors that are not found in humans. In this experiment, the ligand (a PDB file downloaded from the phyre2 service) was utilized as the ligand and TLR8 (PDB ID: 3W3G) was being used as a receptor. The Patchdock server was used for docking, which has a clustering RMSD of 4.0. At specific transformations, Patchdock executes docking and delivers a score for the docked complexes. As a further bonus, it determines the atom's docking area and the atomic contact energy. The following are the top 20 solutions that were initially monitored:

| Solution No	Score	Area	ACE	Transformation
1	18094	2671.30	464.89	-0.94 0.28 2.54 17.37 100.25 25.45
2	17994	2848.60	330.56	-1.26 -0.40 0.94 76.08 32.23 20.52
3	17898	2413.30	244.10	-0.87 0.39 2.70 3.65 101.38 23.51
4	17588	2745.80	281.08	-0.48 -0.25 1.78 61.45 9.71 -9.73
5	17578	2584.00	266.12	3.12 0.30 -0.08 63.47 36.34 68.46
6	17172	2890.00	165.25	-2.49 0.16 -0.39 6.35 41.71 124.36
7	16962	3132.80	345.30	-2.05 0.24 0.49 35.37 -38.47 86.13
8	16858	2559.70	386.27	2.22 -0.31 0.16 52.43 47.59 -2.72
9	16626	2979.50	256.11	2.51 -0.28 -0.53 79.42 35.38 62.32
10	16442	3091.70	495.17	2.01 0.52 2.69 -16.32 10.62 20.75
11	16256	2382.00	309.77	-0.50 0.04 0.56 100.30 6.16 27.51
12	16236	2185.00	374.62	3.06 0.32 0.10 57.28 44.30 66.73
13	16178	3144.80	343.96	-2.66 -0.64 0.06 80.54 26.15 97.71
14	16134	2765.20	469.03	-1.69 0.98 -0.09 8.17 -52.32 78.94
15	16102	2396.70	268.34	-0.43 -0.02 1.31 43.36 50.31 37.05
16	16042	2085.40	411.99	-0.38 -0.42 -1.05 -45.43 -12.76 -33.38
17	16030	3092.90	451.51	-2.32 0.22 -0.56 -3.38 30.33 124.81
18	16018	2373.70	281.89	-1.05 0.35 3.13 -22.20 95.54 32.94
19	15996	2494.80	451.78	-0.51 1.20 -2.88 -34.73 21.86 41.96
20	15850	2574.90	499.11	0.10 1.06 2.79 -33.13 29.47 21.13

Fig 23: Shape Complementarity Principles-Based Molecular Docking Algorithm

According to the obtained results, it is clear that, among the 20 different transformations, the transformation (-0.94,0.28,2.54,17.37,100.25,25.45) has given the highest score of 18094 and the ACE value 464.89 KJmol-1, and covered an area of 2671.30 square angstrom, indicating It is most likely the best combination formed in between TLR8 receptor as well as the chosen vaccination model. The derived protein-ligand complex's PDB structure was examined using the 64-bit client edition of Discovery Studio 2016. The complex docked will be useful for future research, specially molecular dynamics (MD) simulations that are not included in this paper.
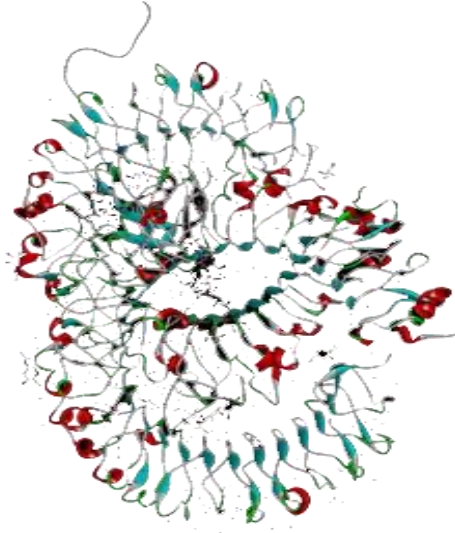
Fig 24: 3D structure of a docked complex containing a simulated vaccination and the TLR8 receptor.

4.13: In-silico immune-simulation Results:

In-silico simulation of the planned vaccine is a critical step, especially in terms of vaccine quality and performance. Immuno simulation was performed using the server of C-immsim and the vaccine's FASTA sequence as input. Many graphs were included in the results, demonstrating the immunoglobulins, concentration of antigen, and immune-cellular epitopes throughout time. The resulting graphs are displayed below, along with their interpretations:
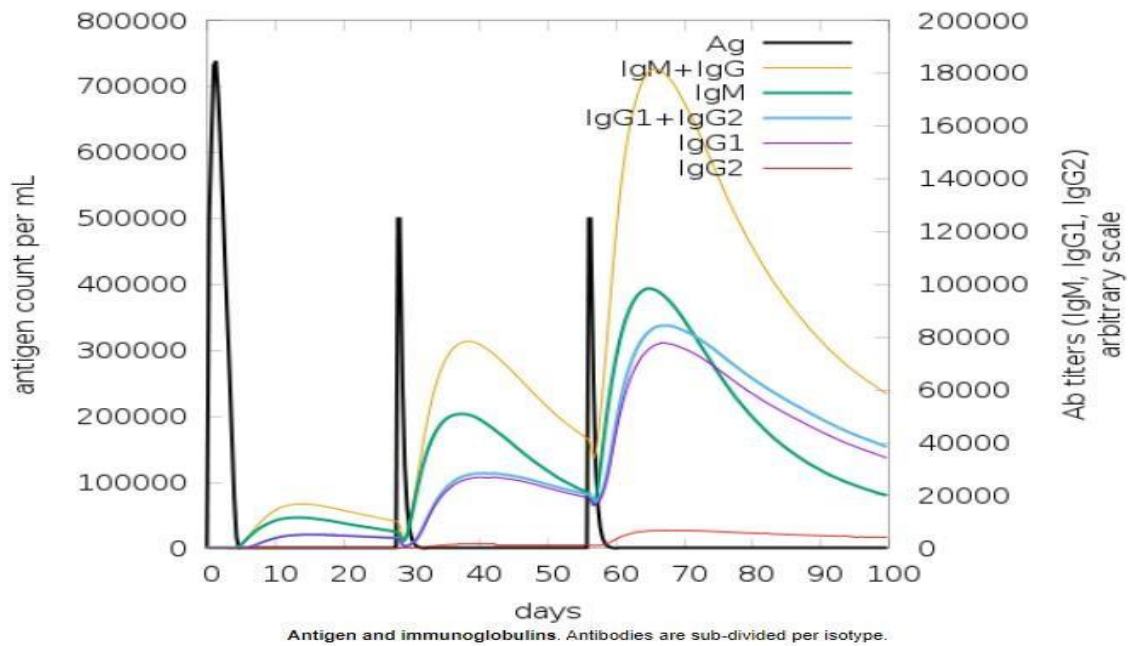
Fig 25: Ag concentration vs time graph.

