An In Silico Method to Developing an Epitope-based Peptide Vaccination Against SARS-CoV-2's Envelope Protein (E)

By
Mohammad Nafees Intesar
ID: 13346017

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

School of Pharmacy

Brac University

June 2022

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

_____

Mohammad Nafees Intesar

Student ID: 13346017

Approval

The thesis titled 'Design of an epitope-based peptide vaccine against Envelope protein (E) of SARS-CoV-2: an in-silico approach' submitted by Mohammad Nafees Intesar (ID: 13346017) of Spring 2021 has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Bachelor of Pharmacy (Hons) on June 09, 2022.

Examining Committee:

Supervisor:
(Member)

Mohammad Kawsar Sharif Siam
Senior Lecturer, School of Pharmacy
Brac University

Program Coordinator:
(Member)

Namara Mariam Chowdhury
Lecturer, School of Pharmacy
Brac University

Deputy Chairperson:
(Member)

Hasina Yasmin, PhD
Professor & Deputy Chairperson, School of Pharmacy
Brac University

Departmental Head:
(Dean)

Eva Rahman Kabir, PhD
Professor & Dean, School 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

The new coronavirus (SARS-CoV-2) pandemic, which has killed millions of people throughout the world, has afflicted millions of people. SARS-CoV-2 therapies were severely limited due to the virus's quick pathogenicity. As a result, immunizations were desperately needed because there were no effective medical therapies. Immunoinformatic approaches were employed in this work to develop a multi-epitope vaccine that has the potential to activate the body's immune system against SARS-CoV-2. The viral structural protein was screened for the first group of epitopes. VaxiJen v2.0, AllerTOP v2.0, and ToxinPred were used to identify probable antigenic, non-toxic, and non-allergenic T-cell and B-cell epitopes, and a projected model was developed. IFNepitope, IL4pred, and IL10pred were used to test cytokine inducing epitopes. One MHC I binding cytotoxic T lymphocyte (CTL) (9-mer) and one MHC II binding helper T lymphocyte (HTL) (9-mer) were tested for T-cell, as both have significant binding affinity and are antigenic, with scores of 0.7476 and 0.5993, respectively. Interferon-gamma, interleukin-4, and interleukin-10 were all induced by the HTL epitope. The chosen B-cell epitope was non-toxic and non-allergenic, with a length of 15 and an antigen score of 0.4992. Epitopes were connected together using appropriate linkers, and biochemical analysis in PROTPARAM revealed the vaccine's instability index (44.39) and GRAVY (-0.023). Through homology modeling, the Phyre2 server projected a PDB model of the final vaccination, which had 100 percent confidence and 47 percent coverage. The z-score (-4.75) was used to determine the overall quality of the model using ProSA online. Patchdock achieved a molecular docking score of 16070 in a 2366.10 square angstrom region by combining complementing form concepts. The C-IMMSIM server was used to examine the proposed vaccine's immunogenic profile. Immune responses, whether tertiary, secondary, or primary, all played a part in vaccination immunity.

Acknowledgments

Without the help of numerous people, this endeavor would not have been feasible. Many thanks to my adviser, Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University, who looked through all of my modifications and helped me make sense of it all. Also, thanks to Rian Rafsan, Student, MS Biotechnology, Department of Mathematics & Natural Sciences, Brac University, who provided direction and support.

Mohammad Nafees Intesar

Id: 13346017

Department of Pharmacy

Brac University

Table Of Contents:

List of Figures

List of Tables

List of Acronyms

SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
MARS-CoV	Middle East Respiratory Syndrome Coronavirus
S	Spike Glycoprotein
M	Membrane
E	Envelope Glycoproteins
N	Nucleocapsid Protein
RBD	Receptor-Binding Domain
ACE2	Angiotensin-Converting Enzyme 2
TMPRSS2	Transmembrane Protease Serine 2
CTL	Cytotoxic T Lymphocyte
HTL	Helper T Lymphocyte
MH	Major Histocompatibility Complex
IFN-gamma	Interferon-Gamma
PI	Isoelectric Point
GRAVY	Grand Average of Hydropathicity
TLR3	Toll-Like Receptor-3's
NK	Natural Killer
DC	Dendritic Cells

Chapter 1

Introduction & literature review:

A Novel Coronavirus (SARS-CoV-2) is responsible for giving rise to the COVID-19 illness, which is thought to be originated in Wuhan, China. The Wuhan health officials uncovered a few instances of unusual pneumonia in mid-December 2019, which was eventually shown to be the result of a novel coronavirus. It is most likely to have moved from the reservoir of animals to people during the first week of November 2019. [1]. It was revealed then that the RNA virus is the causing pathogen which is linked to the identical family of Coronaviruses producing Severe Acute Respiratory Syndrome (SARS) pandemic in 2003, as well as Respiratory Syndrome, is a pandemic of the Middle East (MERS) in 2012 [2]. Throughout the early phases of the pandemic, it was assumed that a viral transmission between an animal and a person happened in November 2019 at one of Wuhan's biggest wet markets. Additional research was focused on determining which animals were accountable for the emerging zoonotic illness. Although it is currently unknown which species serve as the intermediary host, bats are known to be the principal reservoirs for these viruses. They most likely evolved from a nearby wild-animal ranches [3].

1.1: SARS-CoV-2 virions' structure and genome:

Coronaviridae is a massive family of viruses that infect both humans and animals. NL63, 229E, KHU1, OC43, the seven types of human coronavirus that cause respiratory infections are Middle East respiratory syndrome coronavirus (MERS-CoV), severe acute respiratory syndrome coronavirus (SARS-CoV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). MERS-CoV, SARS-CoV, and SARS-CoV-2 are part of the Beta coronavirus genus. They all exhibit significant mutation rates, resulting in viral diversity, flexibility, and adaptation to various

targets [4]. The SARS-CoV-2 is an encapsulated virus of 60 to 140 nm virions that are generally spherical or somewhat pleomorphic. The spike glycoprotein (S), mainly produced lomers by the virgin surface, provides the virus the 'corona,' or the crown-like shape found on the viral membrane in the electron microscope. The membrane (M) and envelope (E) glycoproteins contribute to the ring structure. A spiral nucleocapsid consisting of a nucleocapsid (N) protein and a single positive-strand RNA genome weighing about 30 kb is found within the virion interior [5].

1.2: Replication cycle and pathogenesis of SARS-CoV-2:

The coronavirus is an intracellular obligate virus that uses the host cell system for replication and dissemination. Because virus-host interactions are the foundation of illnesses, it is crucial to understand how they interact, especially when finding essential antivirals targets. The transmembrane spike (S) glycoprotein, which produces homotrimers projecting mostly from the viral layer, is responsible for SARS-CoV-2 entrance into host cells. Coronavirus S protein is made up of two functional subunits: the S1 subunit, which contains the receptor-binding domain (RBD) that binds to host cell surface receptors, and the S2 subunit, which promotes the eventual merging of both the viral as well as host cell membranes [6], [7]. The RBD of SARS-CoV-2 binds to its peptide region of angiotensin-converting enzyme 2 (ACE2), which also serves as SARS-CoV's cell receptor. The RBD region of the SARS-CoV-2 genome is the most varied one. Six RBD amino acids are essential in ACE2 receptor binding, and five of these residues vary between SARS-CoV and SARS-CoV-2 [8]. After the RBD inside the subunit S1 binds to its ACE2 receptor, the S protein of SARS-CoV-2 is fragmented by a cell surface-associated transmembrane protease serine 2 (TMPRSS2) that activates the S2 region, causing the viral and host cell membranes to fuse. The viruses were firmly blocked from penetrating host cells by this anti-CD147 humanized antibody. SARS-CoV-2 and many other coronaviruses penetrate target cells through receptor-mediated

endocytosis. The virus fusion with endosome membranes releases the viral nucleocapsid into the infected cell's cytoplasm [9]. Coronavirus replication begins with the frameshifting viral RNA being released and uncoated in the cytoplasm. Internal viral proteases process polyproteins, a potential therapeutic target crystal structure recently discovered for SARS-CoV-2. Coronavirus RNA replication takes place on a modified endoplasmic reticulum (ER) membrane-based reticulovesicular network generated by the virus [10].

Vaccination is a critical strategy for controlling and eliminating the virus. SARS-CoV-2 vaccine development presents an urgent requirement [11]. Traditional vaccine development procedures take a long time and need much effort. Immunoinformatic tools investigate the host immune response mechanism to produce alternative techniques to generate vaccines against illnesses that are affordable and efficient since predictions in silico can minimize the number of trials required. SARS-CoV-2 epitope-based peptide vaccines have been developed in dozens of experiments [12]. Although immunoinformatic methods have been used to create many vaccines, most of them are spike protein-based. Antibodies that impede SARS-CoV fusion, binding, and neutralizing of the infection might be induced by a vaccine based on the spike protein [13]. However, there are still other challenges. For example, the SARS vaccine based on a spike protein may trigger adverse immune responses, resulting in liver damage in inoculated animals [14]. Other viral proteins are being examined as potential candidates for developing vaccines that are both protective and less damaging to the immune system [15].

In this study, highly promising epitopes from envelope proteins were screened. Multiepitope-based vaccination candidates against SARS-CoV-2 coronavirus infection were generated and suggested, including cytotoxic T lymphocyte (CTL) helper T lymphocyte (HTL) epitopes.

Chapter 2

Materials and Method

A flowchart describing the procedures involved in the design of the multi-epitope peptide vaccine is shown in Figure 1.

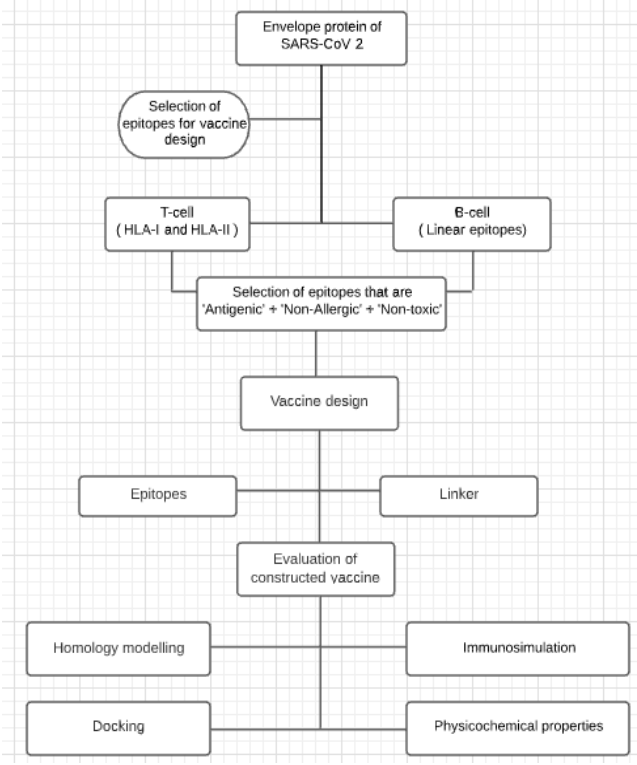


Figure 1: Process of designing a multi-epitope vaccine against SARS-CoV 2

2.1: Retrieval of SARS-CoV-2 Envelope (E) Protein Sequence:

Using the Vipr database, we searched the proteome of SARS CoV-2 for an excellent protein candidate that is highly antigenic. We discovered one protein candidate with high antigenicity while still non-lethal to the host. A non-structural protein cannot be targeted because for up to two weeks, the SARS CoV-2 virus can remain dormant in the host system. For that reason, we looked

at all of the proteins accessible to find an ideal candidate, particularly one that is a structural protein and can be utilized to identify the pathogen immediately [16].

The Envelope protein's whole genome and protein sequences were retrieved in fasta format from the Vipr database. The Vaxijen v2.0 server was used to determine the protein's probable antigenicity (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [17]. "Virus" was chosen as the target organism, and a threshold of 0.5 was set at the server. Because of its validity, length, and quality, the protein sequence selected was used to construct the vaccine further. The protein sequence was then submitted for epitope prediction for helper T-lymphocytes (HTL) and cytotoxic T-lymphocytes (CTL).

2.2: Screening of Cytotoxic T-cell Lymphocytic (CTL) Epitopes:

NetCTL-1.2 has been shown to have excellent predictive ability [18]. SARS-CoV 2's CTL epitopes were predicted with the NetCTL 1.2 server, with good sensitivity and specificity at 0.75 (http://www.cbs.dtu.dk/services/NetCTL/) [18]. MHC class I binding epitopes were predicted, and to expect it, the A1 supertype was chosen using artificial neural networks. Using a half-maximal inhibitory dose (IC50) of 500 nm with a combined score as a guideline, the most promising options for SARS-CoV-2 vaccine development were selected. [19]. The IC50 value of 500 nm indicates that the epitope seems to have a strong affinity toward the receptor. All were predicted using the integrated score, class I binding, TAP transport efficiency, and proteasomal cleavage prediction. At 0.15 and 0.05, respectively, we weighed C-terminal cleavage and TAP transport efficacy.

2.3: MHC I Alleles identification:

We used the NetMHC Pan 4.1 server to discover MHC I alleles specific to CTL epitopes. The stronger an epitope's binding affinity for an allele is, the lower its percentile rank. Epitopes with a

percentile level of 2 were chosen, and the vaccine formulation did not comprise the epitopes anymore that have a higher percentile rank than this level. We used the NetCTL server's input to determine how strong a binding affinity is [21].

2.4: Screening of Helper T Lymphocytic (HTL) epitopes:

The protein sequence's HTL epitopes were predicted with default settings using the NetMHC II pan 4.0's MHC-II epitope prediction module. The primary antigen was used with a default peptide length of 9 being set [22]. The percentile ranks of the generated epitopes were used to rank them. HTL receptors with a lower percentile rank score have a greater binding affinity. The percentile rank was set as 0.5 as a threshold [23].

