A Novel Vaccine Designing Approach Through Analyzing Computational Tools Against Pathogenic *Vibrio Cholerae*'s Outer Membrane Protein of O1 Strain as Adjuvant

By

Ashraf Hossain
ID: 17346023

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
October 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:

_____

ASHRAF HOSSAIN

Student ID: 17346023

Approval

The thesis titled 'A Novel Vaccine Designing Approach Through Analyzing Computational Tools Against Pathogenic *Vibrio Cholerae*'s Outer Membrane Protein of O1 Strain as Adjuvant' submitted by Ashraf Hossain (ID: 17346023) of Summer 2021 has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Bachelor of Pharmacy (Hons).

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

Assistant Dean:
(Member)

Professor Hasina Yasmin, PhD
Assistant Dean and Program Director,
School of Pharmacy
Brac University

Departmental Head:
(Dean)

Professor Eva Rahman Kabir, PhD
Dean, School of Pharmacy
Brac University

Ethics Statement

This study does not include any human or animal trial.

Acknowledgments

Words cannot express the gratitude and thankfulness to my supervisor, Mohammad Kawsar Sharif Siam, who recognized my strengths, and directed me with his comprehensive knowledge of supervising. In addition, I'd like to admire our Dean of School of Pharmacy, Professor Dr. Eva Rahman Kabir for facilitating the importance of undergrad project.

Ashraf Hossain
ID: 17346023
School of Pharmacy
Brac University

Abstract

Cholera epidemic, history denotes it as “The Blue Death", a deoxygenated, dehydrated death of humans due to the bacterial species named *Vibrio Cholerae.* When the sanitation system is extremely devastated by the purity of water and lacks the knowledge of hygiene, inhabitants make themselves vulnerable to this water and airborne microbe, *Vibrio Cholerae.* After getting affected by *Vibrio Cholerae*, a person gets desiccated due to water loss and infecting intestinal lumen (small and large intestine), if it stays untreated, the death of the person might occur in a few hours. Studying this severely spreading bacteria species and the extremity of the causing disease, an “In Silico" approach is made to prepare a vaccine against *Vibrio Cholerae.* Several computational servers and tools were profoundly used during the preparation and justified their results. An In Silico" approach is made to prepare a vaccine against Vibrio Cholerae. Several computational servers and tools were profoundly used during the preparation. The antigenicity result of the selected "Outer Membrane Protein Sequence" is 0.5469, stated as "Probable ANTIGEN". A proposed vaccine was then developed using linkers to be added and tested for antigenicity using Vaxijen V2.0 and the result was 0.6203. After that, the molecular docking was done using Patchdock and achieved a score of 20884 in a 3101.90 square angstrom region.

Key Words: In-Silico, *Vibrio Cholerae*, Antigenicity, Vaccine Design, Molecular Docking

Table of Contents

List of Figures

Figure 18 (c)	Graph showing entity-state of B cells versus days after vaccine administration.
Figure 18 (d)	B cell population expansion in plasma vs. vaccination treatment day
Figure 18 (e)	CD-4 HTL epitopes count. The plot shows the total and memory count.
Figure 18 (f)	CD-4 HTL epitopes count subdivided into per entity-state
Figure 18 (g)	CTL total count (Total and Memory)
Figure 18 (h)	CTL count per entity state
Figure 18 (i)	Total NK cell population count after vaccine administration
Figure 18 (j)	Antigenic peptides can be found in both MHC I and MHC II molecules in the DC population
Figure 18 (k)	Macrophages population growth per entity state
Figure 18 (l)	Epithelial cells total population count per entity-state
Figure 18 (m)	Concentration of cytokines and interleukins. In the inset figure, the danger signal is shown with the leukocyte growth factor IL-2

List of Tables

Chapter 1: Introduction

1.1: *Vibrio Cholerae*:

V. Cholera, the pathogenic bacteria which infects the human intestinal area, both large and small intestines. Around 1.3 to 4 billion people are getting worse scenarios due to death and getting affected by cholera, the disease caused by *Vibrio Cholerae*, where 21 000 to 143 000 people accept death, yearly (*Non-Cholera Vibrio Cholerae Infections | Cholera | CDC*, n.d.). These are people of 69 countries that are stated as the endemic zone of cholera. Apart from these countries, there are another 99 million people in three countries where the inhabitants of India, Bangladesh, China, Ethiopia, and Nigeria are experiencing the crucial threat of the pathogen (*Non-Cholera Vibrio Cholerae Infections | Cholera | CDC*, n.d.). Cholera is a waterborne disease, caused due to the intake of contaminated water, as it spreads through water resources. When a cholera-affected person's fecal matter mixes with edible water resources, the spread of Vibrio Cholerae occurs, contaminating the water, and other people taking this water as a resource, might not purify water properly. As a result, the tremendous risk of getting affected by Vibrio Cholerae pathogen rises till getting infected, and the symptoms start with severe watery diarrhea along with vomiting, through which an infected person can quickly lead to dehydration. (*Gram Negative Bacteria Concept & Examples | What Is Gram Negative Bacteria? | Study.Com*, n.d.)

1.2: Structure and Genome:

Vibrios are Gram-negative bacteria, having their peptidoglycan composed only of canonical muropeptides, having twisted rod like shapes with an isolated hydrophilic appendage which has higher movement functions. They can tolerate alkaline medium, which kills the majority of helpful pathogens in the intestine, the bacteria which are likely to induce protective responses that prevent colony formation and invasion of harmful intestinal microbes since they are acid-sensitive (*Gram Negative Bacteria Concept & Examples | What Is Gram Negative Bacteria? | Study.Com*, n.d.). There is a plethora of free-living vibrios known, some of which are undesirably harmful to our body. Serogroups of V. cholerae strains are defined by the structure of their cell surface lipopolysaccharides, a toxin located inside the bacterial cell. Until 1992, only two serotypes of toxigenic group O 1 V cholera, Inaba (AC) and Ogawa (AB), and two biotypes, classical and El Tor, induced cholera disease into human races. These organisms can be identified by agglutination in O group 1-specific antiserum directed against the lipopolysaccharide component of the cell wall

and indications of their toxin compound producibility. Cholera caused by serogroup O139 (also known as "Bengal," the 139th and most recently discovered serogroup of V cholera) became common in India and Bangladesh since 1992.

1.3: Life cycle and pathogenesis of Vibrio Cholerae:

Cholera is a disease that produces worldwide epidemics and pandemics. A local outbreak can easily become a pandemic as its causing pathogen's, V. cholerae, life cycle enables the pathogen to endure for years in an aquatic environment, a friendly environment for getting reserved naturally, where it adheres to crustaceans, the diverse group of invertebrates like algae, and zooplankton, crabs, lobsters, shrimp, etc. (R. Khan et al., 2019). V. cholerae regularly reproduces and keeps regenerating if the favorable conditions are met. Though if the environment is hostile and threatening to its existence, this disease can remain unconscious, inactive, uncultivable, and chlorine resistant, which diminishes its risk of vulnerability. (Conner et al., 2016)

Cholera-causing comma-shaped bacteria V. cholerae is a severe and possibly deadly diarrheal disease. This disease condition is contracted by ingesting food or water infected with this bacterium. Cholera has nearly passed from sight of affluent countries as maintenance of high sanitation standards and water grade; yet, the disease continues to be a threat in many underdeveloped countries that lack the necessary infrastructure and sanitation. Floods and conflict, which allow to fecal contamination of water reservoir and resources, are frequently related to and exacerbated by disease epidemics (Almagro-Moreno & Taylor, 2013)

Chapter 2: Methodology

2.1: Workflow

A glimpse of the workflow of designing an "In Silico" vaccine with the use of compatible computational tools is given below, using a flowchart. This corresponds to the overview of the whole vaccine design process along with the physicochemical and immunological examination of the proposed vaccine. (Raheem et al., 2021)

Figure 1: Schematical presentation of developing a multi-epitope vaccination against Vibrio Cholerae.