After vaccination, the C-immsim software calculates B cell lymphocyte concentration. B cell epitopes are involved in both humoral and cellular immunity. Memory B cells also help to boost ag specific immunity, which protects the host against subsequent pathogen infections. B cell concentrations are determined by measuring igM, igG1 and igG2 concentrations. Here, graphs are representing state-by-state concentrations of the B cell population. At long last, plasma B cells were discovered. These cells may be useful as therapeutic agents. The following graphs show the B cell status:
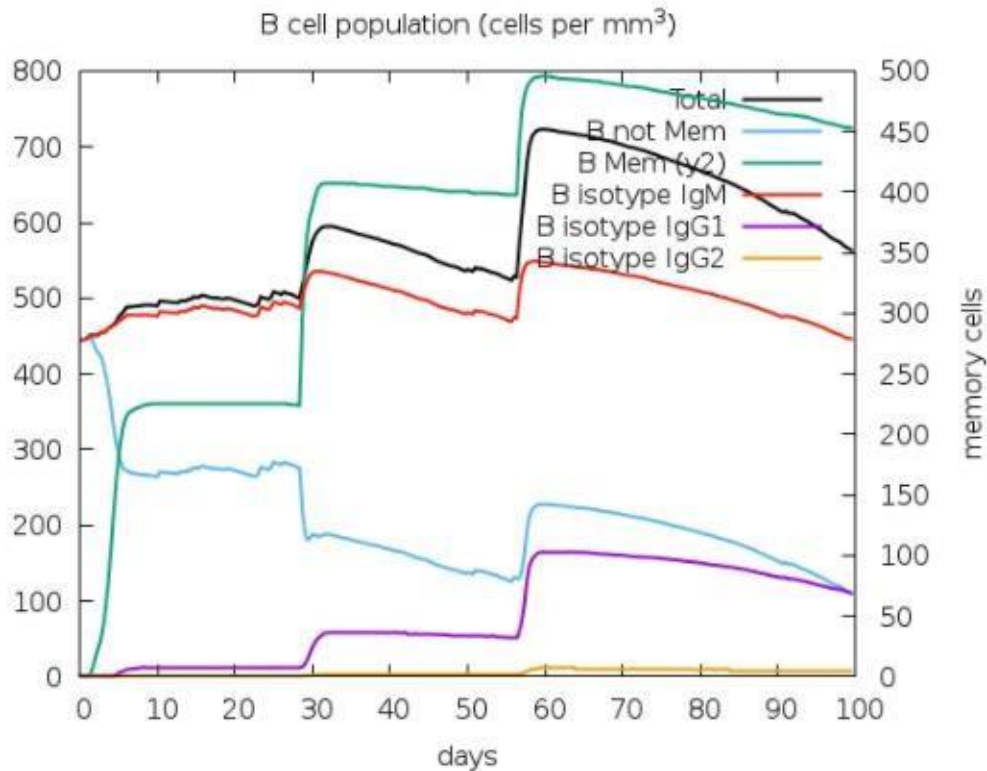
B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM, IgG1 and IgG2.

Fig 26: Diagram demonstrating the concentration of B cells according to subtypes in relation to the number of days following vaccination injection.

B cell population per state (cells per mm³)
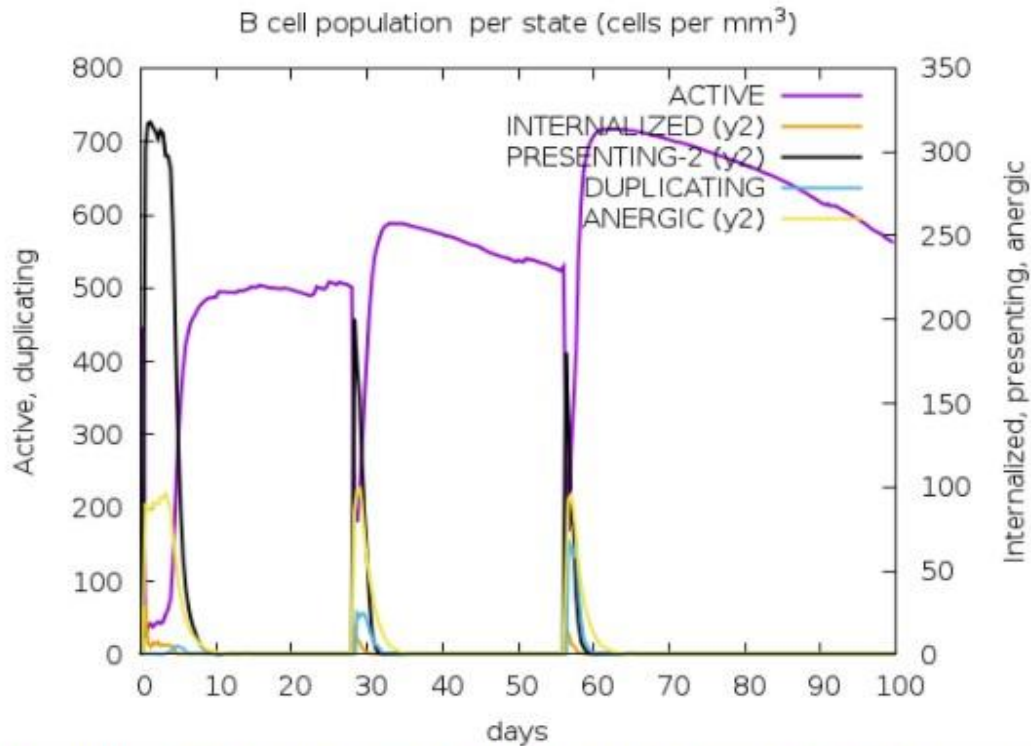
B lymphocytes population per entity-state (i.e., showing counts for active, presenting on class-II, internalized the Ag, duplicating and anergic.

Fig 27: State-by-state breakdown of B cell concentration dependent on the number of days after vaccination.