2.5: Cytokine inducing capability of predicted HTL Epitopes:

The cytokine interferon-gamma (IFN-gamma) plays a key role in antiviral defenses. It triggers both native and targeted immune responses by stimulating macrophages and natural killer cells. In addition, IFN- boosts MHC's antigen response [24]. HTL epitopes were evaluated by predicting IFN epitopes and estimating IL4 productivity and IL10 productivity for screening out the most effective ones. The IFN epitope server, IL4 pred server, and IL10 pred server were utilized.[25].

2.6: Screening of B-cell Epitopes:

The BepiPred linear epitope prediction server (http://tools.iedb.org/bcell/result/) was used to predict Linear B cell epitopes. At a threshold of 0.5, the SARS-CoV-2 protein’s linear B cell epitopes were expected [26].

2.7: Construction of the vaccine:

A multi-epitope polypeptide vaccine was created by combining all screened CTL and HTL epitopes. The adjuvant compound beta-defensin was used to boost the vaccine's immunological response. Using an EAAK linker, the beta-defensin adjuvant was attached to the multi-epitope polypeptide's N terminal, allowing for proper functional domain spacing and efficient production and detection by the host immune system. As epitopes are minimally immunogenic, combining CTL, HTL, and B-cell epitopes through AAY and GPGPG linkers maximizes immunogenicity and epitope expression, resulting in molecular vaccination effectiveness [27].

2.8: Biochemical Analysis of the Constructed Vaccine:

The ProtParam tool was used to assess further the final vaccination protein's physicochemical characteristics (http://web.expasy.org/protparam/). Physicochemical properties investigated included the theoretical isoelectric point (pI), number of amino acids, molecular weight, formula, atomic composition, amino acid composition, extinction coefficients, instability index, estimated half-life, grand average of hydropathicity (GRAVY), and amino acid composition. User-entered sequences were used to calculate the theoretical pI and molecular weight, and the atomic and amino acid compositions were self-evident. The information about a protein's amino acid composition was used to calculate its extinction coefficient [28].

2.9: Prediction of Toxicity and Allergenicity:

The prediction of our proposed vaccine was made with the Toxin and Toxin Target Database (T3DB). This tool focuses on giving toxicity mechanisms and target proteins for each toxin [29].

The vaccine's allergenicity should be non-allergic since the allergenic proteins trigger a detrimental immune response. AllergenOnline server was used to assess the non-allergic nature of the vaccination sequence. [30].

2.10: Homology modeling of vaccine to generate 3D model:

The vaccination was a rebuilt protein with no homology that could be detected. To simulate portions of proteins with no observable homology, Phyre2 uses a structure-based folding simulation. The three-dimensional structure of the intended vaccine was predicted using the Phyre 2 server (http://www.sbg.bio.ic.ac.uk/phyre2/). For creating a protein sequence's full-length 3D model , the program employs modeling of multiple template with simple structure-based folding simulation [31].

2.11: Ramachandran Plotting and Evaluation of the Vaccine's Tertiary Structure for Quality:

The SWISS-MODEL workstation produced a Ramachandran plot to evaluate the constructed vaccine's tertiary structure [32]. The Ramachandran plot reveals favorable locations for the amino acid residues backbone dihedral angles in protein structure. The page of Structure Assessment displays the best scores of Molprobity and allows us to quickly discover where low-quality residues are located in the system or model. After that, the ProSA-web tool was used to validate the protein structure of the vaccine (https://prosa.services.came.ac.at/prosa.php). A positive Z-score indicates that a created 3D protein model piece is incorrect or unpredictable [32].

2.12: SARS-CoV-2 Vaccine's Molecular Docking with Related Antigenic Recognition Receptors

The toll-like receptor-3's (TLR3) antigenic recognition receptors and the immune cell's major histocompatibility complex that the vaccine construct binds to were determined. [33]. The PatchDock server (https://bioinfo3d.cs.tau.ac.il/ PatchDock/) was utilized to confirm the binding affinity of the suggested vaccination construct between these receptors [33]. The server used three algorithms to forecast the possible complex: molecular form representations, filtering, surface patch matching, and scoring.

2.13: Immune Simulations:

The C-IMMSIM server evaluates the immune response and immunogenicity of the developed vaccine. The Celada-Seiden model is used in the C-ImmSim to describe profiles of mammalian immune systems, both humoral and cellular, in response to a specified vaccination. The simulation was run with the default settings, and the simulation took 300 steps to complete. At stages 1, 84, and 168, a tri-dosage technique was used in injection. On the other hand, the immunization was supposed to be administered three times at 28-day intervals [34].

2.14: Remarks on the Materials and Method:

Our comprehensive research was done following the in-silico method, which means that all the predictions and analyses were made using online servers. We cannot confidently say that our final product will be a highly efficient vaccination capable of eradicating the COVID-19 viral infection. We believe that further study is needed as it has the potential to become a vaccine candidate.

Chapter 3

Results

3.1: Antigenicity prediction of Envelope Protein (E):

Of the four structural proteins of SARS-CoV-2, an Envelope protein (E) was selected after a rigorous screening process. The full amino acid sequence (FASTA format) of the envelope protein of the SARS-CoV-2 is given below:

>gb:VIGOR4_HG994158_1_8_26245_26472|ncbiId: VIGOR4_HG994158_1_8_26245_26472|UniProtKB: -N/A-|Organism: Severe acute respiratory syndrome coronavirus 2|Strain Name:1|Protein Name: envelope protein|Gene Symbol: E

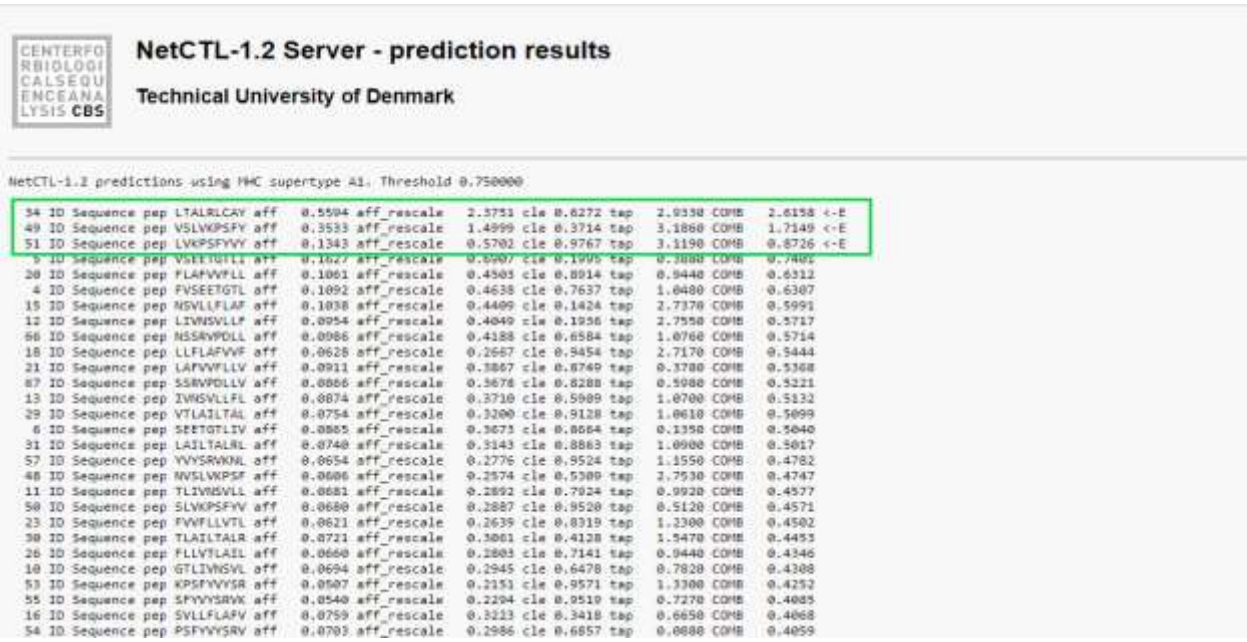
MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSR VKNLNSSRVPDLLV

Then, the VaxiJen v2.0 server examined the core antigen for antigenicity qualities, which showed a score of 0.6025 (Figure 2).

Figure 2: Antigenicity score on the VaxiJen v2.0 server.

3.2: Identification of CTL epitopes:

The Net CTL 1.2 server was used to find CTL epitopes and at a threshold of 0.75, the A1 supertype of MHC I allele epitopes was discovered. In epitope selection, the total score is a significant determinant. This combination score is based on TAP transit efficiency and C terminal cleavage, where the minimal limits were set at 0.15 and 0.05, respectively. The final results are shown in Figure 3.



NetCTL-1.2 Server - prediction results

Technical University of Denmark

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

34	ID	Sequence	pep	LTALRLCAY	aff	0.5504	aff_rescale	2.3751	cle	0.8272	tap	2.9330	COMB	2.6158	<-E
49	ID	Sequence	pep	VSLVKPSFY	aff	0.3533	aff_rescale	1.4899	cle	0.3714	tap	3.1860	COMB	1.7149	<-E
51	ID	Sequence	pep	LVKPSFYVY	aff	0.1343	aff_rescale	0.5702	cle	0.9767	tap	3.1190	COMB	0.8726	<-E
5	ID	Sequence	pep	VSEETGTLL	aff	0.1627	aff_rescale	0.6907	cle	0.1995	tap	0.3080	COMB	0.7401	
20	ID	Sequence	pep	FLAFVVFLL	aff	0.1061	aff_rescale	0.4503	cle	0.8914	tap	0.9440	COMB	0.6312	
4	ID	Sequence	pep	FVSEETGTL	aff	0.1092	aff_rescale	0.4638	cle	0.7637	tap	1.0480	COMB	0.6307	
15	ID	Sequence	pep	NSVLLFLAF	aff	0.1038	aff_rescale	0.4409	cle	0.1424	tap	2.7370	COMB	0.5991	
12	ID	Sequence	pep	LIVNSVLLF	aff	0.0954	aff_rescale	0.4049	cle	0.1956	tap	2.7550	COMB	0.5717	
66	ID	Sequence	pep	NSSRVPDLL	aff	0.0986	aff_rescale	0.4188	cle	0.6584	tap	1.0760	COMB	0.5714	
18	ID	Sequence	pep	LLFLAFVVF	aff	0.0628	aff_rescale	0.2667	cle	0.9454	tap	2.7170	COMB	0.5444	
21	ID	Sequence	pep	LAFVVFLLV	aff	0.0911	aff_rescale	0.3867	cle	0.8749	tap	0.3780	COMB	0.5368	
67	ID	Sequence	pep	SSRVPDLLV	aff	0.0886	aff_rescale	0.3678	cle	0.8288	tap	0.5080	COMB	0.5221	
13	ID	Sequence	pep	IVNSVLLFL	aff	0.0874	aff_rescale	0.3710	cle	0.5989	tap	1.0700	COMB	0.5132	
29	ID	Sequence	pep	VTLAILTAL	aff	0.0754	aff_rescale	0.3200	cle	0.9128	tap	1.0610	COMB	0.5099	
6	ID	Sequence	pep	SEETGTLIV	aff	0.0865	aff_rescale	0.3673	cle	0.8684	tap	0.1350	COMB	0.5040	
31	ID	Sequence	pep	LAILTALRL	aff	0.0740	aff_rescale	0.3143	cle	0.8863	tap	1.0900	COMB	0.5017	
57	ID	Sequence	pep	YVYSRVKNL	aff	0.0654	aff_rescale	0.2776	cle	0.9524	tap	1.1550	COMB	0.4782	
48	ID	Sequence	pep	NVSLVKPSF	aff	0.0606	aff_rescale	0.2574	cle	0.5309	tap	2.7530	COMB	0.4747	
11	ID	Sequence	pep	TLIVNSVLL	aff	0.0681	aff_rescale	0.2892	cle	0.7024	tap	0.9920	COMB	0.4577	
50	ID	Sequence	pep	SLVKPSFYV	aff	0.0680	aff_rescale	0.2887	cle	0.9520	tap	0.5120	COMB	0.4571	
23	ID	Sequence	pep	FVVFLLVTL	aff	0.0621	aff_rescale	0.2639	cle	0.8319	tap	1.2300	COMB	0.4502	
30	ID	Sequence	pep	TLAILTALR	aff	0.0721	aff_rescale	0.3061	cle	0.4128	tap	1.5470	COMB	0.4453	
26	ID	Sequence	pep	FLLVTLAIL	aff	0.0660	aff_rescale	0.2803	cle	0.7141	tap	0.9440	COMB	0.4346	
10	ID	Sequence	pep	GTLIVNSVL	aff	0.0694	aff_rescale	0.2945	cle	0.6478	tap	0.7820	COMB	0.4308	
53	ID	Sequence	pep	KPSFYVYSR	aff	0.0507	aff_rescale	0.2151	cle	0.9571	tap	1.3300	COMB	0.4252	
55	ID	Sequence	pep	SFYVYSRVK	aff	0.0540	aff_rescale	0.2294	cle	0.9519	tap	0.7270	COMB	0.4085	
16	ID	Sequence	pep	SVLLFLAFV	aff	0.0759	aff_rescale	0.3213	cle	0.3418	tap	0.6650	COMB	0.4068	
54	ID	Sequence	pep	PSFYVYSRV	aff	0.0703	aff_rescale	0.2986	cle	0.6857	tap	0.0880	COMB	0.4059	

Figure 3: CTL prediction results on NetCTL-1.2 server

From the results shown in Figure 3 only the epitopes that showed a combined scores of > 0.7 were selected. The selected CTL epitopes are shown in Table 1.