2.2: Retrieval of Protein Sequence:

The UniProt Knowledgebase (UniProtKB) was used to search and explore the proteome of Vibrio Cholerae, specifically the "Outer Membrane Protein Sequence" of the pathogen, screened based on the proteomic antigenicity (Mulder et al., 2007). The outer membrane proteins are the one severely infectable part of gram-negative bacteria like Vibrio Cholerae to its host animal, functionally, its outer membrane is more invadable towards the eukaryotic cell, making it a pickable concern to design a vaccine against its pathogen, (van der Pol et al., 2015) Vibrio Cholerae. After several antigenicity tests, a suitable protein sequence was discovered, showing excellent antigenicity. Firstly, the outer membrane protein's genome and protein sequences were retrieved from the UniProt Knowledgebase database in fasta format. The Vaxijen v2.0 server determined the protein's probable antigenicity (Doytchinova & Flower, 2007). "Bacteria" was chosen as the target organism, and a threshold of 0.5 was set at the server before starting the test. Because of its validity, length, and quality, the selected protein sequence was used to carry out the further process of constructing the vaccine. The protein sequence was then submitted for epitope prediction for helper T-lymphocytes (HTL) and cytotoxic T-lymphocytes (CTL).

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

NetCTL has been shown to have the excellent predictive ability of CTL epitopes. Vibrio Cholerae's CTL epitopes were predicted with the NetCTL 1.2 server, with good sensitivity and specificity at 0.75 (Stranzl et al., 2010). MHC class I binding epitopes were predicted, and to expect it, the A1 supertype was chosen using artificial neural networks. Using a IC50 dose, a measurement of the potency of substance that is going to be used, of 500 nm with a combined score as a guideline, the most promising Vibrio Cholerae vaccine development options were selected (Calis et al., 2013). The epitope appears to have a considerable affinity for the receptor, as indicated by the IC50 value of 500 nm. The integrated score, class I binding, TAP transport efficiency, and proteasomal cleavage prediction were all used to make predictions. TAP transfers peptides from the cytosol towards the endoplasmic reticulum, allowing it to select similar peptides in length and sequence to MHC class I molecules. At 0.15 and 0.05, respectively, we weighed C-terminal cleavage and TAP transport efficacy.

2.4: MHC I Alleles identification:

We discovered MHC class I corresponding alleles specific to CTL epitopes using the NetMHC

Pan 4.1 server. The stronger an epitope's binding affinity for an allele is, the lower its percentile rank. Epitopes with a percentile level of the threshold, ranging up to 0.500, for strong binding peptides, and for weak binding peptides greater than 0.500 to 2.000 threshold range were used. And the epitopes with a higher percentile rank than this threshold were no longer included in the vaccine formulation. Finally, used the NetCTL server's input to determine how strong a binding affinity is (Harding & Unanue, 1990).

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

HTL epitopes in the protein sequence were predicted using the NetMHC II pan 4.0's MHC-II epitope prediction module with its default settings. The main antigen was employed, with a peptide length of 15 specified as the baseline of the tool (Álvaro-Benito et al., 2018). 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, which was denoted as "SB", abbreviated as "Strong Binder". The percentile rank was set as 1 for a strong binder and up to 5 was set for weak binders as a threshold (Reynisson et al., 2020).

2.6: Cytokine inducing capability of predicted HTL Epitopes:

Interferon-gamma (IFN-γ) is a cytokine that is important for both innate and adaptive immunity. It is a key activator of macrophages and also stimulates natural killer cells and neutrophils. 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 purposefully to generate the results needed (Jensen et al., 2018).

2.7: Screening of B-cell Epitopes:

The BepiPred, a linear antigenic determinant server, was used to foretell Linear B cell epitopes (*B Cell Tools*, n.d.). At a threshold of 0.5, the protein's linear B cell epitopes were expected (Jespersen et al., 2017).