PLB cell population (cells per mm³)
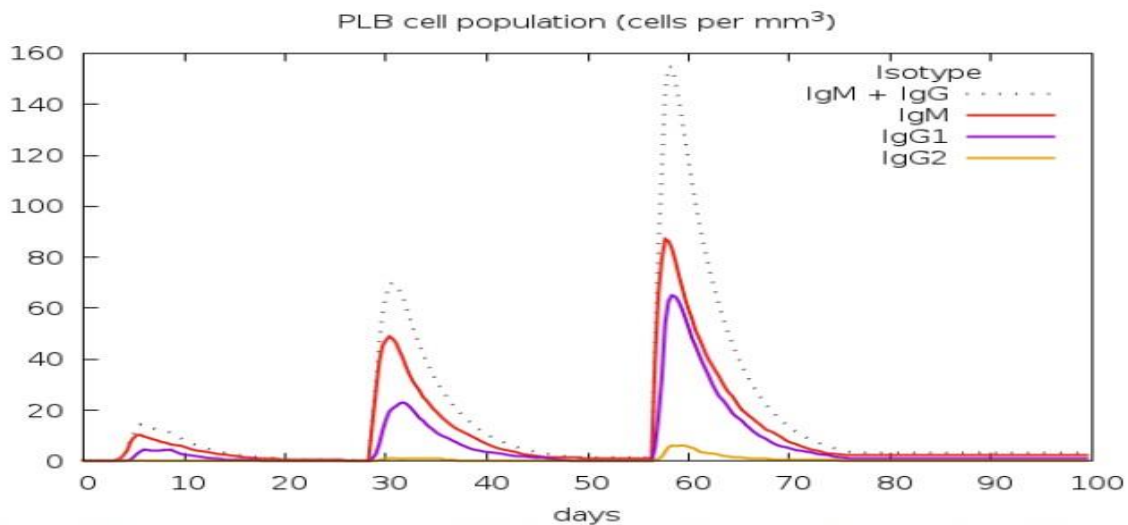
Plasma B lymphocytes count sub-divided per isotype (IgM, IgG1 and IgG2).

Fig 28: B cell population growth in plasma (by isotype) vs days (post vaccination state)

A similar method was used to get graphs showing the concentration of HTL and CTL epitopes from the server. Graphs depicting CTL and HTL concentrations are as follows:
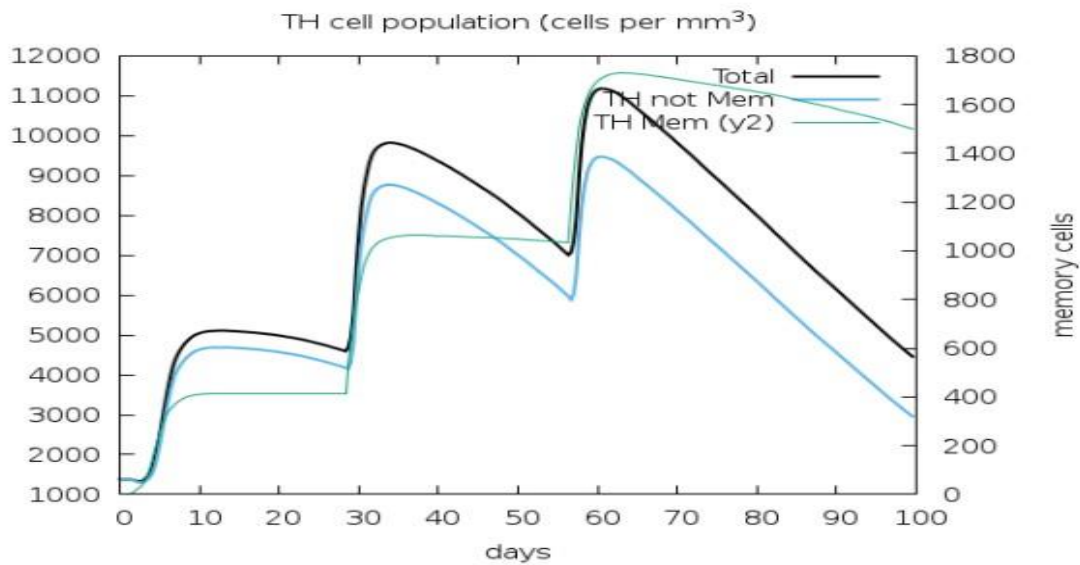
Fig 29: CD-4+ HTL epitopes, including and excluding memory and normal T cells, were detected days after vaccination delivery, inclusive and non-inclusive of B cells and normal T cells.

Fig 30: The CD4+ HTL epitope level in each condition is measured and plotted daily after vaccination.
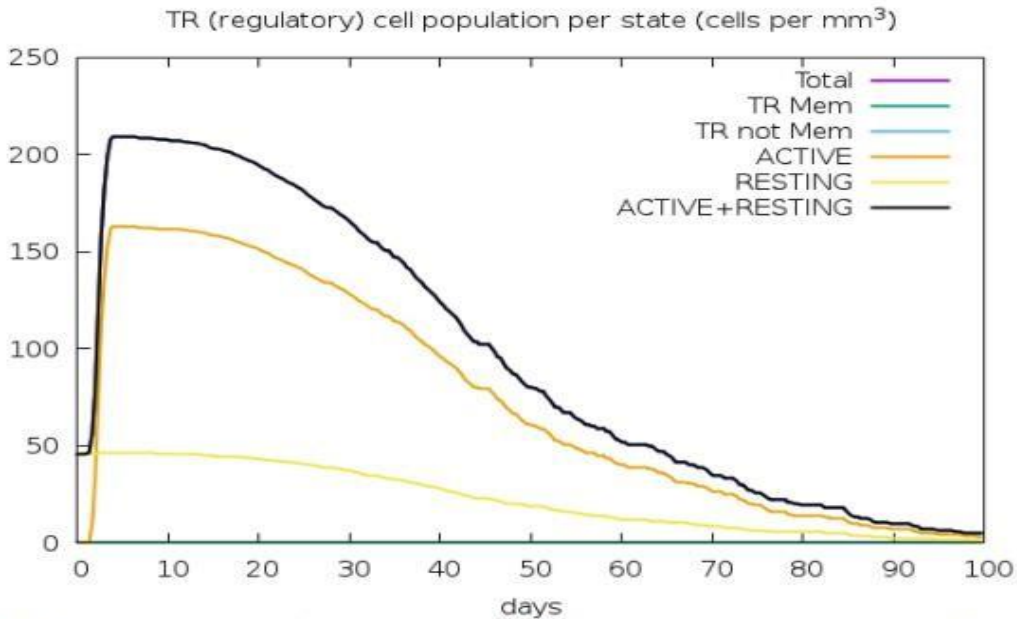
CD4 T-regulatory lymphocytes count. Both total, memory and per entity-state counts are plotted here.