CTL Epitopes	Combined Score
LTALRLCAY	2.6158
VSLVKPSFY	1.7149
LVKPSFYVY	0.8726

Table 1: Combined Score of CTL epitopes

3.3: MHC I alleles specific to CTL epitopes:

Using the NetMHC Pan 4.1 server, the previously specified Epitopes were utilized as input to obtain MHC I alleles. In this scenario, the percentile rank is a metric used in epitope selection, and a greater binding affinity is indicated by lower percentile score and vice versa. For epitope selection, a minimum threshold of 2 was set in this example. A list of CTL epitopes and MHC I allele-specific binding and the associated binding affinity in percentile rank are shown in Table 2.

Strong binding peptides have a threshold of 0.500

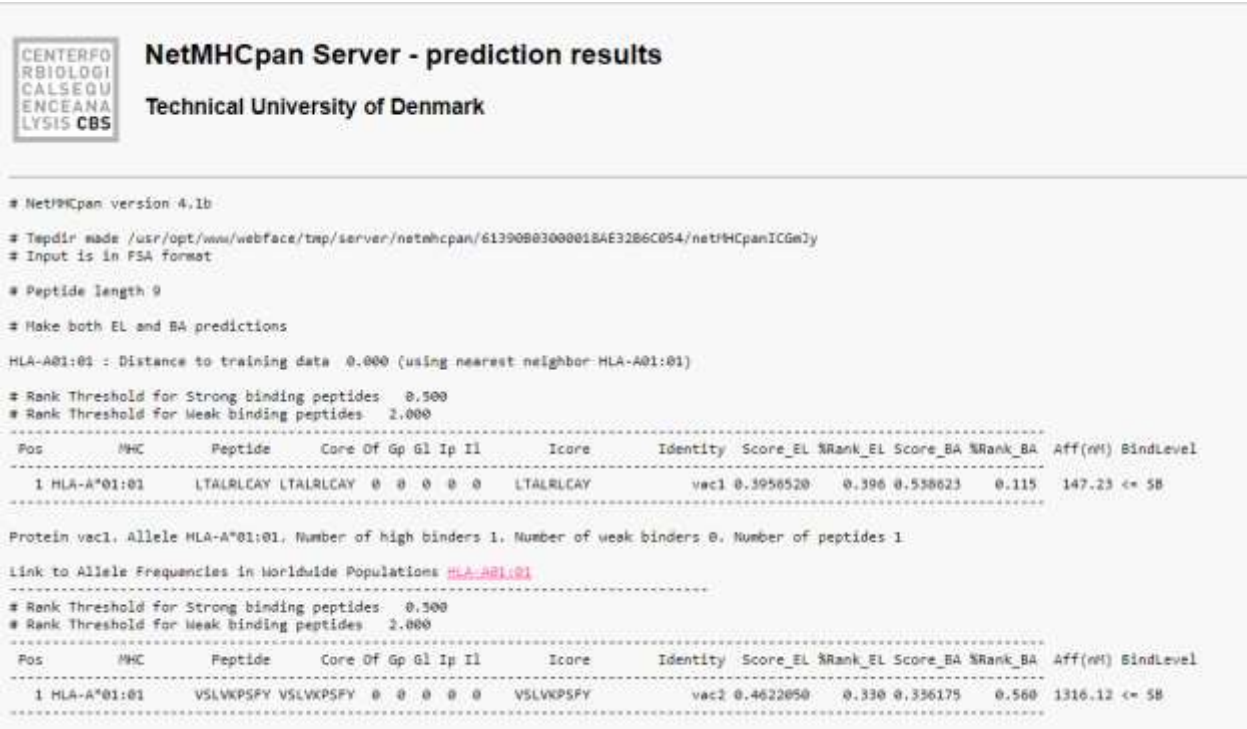
Weak binding peptides have a point of 2.000.

Allele	Peptide	Seq_num	Start	End	Length	Rank
HLA-A*01:01	LTALRLCAY	1	1	9	9	0.27

HLA-A*30:02	VSLVKPSFY	2	1	9	9	0.08
HLA-A*30:02	LVKPSFYVY	3	1	9	9	0.04

Table 2: MHC I allele for specific epitopes along with sequence number, start, end, length, and their percentile rank

MHC I alleles that are specific to the corresponding CTL epitopes were predicted using the NetMHC pan server. The results were shown in Figure 4.



NetMHCpan Server - prediction results

Technical University of Denmark

```
# NetMHCpan version 4.1b

# Tmpdir made /usr/opt/www/webface/tmp/server/netmhcpan/61390B03000018AE32B6C054/netMHCpanICGmJy
# Input is in FSA format

# Peptide length 9

# Make both EL and BA 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 Score_BA %Rank_BA  Aff(nM) BindLevel
-----------------------------------------------------------------------------------------------------------------------
   1 HLA-A*01:01      LTALRLCAY LTALRLCAY  0  0  0  0  0    LTALRLCAY           vac1 0.3956520    0.396 0.558623    0.115   147.23 <= 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 Score_BA %Rank_BA  Aff(nM) BindLevel
-----------------------------------------------------------------------------------------------------------------------
   1 HLA-A*01:01      VSLVKPSFY VSLVKPSFY  0  0  0  0  0    VSLVKPSFY           vac2 0.4622050    0.330 0.336175    0.560  1316.12 <= SB
-----------------------------------------------------------------------------------------------------------------------
```

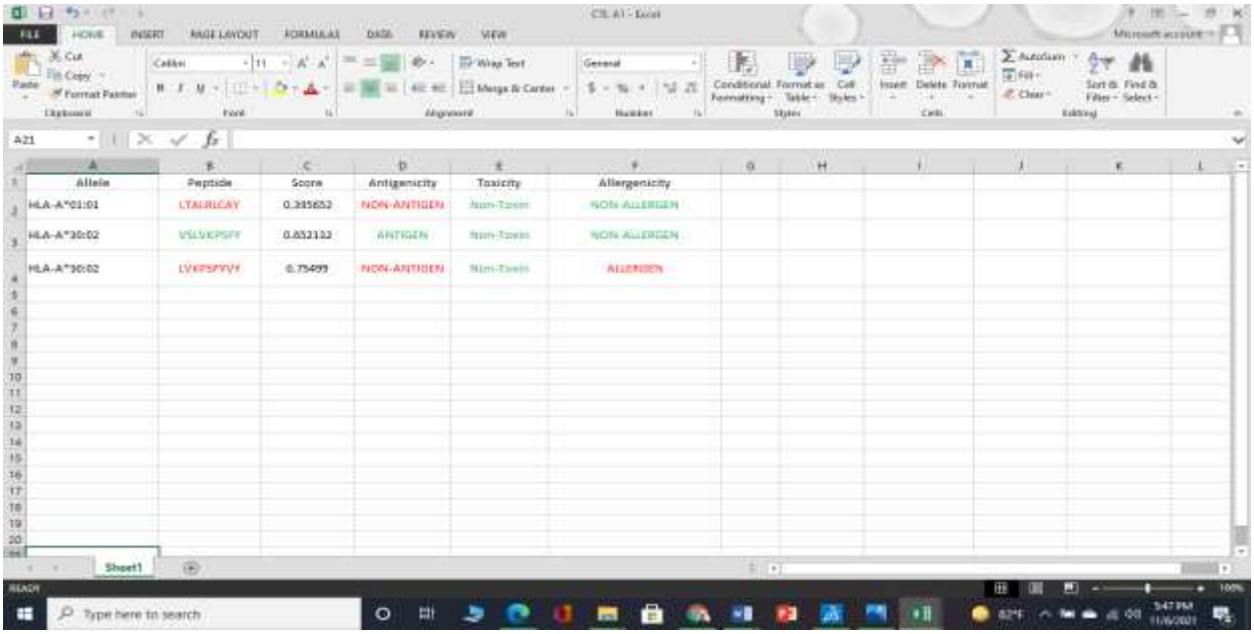
Figure 4: Prediction results for MHC I alleles specific to CTL epitopes in NetMHC pan server

3.4: Antigenicity, Allergenicity, and Toxicity prediction of CTL epitopes:

AllerTOP v2.0 was used to identify the allergenicity of T-cell epitopes, and two of the three CTL epitopes were projected to be non-allergenic. ToxinPred, a support vector machine (SVM)-based approach, was used to assess the toxicity, hydrophobicity, hydropathicity, hydrophilicity, molecular weight and charge of the CTL epitopes (Figure 5).

Figure 5: Prediction of Toxic peptides on ToxinPred server

From figure 5, it was predicted that none of the epitopes tested were hazardous. VaxiJen v2.0 was used to predict epitope antigenicity, and only one epitope was determined to be antigenic (Fig 6).



Allele	Peptide	Score	Antigenicity	Toxicity	Allergenicity
HLA-A*01:01	LTAIRLCAY	0.385652	NON-ANTIGEN	Non-Toxin	NON-ALLERGEN
HLA-A*30:02	VSLVKPSFY	0.652133	ANTIGEN	Non-Toxin	NON-ALLERGEN
HLA-A*30:02	LVKPSFYVY	0.75499	NON-ANTIGEN	Non-Toxin	ALLERGEN

Figure 6: Antigenicity, Toxicity, Allergenicity prediction of CTL epitopes

3.5: MHC II alleles specific to HTL epitopes:

The NetMHCIIpan 4.0 server uses the core antigen as an input to detect MHC II alleles. MHC II alleles can be identified using percentile rank; a greater binding affinity is indicated by a lower percentile rank and vice versa. For the study, only vital binding peptides were selected. For allele identification, a percentile rank of 0.5 was used (Table 3).

The Strong binding peptides threshold (%Rank) 1%

The Weak binding peptides threshold (%Rank) is 5%

Peptide sequence	Core	Alleles	Percentile Rank	Score-EL
VYSRVKNLNSSRVPD	VKNLNSSRV	DRB1_0102	0.72	0.902512

YSRVKNLNSSRVPDL			0.38	0.880447
PSFYVYSRVKNLNSS	YVYSRVKNL	DRB1_0103	0.53	0.746331
KPSFYVYSRVKNLNS			0.72	0.671424
FYVYSRVKNLNSSRV	YSRVKNLNS	DRB1_0401	0.03	0.672255
SFYVYSRVKNLNSSR			0.05	0.663956
FYVYSRVKNLNSSRV	YSRVKNLNS	DRB1_0408	0.04	0.763612
SFYVYSRVKNLNSSR			0.07	0.754284
PSFYVYSRVKNLNSS	YVYSRVKNL	DRB1_0701	0.27	0.891900
KPSFYVYSRVKNLNS			0.24	0.879817
VKPSFYVYSRVKNLN			0.26	0.822794
SFYVYSRVKNLNSSR			0.25	0.659824
KPSFYVYSRVKNLNS	FYVYSRVKN	DRB1_0803	0.85	0.568234
VKPSFYVYSRVKNLN			1.37	0.516976
PSFYVYSRVKNLNSS	YVYSRVKNL	DRB1_0901	0.53	0.798975
KPSFYVYSRVKNLNS			0.53	0.777483
VKPSFYVYSRVKNLN			0.60	0.710573
VYSRVKNLNSSRVPD	VKNLNSSRV	DRB1_1201	3.62	0.625500
YSRVKNLNSSRVPDL			1.65	0.602444

VYSRVKNLNSSRVPD	VKNLNSSRV	DRB1_1501	1.79	0.657755

Table 3: HTL specific allele, percentile rank & binding level.

MHC II alleles that are specific to the corresponding HTL epitopes were predicted using the NetMHCpan 4.0 server. Only those that showed strong binding level and low percentile rank were selected. The results were shown in Figure 7.