2.8: Construction of the vaccine:

By combining all the tested CTL, HTL, and B-cell epitopes, a multi-epitope polypeptide vaccine was constructed. To enhance the vaccine's immune response, the adjuvant chemical outer membrane protein was utilized. The adjuvant was tethered to the multi-epitope polypeptide's N

terminus using an EAAAK linker, allowing for suitable functional domain spacing and effective development and detection by the host immune system. Combining CTL, HTL, and B-cell epitopes via AAY and GPGPG linkers enhances immunogenicity and epitope expression, resulting in molecular vaccination efficiency (M. Khan et al., 2019).

2.9: Biochemical Analysis of the Constructed Vaccine:

The ProtParam tool was employed to further assess the physical properties and chemical features of the proposed vaccine candidate. A brief inclusion of the count of amino acids, molecular mass, chemical formula, atomic composition, amino acid composition, extinction coefficients, instability index, estimated half-life, grand average of hydropathicity (GRAVY), and amino acid composition, among the physicochemical features, was examined. The theoretical pI and molecular weight were calculated, dropping the sequence and the atomic and amino acid constitution were clearly evident. The extinction coefficient of the protein sequence was calculated using the details on resulted amino acid makeup (Gasteiger et al., 2005).

2.10: Prediction of Toxicity and Allergenicity:

The toxicity prediction of the proposed vaccine was made with the Toxin and Toxin Target Database (T3DB). The 'toxic exposome,' as defined by the T3DB, is the entire array of habitually or acutely harmful substances underneath the proposed sequence to which humans may be harmed while exposed (Wishart et al., 2015). The allergic inducibility of the vaccine ought to be non-allergic as allergenic proteins cause a negative susceptible response to the immunity whilst the intake. Based on the Codex Alimentarius Commission's standards, AllergenOnline is a useful peer-reviewed tool for determining the key potential risks of allergy to GMOs and novel foods (Gendel & Jenkins, 2006).

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

The proposed vaccine was a renovated protein sequence with no previously discovered homology known or stored in any database. Phyre2 uses a structure-based folding simulator to simulate sections that have no visible relative similarity. The Phyre 2 server was used to estimate the three-dimensional structure of the planned vaccination. The tool uses several templates and a simple structure-based folding simulation to create a completely three-dimensional model of the inputted protein sequence (Mezulis et al., 2015).

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

To analyze the developed vaccine's tertiary structure, the SWISS-MODEL workstation generated a Ramachandran plot [32]. In protein structure, the Ramachandran plot identifies ideal places for amino acid residue backbone dihedral angles. The Structure Assessment page provides the highest Molprobity scores and allows us to rapidly detect low-quality residues in the system or model. The vaccine's protein structure was then validated using the ProSA-web program. A positive Z-score implies that a 3D protein model piece that has been constructed is erroneous or unpredictable (Waterhouse et al., 2018).

2.13: Vibrio Cholerae Vaccine's Molecular Docking with Related Antigenic Recognition Receptors:

The toll-like receptor-4's (TLR4), 3ULA, Crystal structure of the human TLR2-Diprovocim complex, was used as antigenic recognition receptor and the immune cell's major histocompatibility complex that the vaccine construct binds to were determined (Han et al., 2012). The cluspro v2.0 server was utilized to confirm and check how closely the constricted vaccination bind between the receptors (Nezafat et al., 2016).

2.14: Immune Simulations:

The C-IMMSIM server assesses the vaccine's immune response and immunogenicity. In the C-ImmSim, the Celada-Seiden model is utilized to characterize the profiles of mammalian immune systems, both humoral and cellular, in response to a given vaccination. The simulation was run with the default settings, and it took 300 steps to finish. A tri-dosage approach was applied in injection at stages 1, 84, and 168. The vaccine, on the other hand, was scheduled to be given three times at 28-day intervals (Banagozar et al., 2019)

2.15: Remarks on the Materials and Method:

The research work was extensively investigative, and conducted utilizing the in-silico approach, which entails that all predictions and analyses were performed using internet servers. It can't be guaranteed that product would be a very effective antibacterial eradicating vaccine against Vibrio Cholerae. Yet, the research is needed since it has the potential to become a vaccine candidate (Yang et al., 2021).