Fig 31: Concentration of the regulatory T-cell lymphocyte population (CD-4+) vs time.

The CD8+ CTL epitope concentration was increased in both memory-inducing and non-memoryinducing vaccines. Additionally, the following states of CTL epitope, as well as their concentrations after vaccination, were plotted:
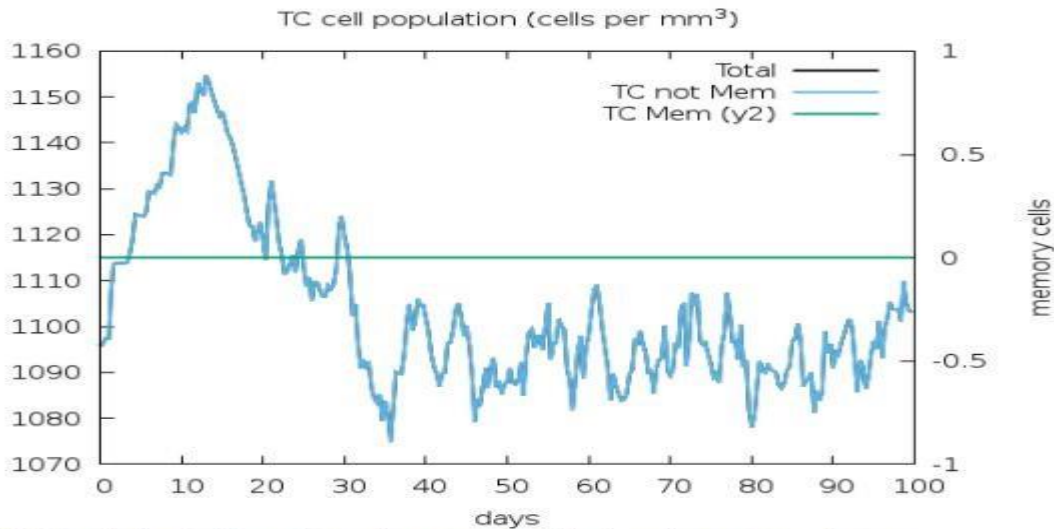
CD8 T-cytotoxic lymphocytes count. Total and memory shown.

Fig 32: Memory, non-memory, and total CTL concentration development over time (days postvaccination)
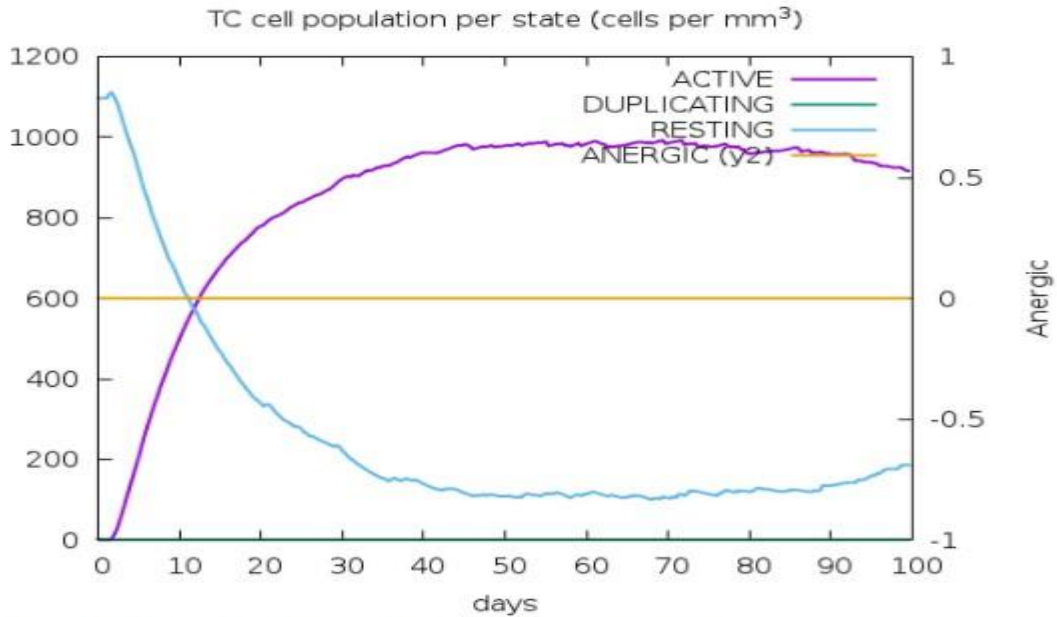
CD8 T-cytotoxic lymphocytes count per entity-state.

Fig 33: Observation of CTL growth by subtype in relation to the number of days from vaccination.

Finally, the host's WBC count after vaccination was monitored by the C-Immsim server, including NK cells, DC, macrophages, and epithelial cells. The graphs are as follows:
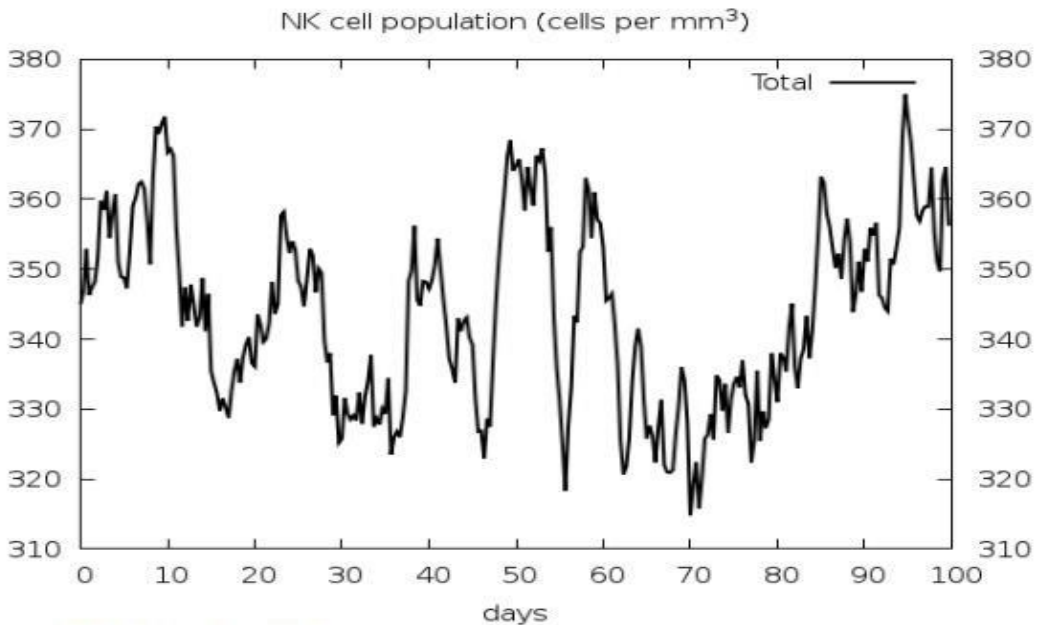
Natural Killer cells (total count).