Allele: DRB1_0102

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
58	DRB1_0102	VYSRVKNLNSSRVPD	4	VKNLNSSRV	1.000	vac	0.902512	0.38	NA	0.646495	45.83	0.72	<=SB
59	DRB1_0102	YSRVKNLNSSRVPDL	3	VKNLNSSRV	1.000	vac	0.860447	0.47	NA	0.674743	33.70	0.38	<=SB
60	DRB1_0102	SRVKNLNSSRVPDLL	2	VKNLNSSRV	1.000	vac	0.716611	1.37	NA	0.626523	56.88	1.11	<=WB
57	DRB1_0102	YVYSRVKNLNSSRVP	5	VKNLNSSRV	1.000	vac	0.678873	1.66	NA	0.645807	46.17	0.74	<=WB
28	DRB1_0102	LVTLAILTALRLCAY	3	LAILTALRL	1.000	vac	0.419100	3.55	NA	0.606944	70.38	1.55	<=WB
27	DRB1_0102	LLVTLAILTALRLCA	4	LAILTALRL	1.000	vac	0.394554	4.27	NA	0.570145	95.90	2.45	<=WB
29	DRB1_0102	VTLAILTALRLCAYC	2	LAILTALRL	1.000	vac	0.347228	4.96	NA	0.616255	63.56	1.32	<=WB
54	DRB1_0102	PSFYVYSRVKNLNSS	3	YVYSRVKNL	0.987	vac	0.237429	7.18	NA	0.521616	176.59	5.38	
2	DRB1_0102	YSFVSEETGTLIVNS	3	VSEETGTLI	0.867	vac	0.195643	8.41	NA	0.479278	279.81	8.97	
26	DRB1_0102	FLLVTLAILTALRLC	5	LAILTALRL	1.000	vac	0.188247	8.68	NA	0.545946	136.01	3.93	
53	DRB1_0102	KPSFVYVSRVKNLNS	4	YVYSRVKNL	0.940	vac	0.181863	8.93	NA	0.502966	216.50	6.75	
1	DRB1_0102	HYSFVSEETGTLIVN	4	VSEETGTLI	0.520	vac	0.165540	9.59	NA	0.490809	246.77	7.80	
61	DRB1_0102	RVKNLNSSRVPDLLV	1	VKNLNSSRV	0.973	vac	0.160352	9.82	NA	0.572889	181.71	2.66	
56	DRB1_0102	FYVYSRVKNLNSSRV	6	VKNLNSSRV	0.567	vac	0.090835	14.18	NA	0.603988	72.65	1.63	
52	DRB1_0102	VKPSFYVYSRVKNLN	5	YVYSRVKNL	0.987	vac	0.083734	14.88	NA	0.481739	272.45	8.70	
30	DRB1_0102	TLAILTALRLCAYCC	1	LAILTALRL	0.993	vac	0.081745	15.10	NA	0.591217	83.34	2.00	
55	DRB1_0102	SFYVYSRVKNLNSSR	2	YVYSRVKNL	0.627	vac	0.080636	15.23	NA	0.532052	158.08	4.72	
25	DRB1_0102	VFLLVTLAILTALRL	6	LAILTALRL	0.960	vac	0.077118	15.64	NA	0.553945	124.74	3.50	
3	DRB1_0102	SFVSEETGTLIVNSV	2	VSEETGTLI	0.927	vac	0.063993	17.38	NA	0.405622	620.82	19.61	
46	DRB1_0102	IVNVSLVKPSFYVYS	3	VSLVKPSFY	0.820	vac	0.012293	39.00	NA	0.523297	171.79	5.27	
22	DRB1_0102	AFVVFLLVTLAILTA	3	VFLLVTLAI	0.713	vac	0.011817	39.68	NA	0.393693	706.35	21.91	
45	DRB1_0102	NIVNVSLVKPSFYVY	4	VSLVKPSFY	0.620	vac	0.009973	42.61	NA	0.528531	179.06	5.47	
51	DRB1_0102	LVKPSFVYVSRVKNL	6	YVYSRVKNL	0.653	vac	0.009370	43.74	NA	0.457618	353.73	11.51	
21	DRB1_0102	LAFVVFLLVTLAILT	3	VVFLLVTLA	0.420	vac	0.008554	45.41	NA	0.377580	840.88	25.32	
47	DRB1_0102	VNVSLVKPSFYVYSR	2	VSLVKPSFY	0.560	vac	0.007329	48.24	NA	0.495408	234.87	7.38	
13	DRB1_0102	IVNSVLLFLAFVVFL	4	VLLFLAFVV	0.967	vac	0.007102	48.80	NA	0.307460	1795.66	43.49	
24	DRB1_0102	VVFLLVTLAILTALR	3	LLVTLAILT	0.760	vac	0.006592	50.24	NA	0.430125	475.21	15.35	
14	DRB1_0102	VNSVLLFLAFVVFLL	3	VLLFLAFVV	0.893	vac	0.006164	51.68	NA	0.323170	1514.97	39.01	
44	DRB1_0102	CNIVNVSLVKPSFYV	3	VNVSLVKPS	0.413	vac	0.005904	52.55	NA	0.490927	246.67	7.80	
12	DRB1_0102	LIVNSVLLFLAFVVF	5	VLLFLAFVV	0.960	vac	0.005597	53.59	NA	0.300670	1932.55	45.42	
20	DRB1_0102	FLAFVVFLLVTLAIL	4	VVFLLVTLA	0.587	vac	0.004836	56.47	NA	0.356328	1058.27	30.35	
16	DRB1_0102	SVLLFLAFVVFLLVT	4	FLAFVVFLL	0.580	vac	0.004026	59.96	NA	0.323890	1506.47	38.87	
15	DRB1_0102	FVVFLLVTLAILTAL	7	VFLLVTLAI	0.413	vac	0.003987	60.17	NA	0.381655	804.61	24.44	

Figure 7: MHC II specific to HTL epitopes predicted by NetMHCIIpan 4.0 server

3.6: Capability of HTL epitopes of inducing cytokine:

At first, we found HTL epitopes' capacity to produce interleukin, namely prediction of IFN epitope, the productivity of IL-4, and productivity of IL-10. These predictions were made using the servers IFN epitope, IL-4pred, and IL-10pred. The SVM approach was used with a default threshold of 0.2 and -0.3 for IL4 and IL10 pred servers.

From all the HTL epitopes that have been selected only the interferon gamma inducing ones were selected. Among them only one were found to be positive that has been shown in Figure 8.

Home | Design | Predict | Scan | Algorithm | Application | Dataset | Help | Team | Contact

Prediction result for the IFNepitope server

Show 50 entries Search:

Serial No.	Epitope Name	Sequence	Method	Result	Score
1	s1	VYSRVKNLNSSRVPD	SVM	NEGATIVE	-0.18705221
2	s2	YSRVKNLNSSRVPDL	SVM	NEGATIVE	-0.43267552
3	s3	PSFYVYSRVKNLNSS	SVM	NEGATIVE	-0.35688916
4	s4	KPSFYVYSRVKNLNS	SVM	NEGATIVE	-0.52339177
5	s5	FYVYSRVKNLNSSRV	SVM	NEGATIVE	-0.1032629
6	s6	SFYVYSRVKNLNSSR	SVM	NEGATIVE	-0.4195908
7	s7	FYVYSRVKNLNSSRV	SVM	NEGATIVE	-0.1032629
8	s8	SFYVYSRVKNLNSSR	SVM	NEGATIVE	-0.4195908
9	s9	PSFYVYSRVKNLNSS	SVM	NEGATIVE	-0.0028082947
10	s10	KPSFYVYSRVKNLNS	SVM	NEGATIVE	-0.17049385
11	s11	VKPSFYVYSRVKNLN	SVM	NEGATIVE	-0.46468316
12	s12	SFYVYSRVKNLNSSR	SVM	NEGATIVE	-0.4195908
13	s13	KPSFYVYSRVKNLNS	SVM	NEGATIVE	-0.52339177
14	s14	VKPSFYVYSRVKNLN	SVM	NEGATIVE	-0.46468316
15	s15	PSFYVYSRVKNLNSS	SVM	NEGATIVE	-0.35688916
16	s16	KPSFYVYSRVKNLNS	SVM	NEGATIVE	-0.52339177
17	s17	VKPSFYVYSRVKNLN	SVM	NEGATIVE	-0.46468316
18	s18	VYSRVKNLNSSRVPD	SVM	POSITIVE	0.065683238
19	s19	YSRVKNLNSSRVPDL	SVM	NEGATIVE	-0.43267552
20	s20	VYSRVKNLNSSRVPD	SVM	NEGATIVE	-0.18705221
Serial No.	Epitope Name	Sequence	Method	Result	Score

Showing 1 to 20 of 20 entries Previous Next

Figure 8: IFN epitope prediction for HTL. (Positive is accepted)

Whether the HTL epitopes were interluking-4 inducing or not it was predicted using IL4pred server Figure 9. From there it was found that every one of the epitopes were IL-4 inducing.

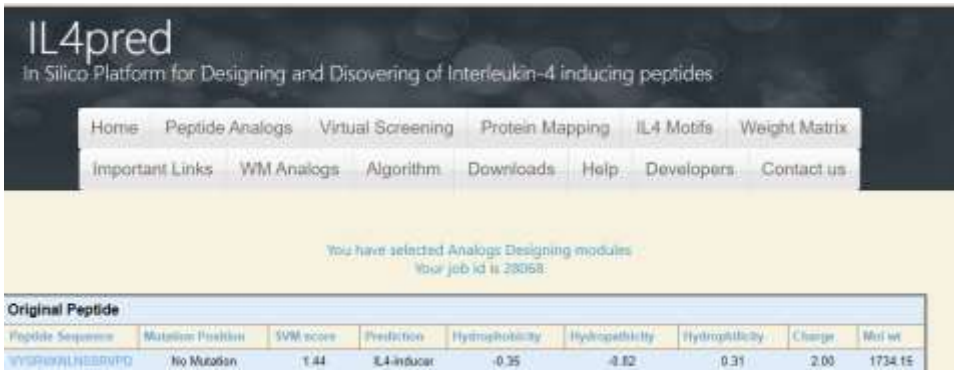
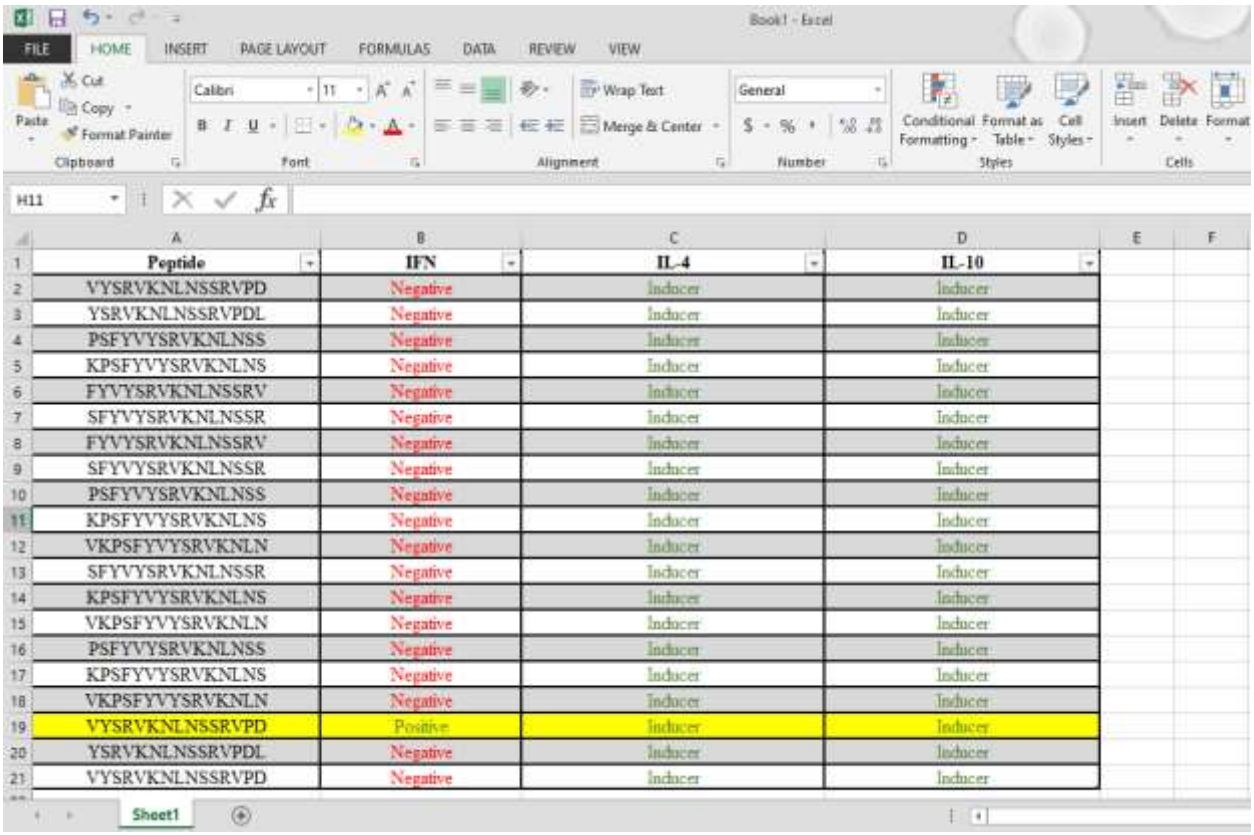


Figure 9: IL4 inducer for HTL epitope.

IL-10 pred server was used to determine the interleukin-10 inducing capabilities of the HTL epitopes (Figure 10). The result showed that all epitopes were capable of inducing interleukin-10.

Figure 10: IL-10 inducer for HTL epitope.

After gathering all the required data regarding the prediction of IFN epitope, the productivity of IL-4, and productivity of IL-10, only those that shoed a positive result was selected for further study. Only one HTL epitope was chosen for vaccine designing as it fulfilled all the criteria for an ideal epitope (Figure 11).



Peptide	IFN	IL-4	IL-10
VYSRVKNLNSSRVPD	Negative	Inducer	Inducer
YSRVKNLNSSRVPDL	Negative	Inducer	Inducer
PSFYVYSRVKNLNSS	Negative	Inducer	Inducer
KPSFYVYSRVKNLNS	Negative	Inducer	Inducer
FYVYSRVKNLNSSRV	Negative	Inducer	Inducer
SFYVYSRVKNLNSSR	Negative	Inducer	Inducer
FYVYSRVKNLNSSRV	Negative	Inducer	Inducer
SFYVYSRVKNLNSSR	Negative	Inducer	Inducer
PSFYVYSRVKNLNSS	Negative	Inducer	Inducer
KPSFYVYSRVKNLNS	Negative	Inducer	Inducer
VKPSFYVYSRVKNLN	Negative	Inducer	Inducer
SFYVYSRVKNLNSSR	Negative	Inducer	Inducer
KPSFYVYSRVKNLNS	Negative	Inducer	Inducer
VKPSFYVYSRVKNLN	Negative	Inducer	Inducer
PSFYVYSRVKNLNSS	Negative	Inducer	Inducer
KPSFYVYSRVKNLNS	Negative	Inducer	Inducer
VKPSFYVYSRVKNLN	Negative	Inducer	Inducer
VYSRVKNLNSSRVPD	Positive	Inducer	Inducer
YSRVKNLNSSRVPDL	Negative	Inducer	Inducer
VYSRVKNLNSSRVPD	Negative	Inducer	Inducer

Figure 11: HTL epitopes from MHC II allele's evaluation

3.7: B-cell epitope prediction:

To find linear B-cell epitopes, the BepiPred linear epitope identification 2.0 was employed, and B-cell epitopes were found at a threshold of 0.5. The starting and ending positions for particular epitopes of B-cell, as well as their lengths, are shown in Figure 12.