Chapter 3: Result

3.1: Antigenicity prediction of Outer Membrane Protein (OMP):

Amongst the two serogroups (O1 and O139) or categories of the pathogenic Vibrio Cholerae bacteria, that can turn out cholera toxin which seeds the disease we name cholera https://www.cdc.gov/cholera/non-01-0139-infections.html, the O1 serotype is found and used to bring out its potential and antigenic "Outer Membrane Protein". This antigenic part of the serotype O1 is collected by a meticulous screening method, method of checking antigenicity. The full amino acid sequence of Outer Membrane Protein of Vibrio Cholerae O1 serotype is conserved in protein FASTA format. Using the format, the protein is presented below;

Outer membrane protein TolC OS=Vibrio Cholerae serotype O1 (strain ATCC 39315 / El Tor Inaba N16961) OX=243277 GN=tolC PE=1 SV=1 https://www.uniprot.org/uniprot/Q9K2Y1

MKKLLPLFVSAALGTLSSAVWAENLAEIYNQAKENDPQLLSVAAQRDAAFEAVTSSRS
ALLPQINLTAGYNINRSDQAPRESDLLSAGINFSQELYQRSSWVSLDTAEKKARQADSQ
YAATQQGLILRVAKAYFEVLRAQDNLEFVRAEKAAVGRQLEQTKQRFEVGLSAITDVH
DAQAQFDGVLADEVLAENSLTNSYEALREITGQEYSKLAVLDTKRFAASRTTESSEALIE
KAQQQNLSLLAARISQDVARDNISLASSGHLPSLTLDGGYNYGNNSNDNAKNTSGEEY
NDFKIGVNLKVPLYTGGNTTSLTKQAEFAYVAASQDLEAAYRSVVKDVRAYNNNINAS
IGALRAYEQAVISAKSALEATEAGFDVGTRTIVDVLDATRRLYDANKNLSNARYDYILS
VLQLRQAIGTLSEQDVMDVNAGLKVAKK

After punctilious collection of the protein sequence, it is run and examined for the result of its antigenicity through the Vaxijen v2.0 server and achieved a score of 0.5453 (Figure 2). The result interprets that the ran sequence was a probable antigen, as its antigenicity is recorded as greater than the threshold level, it was selected for further processes of In-Silico vaccine preparation.

VaxiJen RESULTS
Selected Model: **Bacteria**
Threshold for this model: **0.5**
Overall Prediction for the Protective Antigen= **0.5453 (Probable ANTIGEN)**

Figure 2: "VaxiJen v2.0 server antigenicity score of candidate's proteomes." (Doytchinova & Flower, 2007).

3.2: Identification of CTL epitopes:

3.2.1: MHC Class I Corresponding Alleles

The multi-epitope vaccine preparation starts with achieving Cytotoxic T-Lymphocyte (CTL) epitopes of A1 supertype of major histocompatibility complex (MHC) I, using the protein fasta format of "Outer Membrane protein of Vibrio Cholerae O1 Serotype", with a threshold of 0.75, from NetCTL-1.2 Server. According to the server's combined score, there had been 14 collectively gained A1 Supertypes which were used to run and get the MHC Class I corresponding Alleles.

Table 1: Combined Score and Outcomes of CTL prediction on the NetCTL-1.2 server.

CTL Epitopes	**Combined Score**
WAENLAEIY	2.7858
VLDATRRLY	2.0252
LTLDGGYNY	1.8004
LTKQAEFAY	1.6666
ASQDLEAAY	1.5985
GINFSQELY	1.2830
NLSNARYDY	1.2622
PSLTLDGGY	1.0546
GVNLKVPLY	0.9458

VMDVNAGLK	0.8986
TTESSEALI	0.8732
LREITGQEY	0.8534
LSEQDVMDV	0.8069
AKNTSGEEY	0.7620

3.2.2: MHC I alleles specific to CTL epitopes:

After retrieving the 14 A1 supertypes from NetCTL-1.2 server, their corresponding alleles had been collected from the NetMHC Pan 4.1 server, as Cytotoxic T-Lymphocyte (CTL) epitopes. In this state, the percentile rank is a criterion used in epitope selection, and a greater binding affinity showing epitope is indicated by a lower percentile score and readily selected as less as their value. Whilst selecting epitopes, a minimum threshold of 2.0 was set as a benchmark in this screening. Below is a list of CTL epitopes and MHC I allele-specific binding and the associated binding affinity in percentile rank. The notable points, in which the thresholds were put to denote the strong binders and weak binders from all the inputs, are as like the following criteria;

Strong binding peptides have a threshold of 0.500 Weak binding peptides have a point of 2.000.

Table 2: MHC I allele for specific epitopes along with sequence number, length, their percentile rank and, Score_EL, Bind Level

Allele	**Peptide**	**Length**	**% Rank**	**Score_ EL**	**Bind Level**
HLA-A*01:01	WAENLAEIY	9	0.17	0.6670340	SB
HLA-A*01:01 HLA-A*03:01 HLA-B*15:01	VLDATRRLY	9	0.062 1.061 0.642	0.8855400 0.2364700 0.4045310	SB WB

Allele	Peptide	Length	% Rank	Score_ EL	Bind Level
HLA-A*01:01	LTLDGGYNY	9	0.14	0.7280350	SB
HLA-A*26:01			0.274	0.3645170	SB
HLA-B*58:01			0.564	0.4622460	WB
HLA-B*15:01			0.800	0.3379100	WB
HLA-A*01:01	LTKQAEFAY	9	0.421	0.3780980	SB
HLA-A*26:01					
HLA-A*01:01	ASQDLEAAY	9	0.115	0.7741980	SB
HLA-B*15:01			0.349	0.5771470	SB
HLA-A*01:01	GINFSQELY	9	0.391	0.4007470	SB
HLA-B*15:01			0.499	0.4768780	
HLA-A*01:01	GVNLKVPLY	9	0.732	0.2081660	WB
HLA-A*03:01			0.699	0.3677430	WB
HLA-A*26:01			0.526	0.2112340	WB
HLA-B*15:01			0.438	0.5153090	SB

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

VaxiJen v2.0 was employed to foresee all survived epitopes' immune responsiveness, and three of the epitopes were determined to be antigenic (Fig). AllerTOP v2.0 was used to identify the allergenicity of the Cytotoxic T-Lymphocyte cell epitopes, and three of the three CTL epitopes were projected to be non-allergenic. Furthermore, 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 (Fig). And vigorously all of the CTL epitopes

resulted in non-toxic substances and let them use as best fit CTL epitopes.

Table 3: Antigenicity prediction of CTL epitopes

CTL Epitopes	**Antigenicity Predictions**
WAENLAEIY	-0.4070 (Probable NON-ANTIGEN).
VLDATRRLY	-1.0413 (Probable NON-ANTIGEN).
LTLDGGYNY	0.2955 (Probable NON-ANTIGEN).
LTKQAEFAY	0.0958 (Probable NON-ANTIGEN).
ASQDLEAAY	0.7823 (Probable NON-ANTIGEN).
GINFSQELY	0.5313 (Probable NON-ANTIGEN).
GVNLKVPLY	1.5073 (Probable NON-ANTIGEN).

Table 4: Allergenicity result of selected CTL epitopes

CTL Epitopes	**AllerTOP v. 2.0**
ASQDLEAAY	PROBABLE NON-ALLERGEN
GINFSQELY	PROBABLE NON-ALLERGEN
GVNLKVPLY	PROBABLE NON-ALLERGEN

Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq_1	ASQDLEAAY	-1.48	Non-Toxin	-0.11	-0.38	0.10	-2.00	987.11
seq_1	GINFSQELY	-1.33	Non-Toxin	-0.02	-0.21	-0.52	-1.00	1070.30
seq_1	GVNLKVPLY	-1.17	Non-Toxin	0.06	0.59	-0.63	1.00	1002.36

1/1 50

Figure 4: Toxic epitope prediction using ToxinPred server, (Álvaro-Benito et al., 2018).