Fig 34: The day-to-day rise of the overall NK cell population after immunization was observed.

DC population per state (cells per mm³)
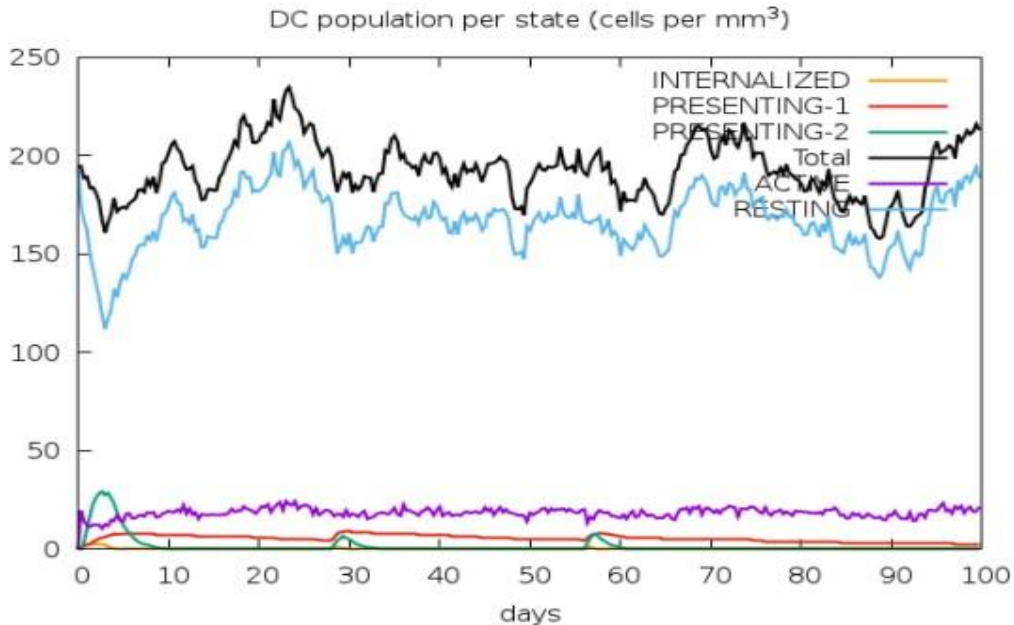
Dendritic cells. DC can present antigenic peptides on both MHC class-I and class-II molecules. The curves show the total number broken down to active, resting, internalized and presenting the ag.

Fig 35: MHC I and MHC II molecules may be used by the DC population to display ag. Based on their active resting and internalized states, cells were depicted in more detail.

MA population per state (cells per mm³)
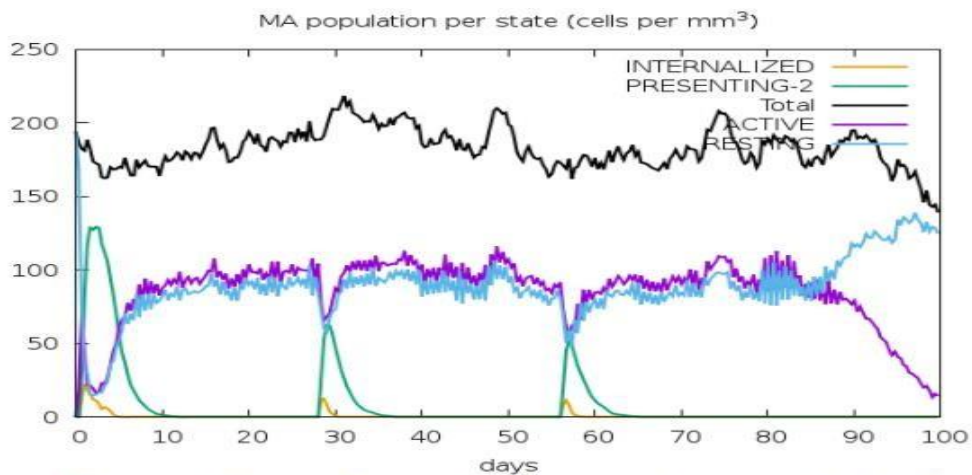
Macrophages. Total count, internalized, presenting on MHC class-II, active and resting macrophages.

Fig 36: Macrophage population expansion measured as active, resting, MHC II presenting, and internalized groups vs vaccination delivery days.

Finally, to depict the proliferation of epithelial cells and cytokines after immunization, graphs were drawn up. Other substances that cause inflammation in the host include interferons and interleukins are appropriate sites for spreading viral infection. The results are shown below:

54
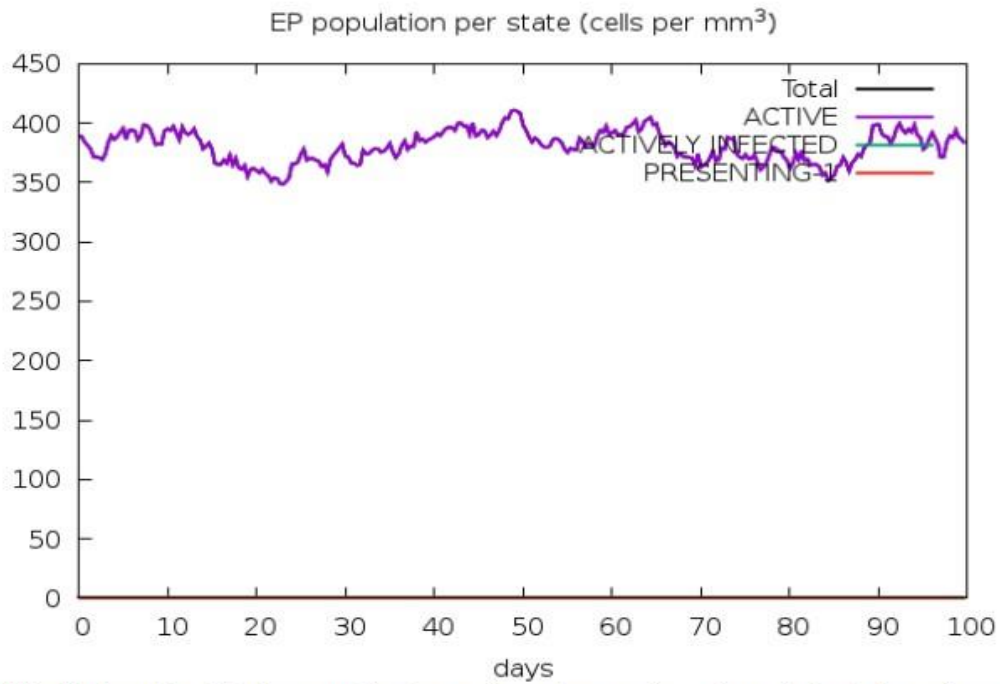
Epithelial cells. Total count broken down to acgive, virus-infected and presenting on class-I MHC molecule.

Fig 37: Comparing pre- and post-vaccination levels of total active MHC I-presented and virally infected cells.
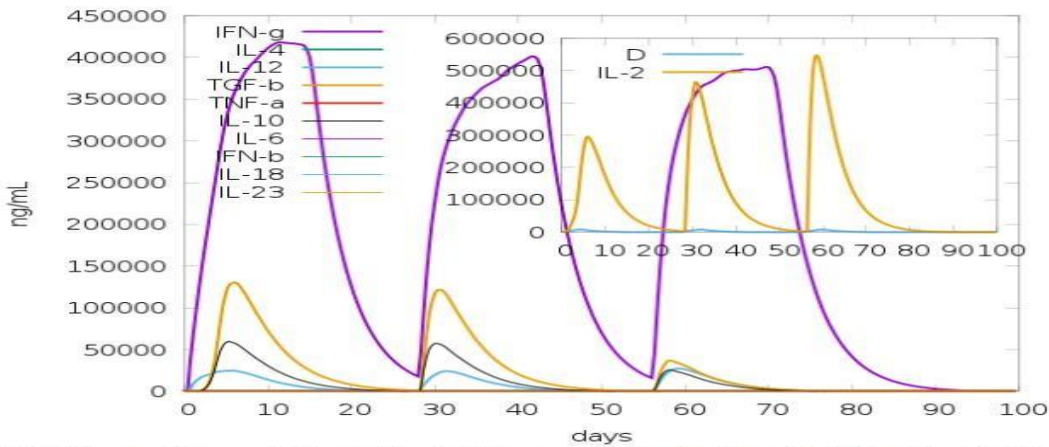
Cytokines. Concentration of cytokines and interleukins. D in the inset plot is danger signal.

Fig 38: The inset shows a graph of the rise of the cytokine population in relation to the number of days.

4.14: Results of in silico simulations:

In this section, an effort is made to understand the findings of the C-immsim software. In-silico simulation analysis is an important step in immunoinformatic vaccinomics as it offers a potentially verified foundation for the research. According to the graphs, after a few days of vaccination administration, the vaccine elicited improved ab-mediated immunity in the 678 host in the form of igM, igG. The server classifies the vaccination as a conventional ag. The immune response of the host is triggered by antibodies produced as a result of the immune stimulation provided by the vaccine. When the vaccine is administered to the 681 host, the concentration of SARS-CoV-2 specific ag, in this instance N protein, as well as the possible epitopes utilized during its creation, increases first. The ag concentration steadily declines over time as a result of the production of SARS-CoV-2-specific ab, which is the major focus of vaccine research. As a consequence of a possible trade-off with ag, ab unique to a given ag develops in a host. Every time a vaccine is administered, ag is neutralized by ab. As previously stated, the vaccine was created using a three-dosage approach. As a result, when each dosage was administered, an antigenic spike was noticed in the host, which was eventually reduced by antibodies generated against the SARS-CoV-2 N protein. Immunoglobulins such as igG and igM, as well as plasma B cells, increased in number in the bloodstream of the patients. Hosts who develop memory B cells may avoid subsequent contacts with ag. Potential antiviral agents derived from the plasma B cells of an individual's blood type are presently being considered According to the graphs, a B-cellmediated host response was stimulated by the vaccination. CD4+ HTL and regulatory T cell lymphocytic epitopes, which are critical for developing immunological tolerance in the host, were discovered. This concentration decreases over time but it does not affect immunity. A similar pattern had been seen in CD8+ cells, specifically in CTL epitopes. Graph that shows WBC concentration of different types demonstrated highly efficient vaccine activity. i.e., DC cells, macrophages, etc., as well as a decrease in viral infections. Finally, the growth of cytokines was shown to be increasing. A key threshold of cytokine development was also identified in the plot, which must be monitored since excessive inflammation have the potential to be hazardous. While interleukins and interferons were growing on the cytokine curve, this indicated that the vaccine was successful in protecting against the SARS-CoV-2 virus.

Chapter 5

Discussion & Research Prospects

An in-silico evaluation of a multiepitope based peptide vaccine against SARS-CoV-2 virus targeting its Nucleocapsid protein was attempted in this study with SARS-CoV-2 viral virulence and host assault increasing exponentially, preventive measures are the best option at this point in the pandemic's life-cycle.

In-silico methods have many advantages over traditional methods in the field of vaccineomics. This method can predict vaccine efficacy and side effects even before human trials or in-vitro investigations. Despite this, in silico approaches might not be adequate for vaccine development. Rather, it is a tool to help the development process of effective vaccines. This research also modestly state that the developed peptide vaccine is a credible attempt to prevent SARS-CoV-2. Although this vaccine has still not been physically manufactured, this study suggests that it should be provided to healthy persons first, and subsequently to those with minor symptoms if necessary.