Bepipred Linear Epitope Prediction 2.0 Results

Input Sequences

```
  1 MYSFVSEETG TLIVNSVLLF LAFVVFLLVT LAILTALRLC AYCCNIVNVS LVKPSFYVYS
 61 RVKNLNSSRV PDLLV
```

Center position: 4 **Threshold:** 0.500 Recalculate

Average: 0.421 **Minimum:** 0.239 **Maximum:** 0.613

Predicted peptides:

No.	Start	End	Peptide	Length
1	6	9	SEET	4
2	57	71	YVYSRVKNLNSSRVP	15

Figure 12: Predicted peptides with start, end, and length

The anticipated B cell epitopes were plotted with the epitopes' residue scores on a graph acquired from the server (Figure 13).

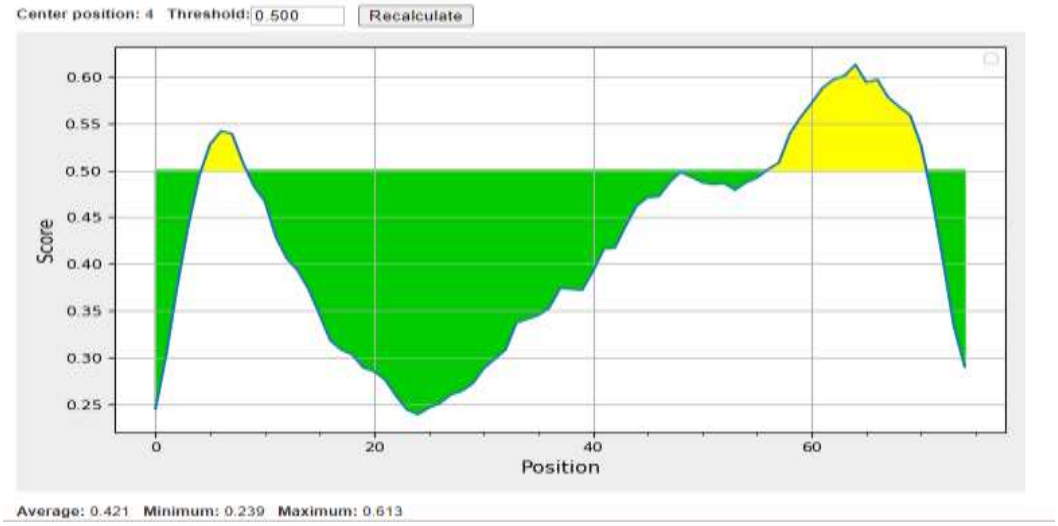


Figure 13: B-cell epitopes score vs. position graph

BepiPred Linear Epitope 2.0 algorithm may also provide the maximum, lowest, and average score produced by the Linear B cell epitopes set. We received a top score of 0.613, a minimum score of 0.239, and an average of 0.421 for our exact B cell epitopes.

3.8: Construction of Final vaccine:

The vaccine was built using the best candidate epitopes that were available. 1 CTL epitope, 1 HTL epitope, and 1 Linear B cell epitope were fused using linker sequences. With the aid of GPGPG linker, the HTL epitopes were combined (The GPGPG linker stimulates the responses of HTL and helpers' immunogenicity is conserved when conformation is taken into account). In contrast, the AAY linker was used for CTL epitopes (AAY linker assists in establishing suitable binding sites for the TAP transporter and increases epitope presentation), and the merging of epitopes of B-cell was done later. Finally, with the aid of the EAAK, the human -defensin-3 sequence was inserted into the vaccine's N-terminal location to enhance immunogenicity of the vaccine. The final constructed vaccine is:

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAKVSLVKPSFYA
AYVYSRVKNLNSSRVPDGPGPGYVYSRVKNLNSSRVP

3.9: Biochemical Analysis of the Constructed Vaccine:

To assess the vaccination, the PROTPARAM program on the Expasy server was utilized to conduct biochemical studies. The server delivers the results based on a molecular formula, molar mass, instability index, aliphatic index, theoretical PI, GRAVY, and other parameters.

From the server it was found that the number of amino acids were 96, molecular weight 10732.54, theoretical pI was 10.18 (Figure 14).