This research, on the other hand, do not claim that the vaccine should be administered to patients who are severely impacted by COVID-19. Such individuals already have an unusual high level of immunological reaction within their body that is why they are considered at high-risk. Regarding these facts, this research tried to develop an in-silico based multiepitope subunit peptide vaccine against SARS-CoV-2 using computational biology. After all the necessary efforts had given, this study was able to develop a candidate capable of combatting SARS-CoV-2. As the final vaccine appears unstable in nature, chaperones can be added to ensure the stability of unfolded polypeptide chains when they enter subcellular organelles. Molecular chaperones are a class of protein that binds with various protein substrates to help in their folding, preventing misfolded or otherwise damaged molecules from developing. By stabilizing partially or completely unfolded protein polypeptides, chaperones prevent aggregation and improper folding (Blagosklonny, 2001; Bolhassani & Rafati, 2013; Corigliano et al., 2021).

References:

Abraham Peele, K., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–9. https://doi.org/10.1080/07391102.2020.1770127

Akhand, M. R. N., Azim, K. F., Hoque, S. F., Moli, M. A., Joy, B. Das, Akter, H., Afif, I. K., Ahmed, N., & Hasan, M. (2020). Genome based evolutionary lineage of SARS-CoV-2 towards the development of novel chimeric vaccine. *Infection, Genetics and Evolution*, *85*. https://doi.org/10.1016/j.meegid.2020.104517

Alazmi, M., & Motwalli, O. (2021). In silico virtual screening, characterization, docking and molecular dynamics studies of crucial SARS-CoV-2 proteins. *Journal of Biomolecular Structure and Dynamics*, *39*(17), 6761–6771. https://doi.org/10.1080/07391102.2020.1803965

Barreto-Vieira, D. F., da Silva, M. A. N., Garcia, C. C., Miranda, M. D., Matos, A. da R., Caetano, B. C., Resende, P. C., Motta, F. C., Siqueira, M. M., Girard-Dias, W., Archanjo, B. S., & Barth, O. M. (2021). Morphology and morphogenesis of sars-cov-2 in vero-e6 cells. *Memorias Do Instituto Oswaldo Cruz*, *116*(1), 1–6. https://doi.org/10.1590/007402760200443

Blagosklonny, M. V. (2001). Re: Role of the heat shock response and molecular chaperones in oncogenesis and cell death [4]. *Journal of the National Cancer Institute*, *93*(3), 239–240. https://doi.org/10.1093/jnci/93.3.239-a

Bolhassani, A., & Rafati, S. (2013). Mini-chaperones: Potential immuno-stimulators in vaccine design. *Human Vaccines and Immunotherapeutics*, *9*(1), 153–161. https://doi.org/10.4161/hv.22248

Castiglione, F., Deb, D., Srivastava, A. P., Liò, P., & Liso, A. (2021). From Infection to Immunity:

Understanding the Response to SARS-CoV2 Through In-Silico Modeling. *Frontiers in Immunology*, *12*(September), 1–16. https://doi.org/10.3389/fimmu.2021.646972

Chakravarty, M., & Vora, A. (2021). Nanotechnology-based antiviral therapeutics. *Drug Delivery and Translational Research*, *11*(3), 748–787. https://doi.org/10.1007/s13346-020-00818-0

Chen, H. Z., Tang, L. L., Yu, X. L., Zhou, J., Chang, Y. F., & Wu, X. (2020). Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. *Infectious Diseases of Poverty*, *9*(1), 1–10. https://doi.org/10.1186/s40249-020-00713-3

Chukwudozie, O. S., Gray, C. M., Fagbayi, T. A., Chukwuanukwu, R. C., Oyebanji, V. O., Bankole, T. T., Adewole, R. A., & Daniel, E. M. (2021). Immuno-informatics design of a multimeric epitope peptide based vaccine targeting SARS-CoV-2 spike glycoprotein. *PLoS ONE*, *16*(3 March), 1–25. https://doi.org/10.1371/journal.pone.0248061

Corigliano, M. G., Sander, V. A., Sánchez López, E. F., Ramos Duarte, V. A., Mendoza Morales, L. F., Angel, S. O., & Clemente, M. (2021). Heat Shock Proteins 90 kDa: Immunomodulators and Adjuvants in Vaccine Design Against Infectious Diseases. *Frontiers in Bioengineering and Biotechnology*, *8*(January). https://doi.org/10.3389/fbioe.2020.622186

Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, *17*(3), 181–192. https://doi.org/10.1038/s41579-018-0118-9

Dash, P., Turuk, J., Behera, S. K., Palo, S. K., Raghav, S. K., Ghosh, A., Sabat, J., Rath, S., Subhadra, S., Rana, K., Bhattacharya, D., Kanungo, S., Kshatri, J. S., Mishra, B. K., Dash, S., Parida, A., & Pati, S. (2021). Sequence analysis of Indian SARS-CoV-2 isolates shows a stronger interaction of mutant receptor-binding domain with ACE2. *International Journal of Infectious Diseases*, *104*, 491–500. https://doi.org/10.1016/j.ijid.2021.01.020

Dimitrov, I., Garnev, P., Flower, D. R., & Doytchinova, I. (2010). MHC class II binding predictionA little help from a friend. *Journal of Biomedicine and Biotechnology*, *2010*. https://doi.org/10.1155/2010/705821

Dutta, N. K., Mazumdar, K., & Gordy, J. T. (2020). The Nucleocapsid Protein of SARS–CoV-2: a Target for Vaccine Development. *Journal of Virology*, *94*(13), 1–2. https://doi.org/10.1128/jvi.00647-20

Gao, T., Gao, Y., Liu, X., Nie, Z., Sun, H., Lin, K., Peng, H., & Wang, S. (2021). Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC Microbiology*, *21*(1), 1–10. https://doi.org/10.1186/s12866-021-02107-3

Grudlewska-Buda, K., Wiktorczyk-Kapischke, N., Wałecka-Zacharska, E., Kwiecińska-Piróg, J., Buszko, K., Leis, K., Juszczuk, K., Gospodarek-Komkowska, E., & Skowron, K. (2021). SARS-CoV-2—Morphology, Transmission and Diagnosis during Pandemic, Review with Element of Meta-Analysis. *Journal of Clinical Medicine*, *10*(9), 1962. https://doi.org/10.3390/jcm10091962

Harrison, A. G., Lin, T., & Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in Immunology*, *41*(12), 1100–1115. https://doi.org/10.1016/j.it.2020.10.004

He, J., Huang, F., Zhang, J., Chen, Q., Zheng, Z., Zhou, Q., Chen, D., Li, J., & Chen, J. (2021). Vaccine design based on 16 epitopes of SARS-CoV-2 spike protein. *Journal of Medical Virology*, *93*(4), 2115–2131. https://doi.org/10.1002/jmv.26596 Henry, R., & T. (2019). *Strain Infecting Cattle Tick*. 30329.