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
GIINTLQKYY CRVRGGRCAV LSCLPKEEQI GKCSTRGRKC CRRKKEAAKV SLVKPSFYAA

        70         80         90
YVYSRVKNLN SSRVPDGPGP GYVYSRVKNL NSSRVP
```

References and documentation are available.

Number of amino acids: 96

Molecular weight: 10732.54

Theoretical pI: 10.18

Figure 14: Constructed Vaccine's number of molecular weight, amino acids, and theoretical pI

From the Figure 15 the amino acid composition was found of the constructed vaccine. Alongside it the total number of negatively and positively charged residues were also found.

Amino acid composition: CSV format

Ala (A)	5	5.2%
Arg (R)	11	11.5%
Asn (N)	5	5.2%
Asp (D)	1	1.0%
Cys (C)	6	6.2%
Gln (Q)	2	2.1%
Glu (E)	3	3.1%
Gly (G)	8	8.3%
His (H)	0	0.0%
Ile (I)	3	3.1%
Leu (L)	6	6.2%
Lys (K)	10	10.4%
Met (M)	0	0.0%
Phe (F)	1	1.0%
Pro (P)	6	6.2%
Ser (S)	10	10.4%
Thr (T)	2	2.1%
Trp (W)	0	0.0%
Tyr (Y)	7	7.3%
Val (V)	10	10.4%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 4
Total number of positively charged residues (Arg + Lys): 21

Figure 15: Constructed vaccine's amino acid composition, the total number of negatively and positively charged residues.

The instability index value shown for the vaccine was 44.39, which according to the server, is unstable as it shows the importance of >40 (Figure 16). The vaccine is hydrophilic as the Grand Average of Hydropathicity (GRAVY) value showed -0.023. Hydrophilic vaccines are preferable as hydrophobic can cause an increased risk of contamination and loss of functionality.

Formula: $C_{928}H_{1522}N_{258}O_{255}S_{10}$
Total number of atoms: 2973

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that
this could result in more than 10% error in the computed extinction coefficient.

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

Ext. coefficient 16890
Abs 0.1% (=1 g/l) 0.816, assuming all pairs of Cys residues form cystines

Ext. coefficient 16390
Abs 0.1% (=1 g/l) 0.792, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is E (Glu).

The estimated half-life is: 1 hours (mammalian reticulocytes, in vitro).
30 min (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 44.39
This classifies the protein as unstable.

Aliphatic index: 96.81

Grand average of hydropathicity (GRAVY): -0.023

Figure 16: Constructed vaccine's formula, number of atoms, half-life, instability index, aliphatic index, and GRAVY value.

3.10: Constructed Vaccine's Allergenicity and Toxicity Evaluation:

The Allergen web server identified the vaccine's allergenic nature using a hybrid method. To efficiently measure the protein's efficacy, we used a value of 0.5 based on z-score analysis and the Full FASTA 36 technique was applied (Figure 17).

Figure 17: Allergenicity of the constructed vaccine from Allergen Online server.

Similarly, the toxicity of constructed vaccine was assessed using the T3DB server, which showed non-toxic in figure 18.

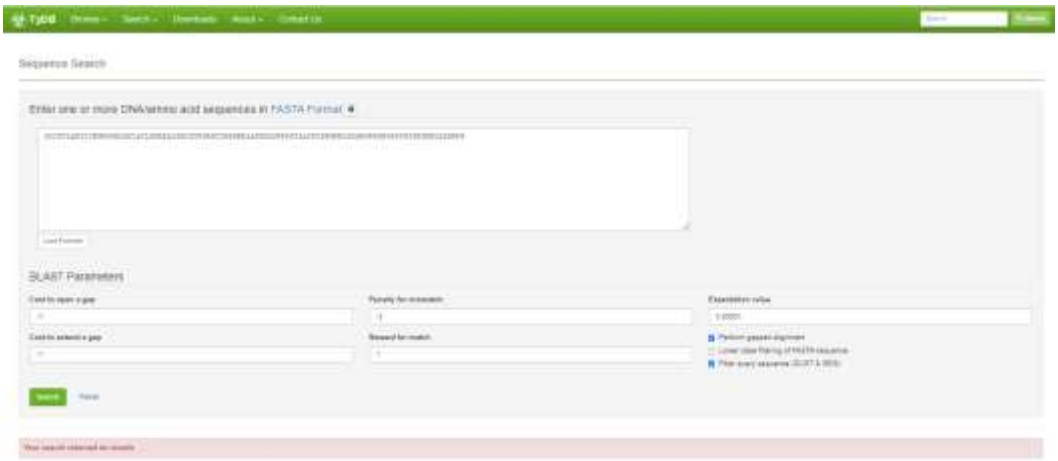


Figure 18: Toxicity prediction of the constructed vaccine.

3.11: Homology modeling of vaccine:

To further continue our research, it is imperative to get a 3D structure of our vaccine. In the in-silico approach, we could create a 3D design in the form of a PDB file. The homology modeling technique was used to construct this PDB file, and the top sorting template was used to model 45 residues with 100% confidence (Figure 19).

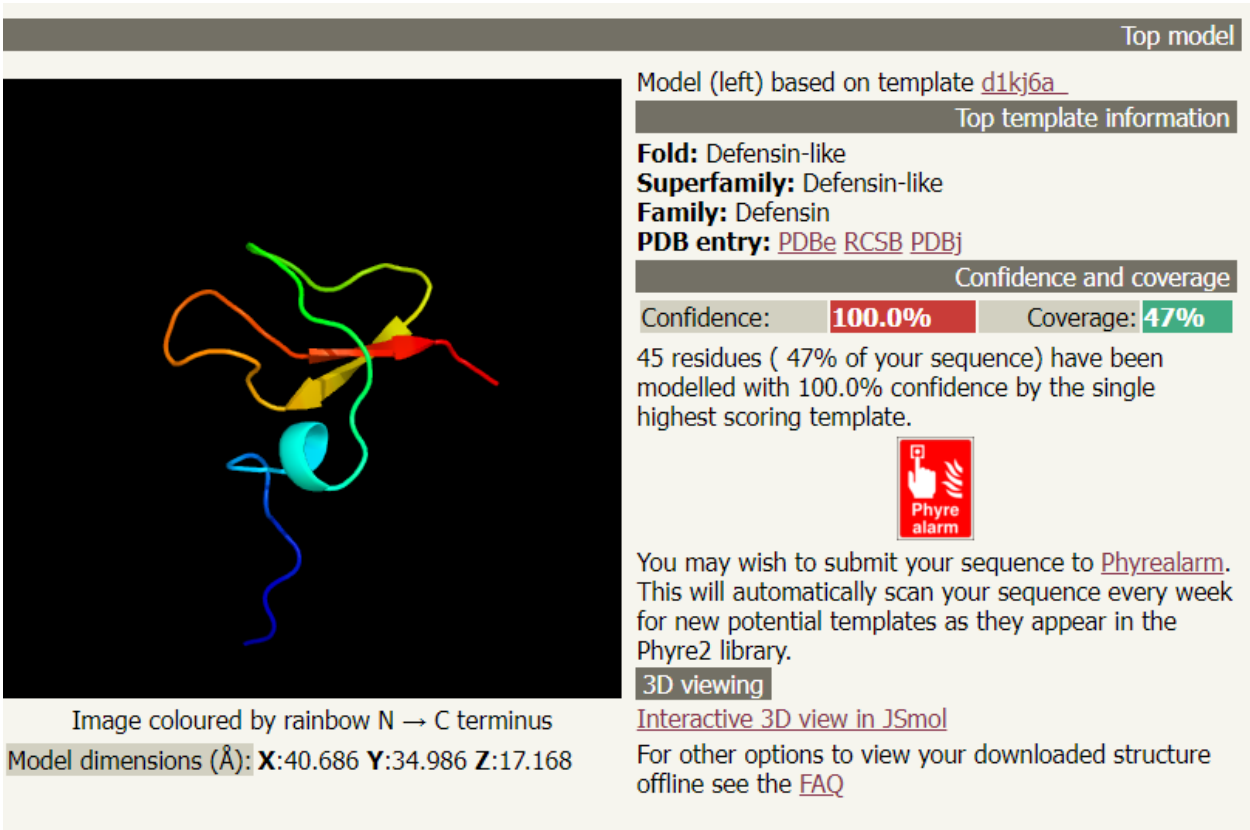


Figure 19: Phyre2 server was used to create a 3D model of the vaccination.

3.12: Homologous vaccine model's analysis:

Our vaccine's PDB structure was examined further, obtained from the phyre2 server. The SWISS PDB plotter was used to do the Ramachandran plot analysis (figure 20), and the PROSA webserver was used to create a Z-score versus residue analysis curve (figure 24).

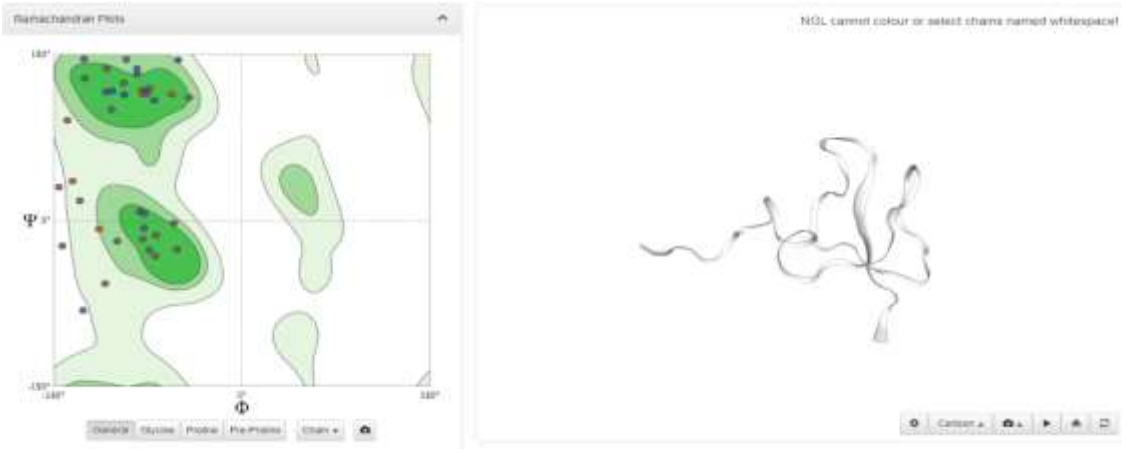


Figure 20: Ramachandran plot using SWISS PDB plotter.

Further analysis of the Ramachandran plotting is shown in figure 21:

MolProbity Results

MolProbity Score	3.18	
Clash Score	80.86	(23 CYS- 41 CYS), (11 CYS- 40 CYS), (10 TYR- 30 ILE), (11 CYS- 30 ILE), (33 CYS- 39 LYS), (27 GLU- 43 ARG), (23 CYS- 24 LEU), (34 SER- 35 THR)
Ramachandran Favoured	74.42%	
Ramachandran Outliers	4.65%	35 THR, 14 ARG
Rotamer Outliers	0.00%	
C-Beta Deviations	0	
Bad Bonds	3 / 361	23 CYS- 41 CYS, 11 CYS- 40 CYS, 18 CYS- 33 CYS
Bad Angles	5 / 478	(11 CYS- 40 CYS), (18 CYS- 33 CYS), (23 CYS- 41 CYS)

Results obtained using MolProbity version 4.4

Figure 21: MolProbity results of Ramachandran plotting.

The results from the plotting from figure 21 showed MolProbity score as 3.18, Clash score as 80.86, Ramachandran Favored as 74.42%, and Ramachandran Outliers as 4.65%.

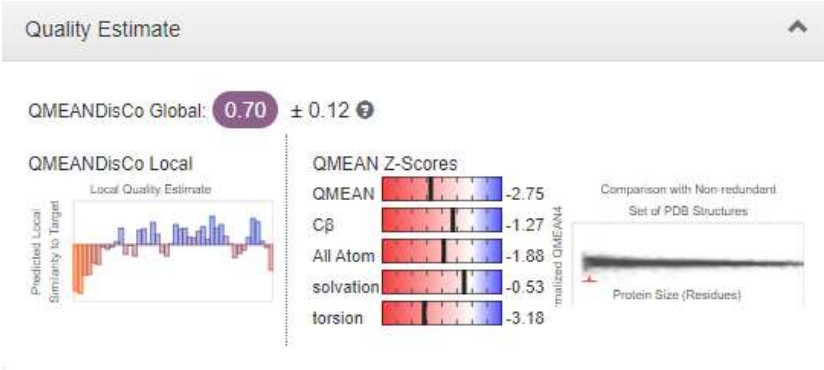


Figure 22: Quality estimation of Ramachandran plot

From the figure 22, we found that the QMEANDisCo's Global score was 0.70.

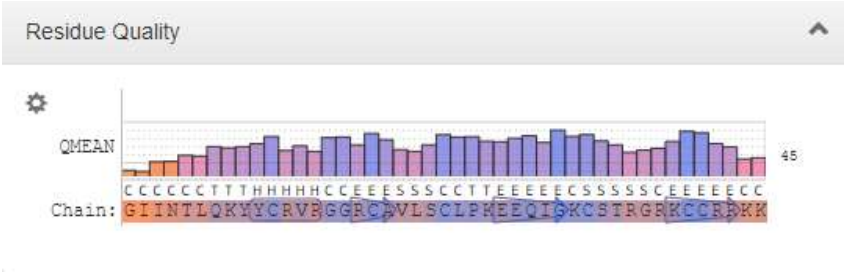


Figure 23: Residue quality estimation of Ramachandran plot

Figure 23 shows the estimation of the residue quality of the constructed vaccine through Ramachandran plotting.

ProSA-web, which predicts the overall quality of the model in the form of a z-score, was used to examine the structural validation of the multiple epitope vaccination. ProSA-web, which predicts the overall quality of the model in the form of a z-score, was used to examine the structural validation of the multiple epitope vaccination. The erroneous structure is indicated if the projected model's z-scores are beyond the characteristic range for natural proteins. The vaccination projected

model had a Z-score of -4.75, suggesting that it was a decent model (Figure 24). The local model quality of the protein was also generated (Figure 25)

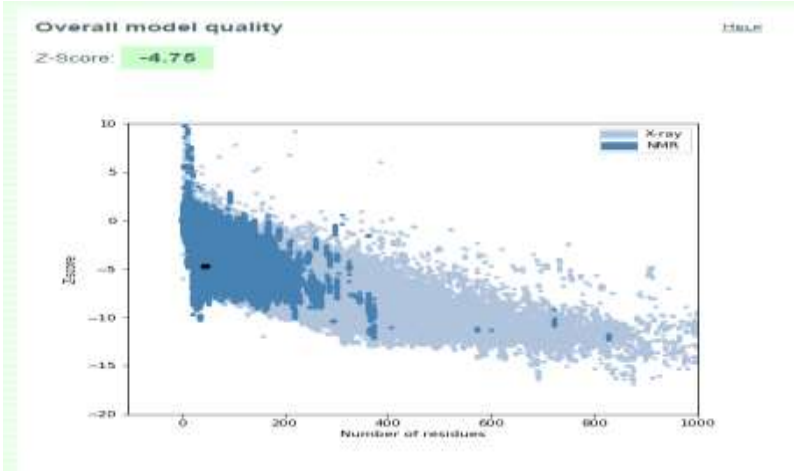


Figure 24: Overall model quality: Z-score analysis

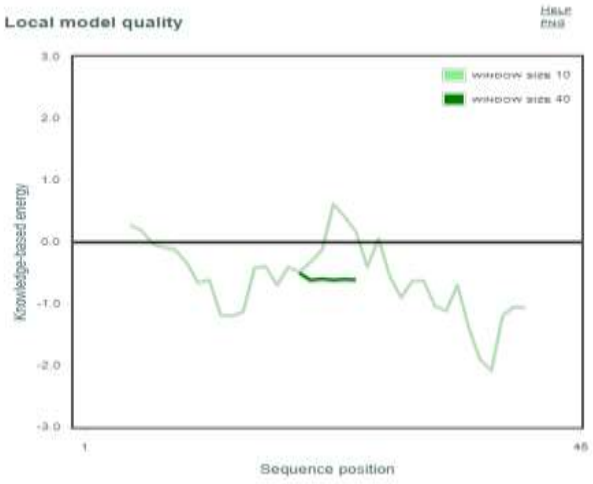


Figure 25: Knowledge-based energy versus sequence position in local model quality

3.13: Molecular Docking of the Relatively Antigenic Receptor with the Final Vaccine Construct:

The binding affinity between the suggested vaccine design and the relevant TLR8 was further investigated using molecular docking, which belongs to the family of toll-like receptor (TLR). The innate immune response is activated by this receptor family, which comprises of protein-rich receptors. Because they are physically designated as a single-pass membrane-spanning receptors, they are frequently discovered to be reveal in cells, which are largely in charge of pathogen elimination that are invading the body. The numbers 1 to 13 on the labels of TLRs distinguish them. Except for TLR 11, 12, and 13, all the other TLR are among the receptors found in both animals and humans. Our research used the ligand (PDB file retrieved from the phyre2 server) as the ligand and TLR8 (PDB ID:3W3G) as the receptor. Patchdock server was used for the docking, and it also provides scores of specific docked complexes. The top 20 solutions of patchdock are monitored in the figure 26.

Receptor	Ligand	Complex Type	Clustering RMSD	User e-mail	Receptor Site	Ligand Site	Distance Constraints
3w3g	Final_vac_2.pdb	Default	4.0	nafeesintesar03@gmail.com	-	-	-

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	16070	2366.10	331.53	-2.60 -0.86 1.58 10.10 4.29 37.61	result.1.pdb
2	15826	2238.10	290.69	-1.35 -0.54 0.89 12.07 26.18 30.79	result.2.pdb
3	15220	2378.20	255.83	0.38 0.65 -1.19 7.15 22.17 33.07	result.3.pdb
4	14950	2459.90	133.06	1.75 0.31 1.93 11.28 6.43 35.89	result.4.pdb
5	14816	2148.30	381.73	-2.61 -0.43 2.12 13.59 3.74 35.52	result.5.pdb
6	14386	1905.10	435.22	0.37 0.46 -1.66 12.35 22.55 35.20	result.6.pdb
7	14346	1859.50	279.24	0.71 0.59 2.21 19.59 1.03 48.32	result.7.pdb
8	14346	1711.50	261.34	-2.93 0.12 -1.57 20.99 3.93 34.66	result.8.pdb
9	14328	1991.40	391.95	-1.21 -0.50 -2.31 9.10 -23.34 20.62	result.9.pdb
10	14304	2164.60	129.32	-2.88 -1.07 1.00 12.77 2.22 38.45	result.10.pdb
11	14236	2113.50	99.02	2.54 -0.59 1.91 5.95 29.98 17.39	result.11.pdb
12	14222	2726.30	402.61	-0.22 -0.20 0.76 -3.07 45.89 47.86	result.12.pdb
13	14200	1795.50	488.67	-1.42 0.99 -1.72 16.24 26.41 15.87	result.13.pdb
14	14178	1991.30	3.92	1.67 -1.38 -2.28 14.48 8.98 38.88	result.14.pdb
15	14168	2145.10	359.90	-2.96 -0.25 1.42 13.84 23.12 21.33	result.15.pdb
16	14136	2079.00	290.09	0.46 0.28 -2.21 4.42 24.89 12.31	result.16.pdb
17	14124	1912.00	465.71	-0.17 -0.94 -1.21 16.40 22.97 25.64	result.17.pdb
18	14100	1774.90	5.79	-1.78 -0.31 1.12 13.78 30.46 21.80	result.18.pdb
19	14064	1801.20	417.14	2.00 0.02 -0.04 17.75 1.83 37.97	result.19.pdb
20	14072	2160.00	482.23	-3.04 0.00 -1.99 8.46 15.47 35.73	result.20.pdb

show next 20 >>

Figure 26: Molecular Docking Algorithm Based on Shape Complementarity Principles.

The results show in figure 26 is the best complex between TLR8 and our proposed vaccine gave the highest score of 16070 with a transformation of (-2.60 -0.86 1.58 10.10 4.29 37.6) 331.53 KJmol-1 was the value of ACE, which covered the area of 2366.10 square angstroms. With the 64-bit client version of Discovery Studio 2016, the PDB structure of the produced protein-ligand combination can be viewed (Figure 27).

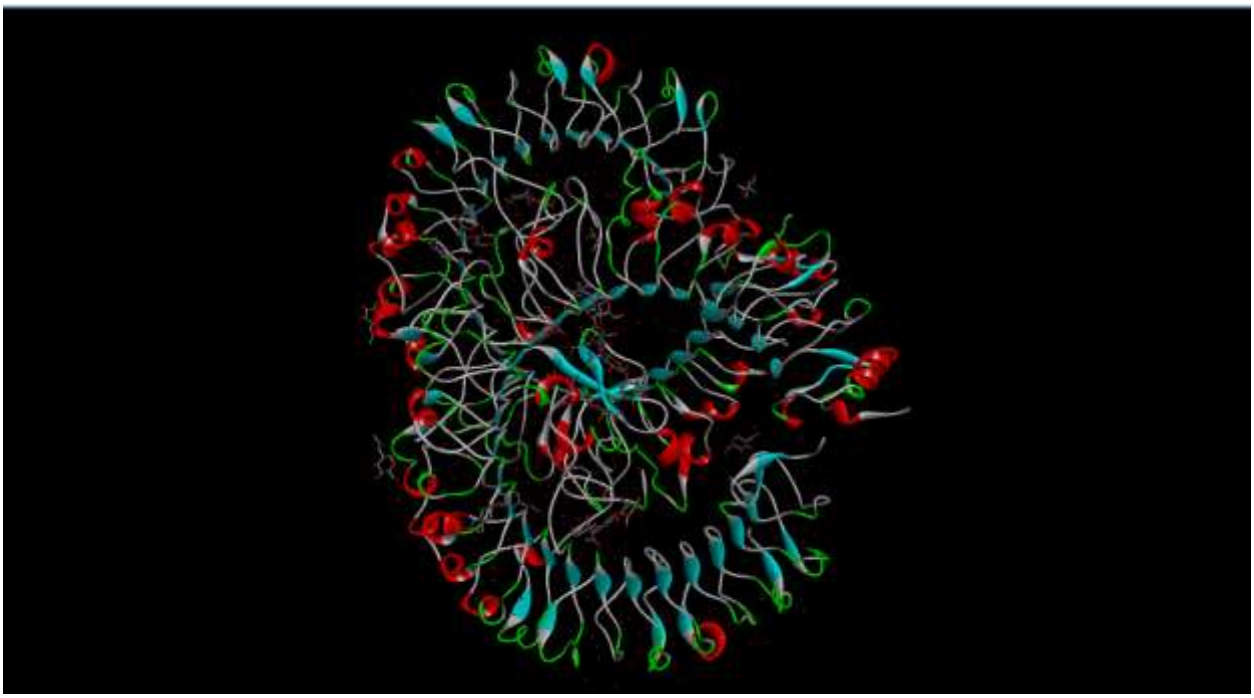


Figure 27: The docked complex between the TLR8 receptor and the proposed vaccination in 3D.

3.14: Immune Simulation in silico for the immune response:

The final vaccination's immune stimulation was carried out utilizing the C-ImmSim web server, which provides immunological profiles for the intended vaccine. IgG1 + IgG2 and IgM were used to identify proliferation in the secondary and tertiary immune responses and a decrease in the antigen count (IgG + IgM), indicating that the immune response had proliferated (Figure 28).

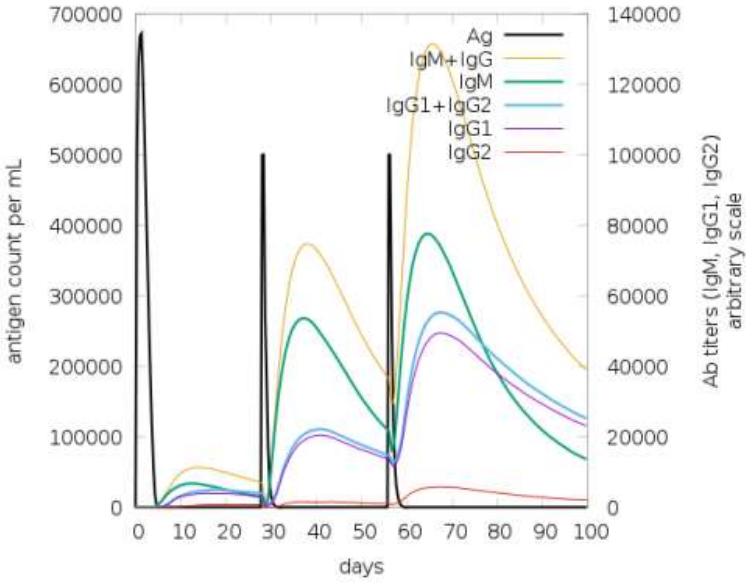


Figure 28: The virus, the immunoglobulins, and the immunocomplexes.

The C-immsim website also calculates the B cell lymphocyte concentration following immunization. Both humoral and cellular immunity relies on B cell epitopes, and IgM, IgG1and IgG2 concentrations determine B cell concentrations. From figure 29 Graphs depicting B cell population densities in each state. Last but not least, plasma B cells were discovered. These cells have the potential to be used as medicinal agents.

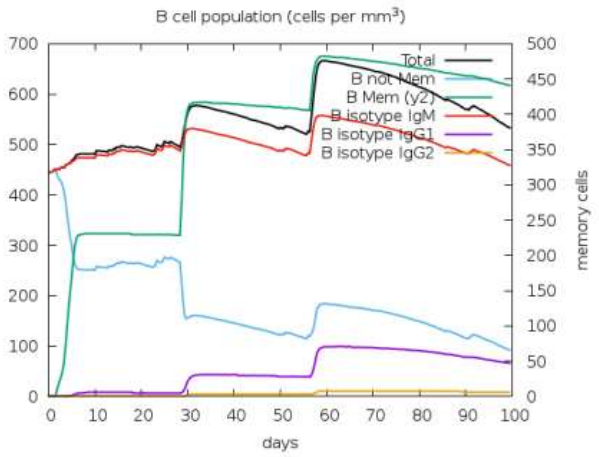


Figure 29: Graph showing the concentration of B cells based on subtypes and administration

In, Figure 30 the graphs show the concentration of B cells is increasing based on state versus days passed after vaccine administration.

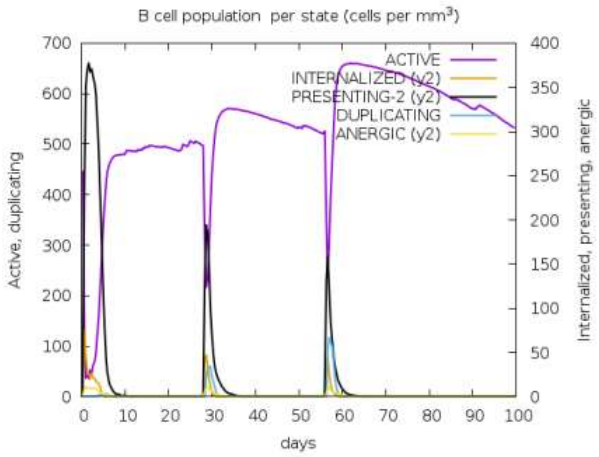


Figure 30: Graph showing entity-state of B cells versus days after vaccine administration.

From figure 31 graph it can be seen that B cell increases in plasma after the vaccine is administered and it slowly decreases before another dose. And the increase of B cell per dose is higher than the previous administration.

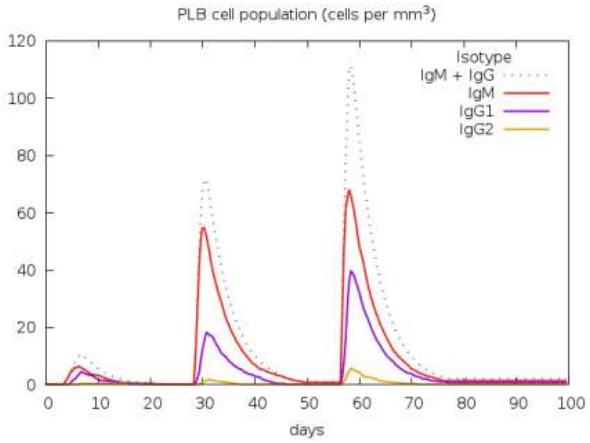


Figure 31: B cell population expansion in plasma vs. vaccination treatment days

Graphs demonstrating CTL and HTL epitope concentrations were received from the server in the same way as plots exhibiting B cell concentrations were generated. Graphs depicting CTL and HTL concentrations are shown in Figure 32-33.

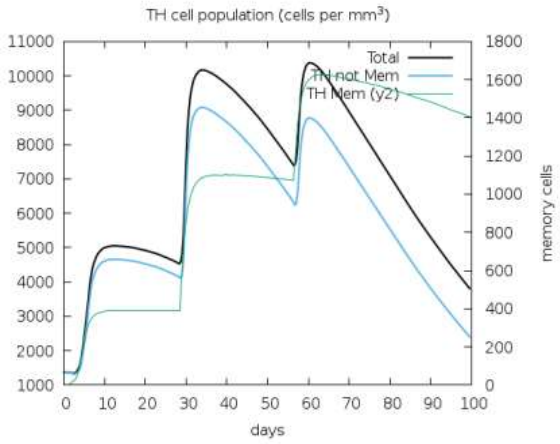


Figure 32: CD-4 HTL epitopes count. The plot shows the total and memory count.

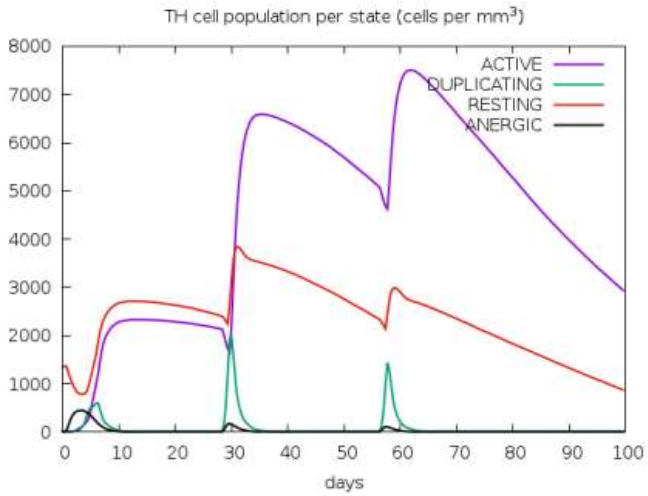


Figure 33: CD-4 HTL epitopes count subdivided into per entity-state

Figure 34-35 shows the potential CTL epitopes evolve in response to vaccination. The CTL epitope's CD-8+ concentration was more significant in memory and non-memory inducing states.

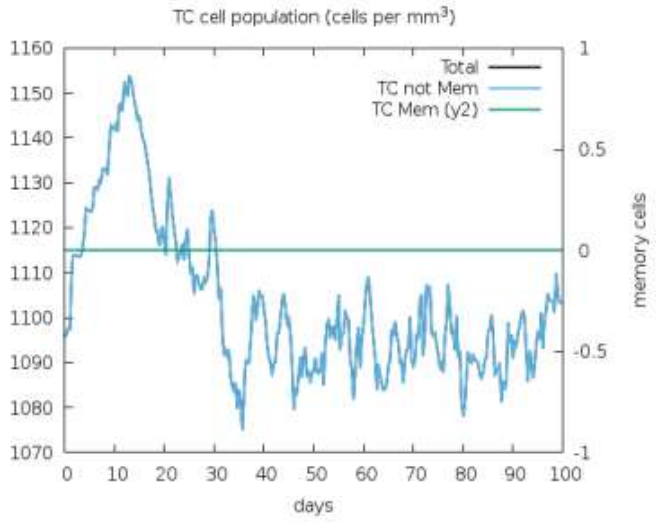


Figure 34: CTL total count (Total and Memory)

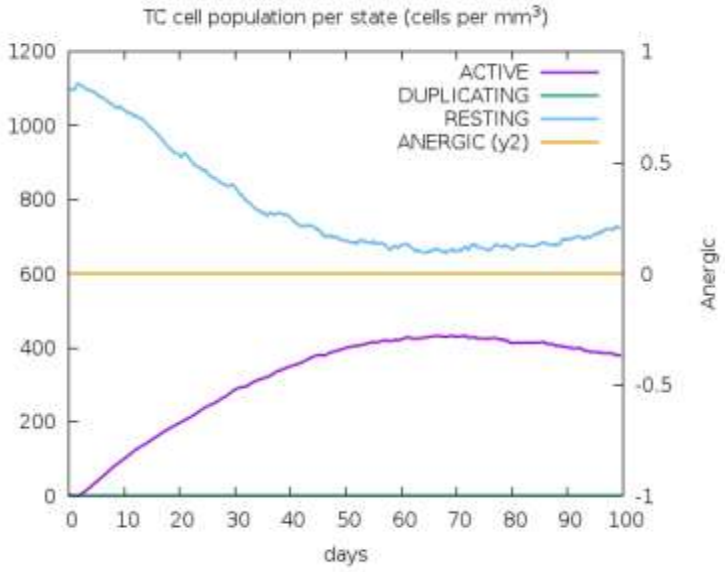


Figure 35: CTL count per entity state

C-Immsim server also shows the host Natural Killer (NK) cells population growth (Figure 36) that was observed on day-to-day basis of post vaccination.

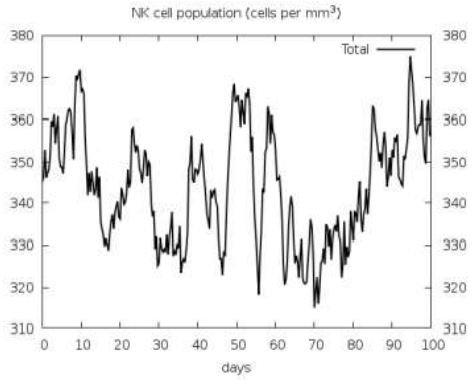


Figure 36: Total NK cell population count after vaccine administration

Dendritic cells (DC) represent antigenic peptides on both MHC class-l and class-11 molecules. The curves in figure 37 show the total number broken down to active, resting, internalized and presenting the ag.

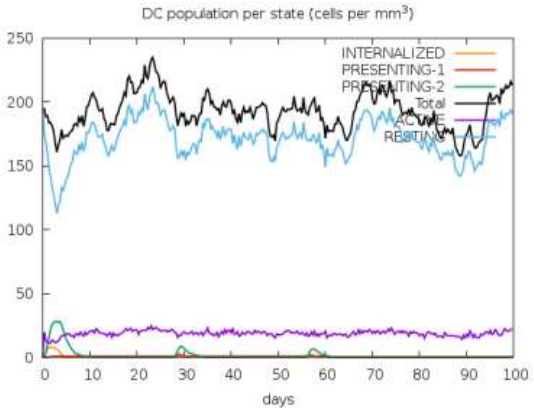


Figure 37: Antigenic peptides can be found in both MHC I and MHC II molecules in the DC population.

Graph of figure 38 shows Macrophage population growth observed by classification into active, resting, MHC II presenting and internalized groups along with total growth of post vaccination.

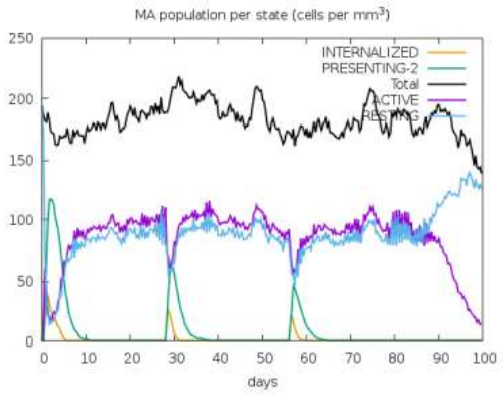


Figure 38: Macrophages population growth per entity state

The populations of interferon and interleukin and other substances that cause inflammation in the host are good places for a viral infection to spread. Graphs of figure 39 depict epithelial cell count after each state.

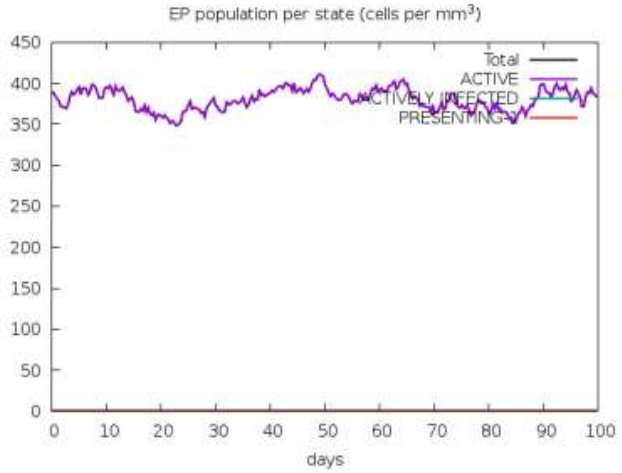


Figure 39: Epithelial cells total population count per entity-state

From figure 40, the graph depicts interferon and interleukin populations, together with other chemicals that produce inflammation within the host, are appropriate sites for viral infection spread.

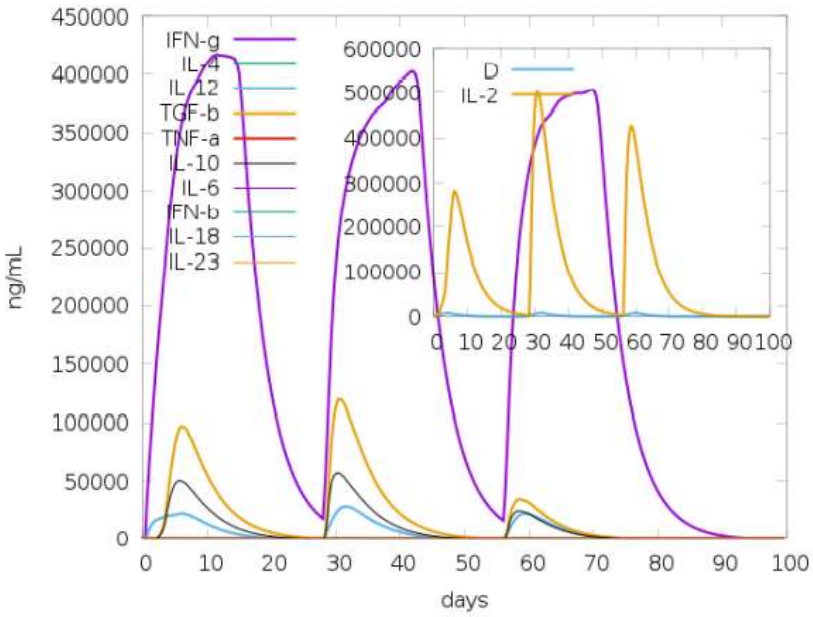


Figure 40: Concentration of cytokines and interleukins. In the inset figure, the danger signal is *shown with the leukocyte growth factor IL-2.*

Chapter 4

Discussion:

The advent of SARS-CoV-2 is a frightening condition for the entire population; hence treatments and preventative measures are critical. The SARS-CoV-2 virus lives in the lungs, causing fever, cough, and dyspnea. SARS-CoV-2 symptoms can appear within 2 to 24 days, according to the WHO, and the virus can be transferred from person to person or by contact with contaminated surfaces and objects [35]. Immune epitopes must be identified as soon as possible. Envelope (E) protein has the highest antigenicity. Also, it possesses a highly concentrated amino acid sequence compared to SARS-CoV 2's spike (S) protein, which has been undergone several amino acid sequence changes throughout 2020-2021. As a result, E-protein might be used as a vaccine target against SARS-CoV2 [36].

Selecting the E protein components exposed on the membrane surface can improve the specificity of "epitope-based vaccinations." [37]. SARS-CoV-2 vaccines are being developed by medical biotechnology regularly. In-silico immune-informatics, on the other hand, can save time and money, making it an essential approach in immunogenic analysis and vaccine development.

We used an 'In-silico' technique with rigorous criteria to discover E protein targeting B-cell and T-cell epitopes that may help promote immune response inside the host cell in this work. Using computational technology, we attempted to build an in-silico peptide-based vaccination against SARS-CoV-2, and we believe we have found a candidate capable of combating SARS-CoV-2 despite all efforts. The final vaccine that we constructed was unstable, so we believe that stability can be improved with the help of chaperons. Molecular chaperones are present in all species and are required for cell viability. One of the primary functions of molecular chaperones is to aid in