Hoque, H., Islam, R., Ghosh, S., Rahaman, M. M., Jewel, N. A., & Miah, M. A. (2021). Implementation of in silico methods to predict common epitopes for vaccine development against Chikungunya and Mayaro viruses. *Heliyon*, *7*(3), e06396. https://doi.org/10.1016/j.heliyon.2021.e06396

Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., & Whittaker, G. R. (2020). Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *Journal of Molecular Biology*, *432*(10), 3309–3325. https://doi.org/10.1016/j.jmb.2020.04.009

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29. https://doi.org/10.1093/nar/gkx346

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, *10*(1), 1–24. https://doi.org/10.1038/s41598-020-67749-1

Khairkhah, N., Aghasadeghi, M. R., Namvar, A., & Bolhassani, A. (2020). Design of novel multiepitope constructs-based peptide vaccine against the structural S, N and M proteins of human COVID-19 using immunoinformatics analysis. *PLoS ONE*, *15*(10 October), 1–28. https://doi.org/10.1371/journal.pone.0240577

Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., & Chlanda, P. (2020). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-19619-7

Kleine-Tebbe, J., & Mailänder, C. (2020). Patterns of allergen sensitization in patients with severe asthma in Germany. *Journal of Allergy and Clinical Immunology: In Practice*, *8*(2), 744746.e3. https://doi.org/10.1016/j.jaip.2019.11.038

Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M., & Bergeron, J. R. C. (2020). Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Frontiers in Molecular Biosciences*, *7*. https://doi.org/10.3389/fmolb.2020.605236

Mittal, A., Manjunath, K., Ranjan, R. K., Kaushik, S., Kumar, S., & Verma, V. (2020). COVID19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARSCoV-2. *PLoS Pathogens*, *16*(8), e1008762. https://doi.org/10.1371/journal.ppat.1008762

Nain, N., Surabhi, R., Yathish, N. V., Krishnamurthy, V., Deepa, T., & Tharannum, S. (2019). Enhancement in strength parameters of concrete by application of Bacillus bacteria. *Construction and Building Materials*, *202*, 904–908. https://doi.org/10.1016/j.conbuildmat.2019.01.059

Phan, I. Q., Subramanian, S., Kim, D., Murphy, M., Pettie, D., Carter, L., Anishchenko, I., Barrett, L. K., Craig, J., Tillery, L., Shek, R., Harrington, W. E., Koelle, D. M., Wald, A., Veesler,

D., King, N., Boonyaratanakornkit, J., Isoherranen, N., Greninger, A. L., … Van Voorhis, W. C. (2021). In silico detection of SARS-CoV-2 specific B-cell epitopes and validation in ELISA for serological diagnosis of COVID-19. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-83730-y

Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., Gu, Z., Kumar, S., Zaremba, S., Larsen, C. N., Jen, W., Klem, E. B., & Scheuermann, R. H. (2012). Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. *Viruses*, *4*(11), 3209–3226. https://doi.org/10.3390/v4113209

Pollett, S., Conte, M. A., Sanborn, M., Jarman, R. G., Lidl, G. M., Modjarrad, K., & Maljkovic Berry, I. (2021). A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-96626-8

Poran, A., Harjanto, D., Malloy, M., Arieta, C. M., Rothenberg, D. A., Lenkala, D., Van Buuren, M. M., Addona, T. A., Rooney, M. S., Srinivasan, L., & Gaynor, R. B. (2020). Sequencebased prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes. *Genome Medicine*, *12*(1), 1–15. https://doi.org/10.1186/s13073-020-00767-w

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2021). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1), W449–W454. https://doi.org/10.1093/NAR/GKAA379

Sarkar, M., & Saha, S. (2020). Structural insight into the role of novel SARSCoV-2 E protein: A potential target for vaccine development and other therapeutic strategies. *PLoS ONE*, *15*(8 August), 1–25. https://doi.org/10.1371/journal.pone.0237300

Sayed, S. Bin, Nain, Z., Abdullah, F., Haque, Z., Rahman, S. M. R., Tasmin, R., & Adhikari, U. K. (2019). Immunoinformatics-guided Designing of Peptide Vaccine against Lassa Virus with Dynamic and Immune Simulation Studies. *Preprints.Org*, *September*. https://doi.org/10.20944/preprints201909.0076.v2

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(21), 1–8. https://doi.org/10.1073/pnas.2003138117

Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., Asa, B. F., Ngale, F. N., Vanhamme, L., & Souopgui, J. (2019). In-silico design of a multiepitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, *9*(1), 1–18. https://doi.org/10.1038/s41598-019-40833-x

Shu, C., Huang, X., Brosius, J., & Deng, C. (2020). Exploring potential super infection in SARSCoV2 by genome-wide analysis and receptor – ligand docking. *Preprints*, *March*, 2020030310.

Sohail, M. S., Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2021). In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. *Advanced Drug Delivery Reviews*, *171*, 29–47. https://doi.org/10.1016/j.addr.2021.01.007

Trougakos, I. P., Stamatelopoulos, K., Terpos, E., Tsitsilonis, O. E., Aivalioti, E., Paraskevis, D., Kastritis, E., Pavlakis, G. N., & Dimopoulos, M. A. (2021). Insights to SARS-CoV-2 life cycle, pathophysiology, and rationalized treatments that target COVID-19 clinical complications. *Journal of Biomedical Science*, *28*(1), 1–18. https://doi.org/10.1186/s12929020-00703-5

Usmani, S. S., Kumar, R., Bhalla, S., Kumar, V., & Raghava, G. P. S. (2018). In Silico Tools and Databases for Designing Peptide-Based Vaccine and Drugs. In *Advances in Protein Chemistry and Structural Biology* (1st ed., Vol. 112). Elsevier Inc. https://doi.org/10.1016/bs.apcsb.2018.01.006

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Wang, M., Xiong, H., Chen, H., Li, Q., & Ruan, X. Z. (2020). Renal Injury by SARS-CoV-2 Infection: A Systematic Review. *Kidney Diseases*, 1–11. https://doi.org/10.1159/000512683

Woo, P. C. Y., Lau, S. K. P., Huang, Y., & Yuen, K. Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine*, *234*(10), 1117–1127. https://doi.org/10.3181/0903-MR-94

Yang, Z., Bogdan, P., & Nazarian, S. (2021). An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Scientific Reports*, *11*(1), 1–21. https://doi.org/10.1038/s41598-021-81749-9