protein folding. Hsp60s, Hsp70s, Hsp90s, and sHsps are molecular chaperones that aid via stabilizing folding stages and avoiding protein aggregation and misfolding in unfolded and misfolded polypeptides [38].

In conclusion, we believe that the vaccine is still in its primary stage. More research needs to be done to create a vaccine for SARS-CoV-2 and ensure its safety in terms of in vivo, in vitro, and a clinical trial is a must.

References:

[1] A. Shanker, "The Possible Origins of the Novel Coronavirus SARS-CoV-2," 2020, doi: 10.31219/OSF.IO/A83R4.

[2] R. Lu *et al.*, "Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *Lancet*, vol. 395, no. 10224, pp. 565–574, Feb. 2020, doi: 10.1016/S0140-6736(20)30251-8.

[3] G. R *et al.*, "Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses," *Cell. Mol. Life Sci.*, vol. 78, no. 4, pp. 1655–1688, Feb. 2021, doi: 10.1007/S00018-020-03603-X.

[4] W. AC, P. YJ, T. MA, W. A, M. AT, and V. D, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281-292.e6, Apr. 2020, doi: 10.1016/J.CELL.2020.02.058.

[5] L. E. Gralinski and V. D. Menachery, "Return of the Coronavirus: 2019-nCoV," *Viruses 2020, Vol. 12, Page 135*, vol. 12, no. 2, p. 135, Jan. 2020, doi: 10.3390/V12020135.

[6] L. W *et al.*, "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus," *Nature*, vol. 426, no. 6965, pp. 450–454, Nov. 2003, doi: 10.1038/NATURE02145.

[7] H. M *et al.*, "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor," *Cell*, vol. 181, no. 2, pp. 271-280.e8, Apr. 2020, doi: 10.1016/J.CELL.2020.02.052.

[8] L. L *et al.*, "Epithelial cells lining salivary gland ducts are early target cells of severe acute respiratory syndrome coronavirus infection in the upper respiratory tracts of rhesus macaques," *J. Virol.*, vol. 85, no. 8, pp. 4025–4030, Apr. 2011, doi: 10.1128/JVI.02292-10.

[9] K. Knoops *et al.*, "SARS-Coronavirus Replication Are Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum," *PLOS Biol.*, vol. 6, no. 9, p. e226, Sep. 2008, doi: 10.1371/JOURNAL.PBIO.0060226.

[10] N. BW *et al.*, “A structural analysis of M protein in coronavirus assembly and morphology,” *J. Struct. Biol.*, vol. 174, no. 1, pp. 11–22, Apr. 2011, doi: 10.1016/J.JSB.2010.11.021.

[11] S. Lu, “Timely development of vaccines against SARS-CoV-2,” *Emerg. Microbes Infect.*, vol. 9, no. 1, p. 542, Jan. 2020, doi: 10.1080/22221751.2020.1737580.

[12] T. N and D. RK, “Immunoinformatics: a brief review,” *Methods Mol. Biol.*, vol. 1184, pp. 23–55, 2014, doi: 10.1007/978-1-4939-1115-8_3.

[13] E. Ortiz-Prado *et al.*, “Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review,” *Diagn. Microbiol. Infect. Dis.*, vol. 98, no. 1, p. 115094, 2020, doi: 10.1016/j.diagmicrobio.2020.115094.

[14] C. M, W. H, C. S, H. R, and C. J, “Evaluation of modified vaccinia virus Ankara based recombinant SARS vaccine in ferrets,” *Vaccine*, vol. 23, no. 17–18, pp. 2273–2279, Mar. 2005, doi: 10.1016/J.VACCINE.2005.01.033.

[15] W. Liu, Y. Lu, and Y. Chen, “Bioinformatics analysis of SARS-Cov M protein provides information for vaccine development *,” *Prog. Nat. Sci.*, vol. 13, no. 11, pp. 844–847, Nov. 2003, doi: 10.1080/10020070312331344530.

[16] D. S. Mishra, “T Cell Epitope-Based Vaccine Design for Pandemic Novel Coronavirus 2019-nCoV,” Apr. 2020, doi: 10.26434/CHEMRXIV.12029523.V2.

[17] I. A. Doytchinova and D. R. Flower, “VaxiJen: a server for predicting protective antigens, tumor antigens and subunit vaccines,” *BMC Bioinforma. 2007 81*, vol. 8, no. 1, pp. 1–7, Jan. 2007, doi: 10.1186/1471-2105-8-4.

[18] M. V Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, and M. Nielsen, “Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction,” *BMC Bioinforma. 2007 81*, vol. 8, no. 1, pp. 1–12, Oct. 2007, doi: 10.1186/1471-2105-8-424.

[19] L. MV *et al.*, “An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions,” *Eur. J. Immunol.*, vol. 35, no. 8, pp. 2295–2303, Aug. 2005, doi:

10.1002/EJI.200425811.

[20] O. Lund *et al.*, “Definition of supertypes for HLA molecules using clustering of specificity matrices,” *Immunogenetics*, vol. 55, no. 12, pp. 797–810, Mar. 2004, doi: 10.1007/S00251-004-0647-4.

[21] J. J. A. Calis *et al.*, “Properties of MHC Class I Presented Peptides That Enhance Immunogenicity,” *PLOS Comput. Biol.*, vol. 9, no. 10, p. e1003266, Oct. 2013, doi: 10.1371/JOURNAL.PCBI.1003266.

[22] M. Wieczorek *et al.*, “Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation,” *Front. Immunol.*, vol. 8, no. MAR, Mar. 2017, doi: 10.3389/FIMMU.2017.00292.

[23] P. S *et al.*, “Development and validation of a broad scheme for predicting HLA class II-restricted T cell epitopes,” *J. Immunol. Methods*, vol. 422, pp. 28–34, Jul. 2015, doi: 10.1016/J.JIM.2015.03.022.

[24] C. D. Russell, S. A. Unger, M. Walton, and J. Schwarze, “The Human Immune Response to Respiratory Syncytial Virus Infection,” *Clin. Microbiol. Rev.*, vol. 30, no. 2, p. 481, Apr. 2017, doi: 10.1128/CMR.00090-16.

[25] J. KK *et al.*, “Improved methods for predicting peptide binding affinity to MHC class II molecules,” *Immunology*, vol. 154, no. 3, pp. 394–406, Jul. 2018, doi: 10.1111/IMM.12889.

[26] J. MC, P. B, N. M, and M. P, “BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes,” *Nucleic Acids Res.*, vol. 45, no. W1, pp. W24–W29, Jul. 2017, doi: 10.1093/NAR/GKX346.

[27] M. Khan *et al.*, “Immunoinformatics approaches to explore Helicobacter Pylori proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine,” *Sci. Reports 2019 91*, vol. 9, no. 1, pp. 1–13, Sep. 2019, doi: 10.1038/s41598-019-49354-z.

[28] K. M *et al.*, “Template-based protein structure modeling using the RaptorX web server,” *Nat. Protoc.*, vol. 7, no. 8, pp. 1511–1522, Aug. 2012, doi: 10.1038/NPROT.2012.085.

[29] W. D *et al.*, “T3DB: the toxic exposome database,” *Nucleic Acids Res.*, vol. 43, no. Database issue, pp. D928–D934, Jan. 2015, doi: 10.1093/NAR/GKU1004.

[30] G. SM and J. JA, “Allergen sequence databases,” *Mol. Nutr. Food Res.*, vol. 50, no. 7, pp. 633–637, Jul. 2006, doi: 10.1002/MNFR.200500271.

[31] K. LA, M. S, Y. CM, W. MN, and S. MJ, “The Phyre2 web portal for protein modeling, prediction and analysis.,” *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, May 2015, doi: 10.1038/NPROT.2015.053.

[32] W. A *et al.*, “SWISS-MODEL: homology modeling of protein structures and complexes,” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, Jul. 2018, doi: 10.1093/NAR/GKY427.

[33] S.-D. D, I. Y, N. R, and W. HJ, “PatchDock and SymmDock: servers for rigid and symmetric docking,” *Nucleic Acids Res.*, vol. 33, no. Web Server issue, Jul. 2005, doi: 10.1093/NAR/GKI481.

[34] C. Rueckert and C. A. Guzmán, “Vaccines: From Empirical Development to Rational Design,” *PLOS Pathog.*, vol. 8, no. 11, p. e1003001, Nov. 2012, doi: 10.1371/JOURNAL.PPAT.1003001.

[35] C. N *et al.*, “Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study,” *Lancet (London, England)*, vol. 395, no. 10223, pp. 507–513, Feb. 2020, doi: 10.1016/S0140-6736(20)30211-7.

[36] S. Bhattacharya, A. Banerjee, and S. Ray, “Development of new vaccine target against SARS-CoV2 using envelope (E) protein: An evolutionary, molecular modeling and docking based study,” 2020, doi: 10.1016/j.ijbiomac.2020.12.192.

[37] Ghana, P. Farnia, and H. Ghomi, “The effectiveness of cold atmospheric plasma by inhaling anesthetic mask or through bronchoscopy against COVID-19,” *Biomed. Biotechnol. Res. J.*, vol. 4, no. 1, p. 1, Jan. 2020, doi: 10.4103/BBRJ.BBRJ_27_20.

[38] J. L. Camberg, S. M. Doyle, D. M. Johnston, and S. Wickner, “Molecular Chaperones,” *Brenner’s Encycl. Genet. Second Ed.*, pp. 456–460, Feb. 2013, doi: 10.1016/B978-0-12-374984-0.00221-7.