3.3.1: MHC II alleles specific to HTL epitopes:

The second type of T-Cell epitopes, named Helper T-Lymphocytes were retrieved from NetMHCIIpan 4.0 server, in which the core antigenic part (Outer Membrane Protein) as an input to detect MHC II alleles. MHC II alleles can be identified by considering the criteria of the percentile rank; a greater binding affinity is indicated by a lower percentile rank, the strong binders were collected and preserved for further credibility checking percentile rank, and vice versa. For the study, only the best binding peptides were selected. For allele identification, a percentile rank of 0.5 was used in this example. The Strong binding peptides threshold (%Rank) 1% The Weak binding peptides threshold (%Rank) is 5%.

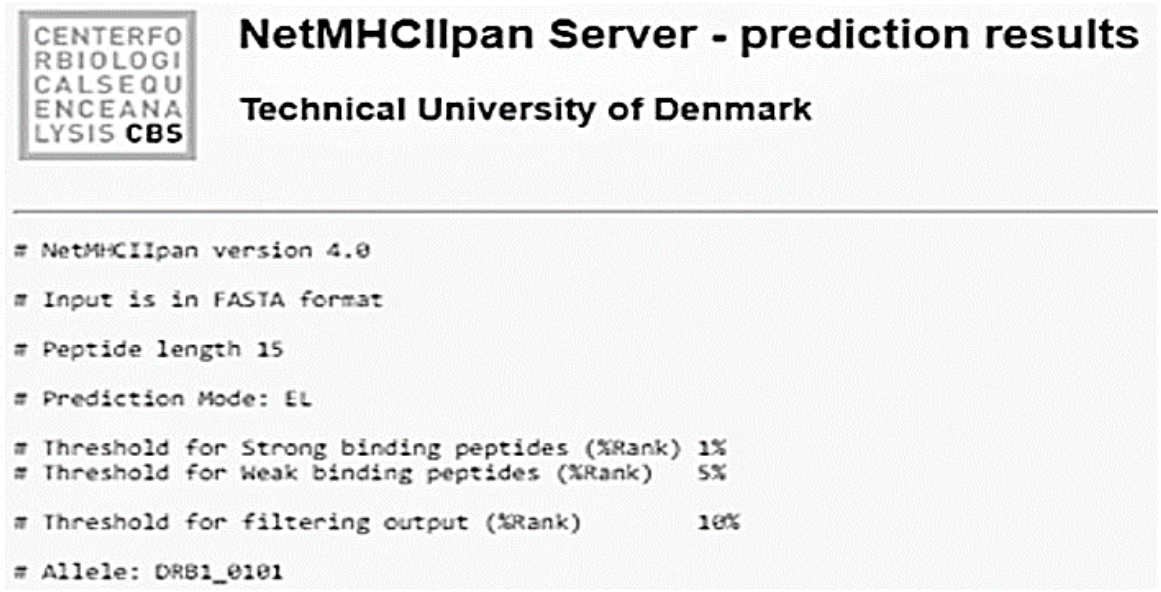


Figure 6: NetMHCIIpan 4.0 to discover MHC II specific to HTL epitopes, (Reynisson et al., 2020).

The Helper T-lymphocyte cells are preserved according to their strong binding affinity, which demonstrate their binding ability to most regions as alleles, along with eluted ligand mass spectrometry (EL) score.

Table 5: Strongest binding HTL specific allele, their core, Score-EL, and percentile rank.

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

The toxicity of the retrieved HTL epitopes of the "Outer membrane protein sequence" was checked using the ToxinPred server, where all of the epitopes were found non toxin and were allowed to use as input for further screening.

Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq_1	VYSRVKNLNSSRVPD	-0.80	Non-Toxin	-0.35	-0.82	0.31	2.00	1734.15
seq_1	YSRVKNLNSSRVPDL	-0.66	Non-Toxin	-0.35	-0.85	0.29	2.00	1748.18
seq_1	PSFYVYSRVKNLNSS	-1.17	Non-Toxin	-0.20	-0.52	-0.29	2.00	1761.17
seq_1	KPSFYVYSRVKNLNS	-1.25	Non-Toxin	-0.26	-0.73	-0.11	3.00	1802.27
seq_1	FYVYSRVKNLNSSRV	-1.03	Non-Toxin	-0.26	-0.38	-0.21	3.00	1832.30
seq_1	SFYVYSRVKNLNSSR	-0.98	Non-Toxin	-0.31	-0.71	-0.09	3.00	1820.24
seq_1	FYVYSRVKNLNSSRV	-1.03	Non-Toxin	-0.26	-0.38	-0.21	3.00	1832.30
seq_1	SFYVYSRVKNLNSSR	-0.98	Non-Toxin	-0.31	-0.71	-0.09	3.00	1820.24
seq_1	PSFYVYSRVKNLNSS	-1.17	Non-Toxin	-0.20	-0.52	-0.29	2.00	1761.17
seq_1	KPSFYVYSRVKNLNS	-1.25	Non-Toxin	-0.26	-0.73	-0.11	3.00	1802.27
seq_1	VKPSFYVYSRVKNLN	-1.31	Non-Toxin	-0.20	-0.39	-0.23	3.00	1814.33
seq_1	SFYVYSRVKNLNSSR	-0.98	Non-Toxin	-0.31	-0.71	-0.09	3.00	1820.24
seq_1	KPSFYVYSRVKNLNS	-1.25	Non-Toxin	-0.26	-0.73	-0.11	3.00	1802.27
seq_1	VKPSFYVYSRVKNLN	-1.31	Non-Toxin	-0.20	-0.39	-0.23	3.00	1814.33
seq_1	PSFYVYSRVKNLNSS	-1.17	Non-Toxin	-0.20	-0.52	-0.29	2.00	1761.17
seq_1	KPSFYVYSRVKNLNS	-1.25	Non-Toxin	-0.26	-0.73	-0.11	3.00	1802.27
seq_1	VKPSFYVYSRVKNLN	-1.31	Non-Toxin	-0.20	-0.39	-0.23	3.00	1814.33
seq_1	VYSRVKNLNSSRVPD	-0.80	Non-Toxin	-0.35	-0.82	0.31	2.00	1734.15
seq_1	YSRVKNLNSSRVPDL	-0.66	Non-Toxin	-0.35	-0.85	0.29	2.00	1748.18
seq_1	VYSRVKNLNSSRVPD	-0.80	Non-Toxin	-0.35	-0.82	0.31	2.00	1734.15

1/1 50

Figure 5: Toxic HTL epitope prediction using ToxinPred server, (Álvaro-Benito et al., 2018)

3.3.2: 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.

The retrieved data from IL-4Pred, which predicts the interleukin-4 inducing peptides for the vaccine, are shown below. This data demonstrates the inducers and SVM score, hydrophobicity, hydropathicity, hydrophilicity, and molecular weight of them accordingly.

The retrieved data from IL-10Pred, which predicts the interleukin-10 inducing peptides for the vaccine, are shown below. This data demonstrates the inducers and scores them accordingly.

Table 6: IFN, IL4, IL10 epitope prediction for HTL

Sequence	**Method**	**Result**		
		IFN	**IL4**	**IL10**
QDNLEFVRAEKAAVG	SVM	POSITIVE	Inducer	Inducer
AQDNLEFVRAEKAAV	SVM	POSITIVE	Inducer	Inducer
DNLEFVRAEKAAVGR	SVM	POSITIVE	Inducer	Inducer
RAQDNLEFVRAEKAA	SVM	POSITIVE	Inducer	Inducer
QQNLSLLAARISQDV	SVM	POSITIVE	Inducer	Inducer
QQQNLSLLAARISQD	SVM	POSITIVE	Inducer	Inducer
NLEFVRAEKAAVGRQ	SVM	POSITIVE	Inducer	Inducer
LEFVRAEKAAVGRQL	SVM	POSITIVE	Inducer	Inducer
QAVISAKSALEATEA	SVM	POSITIVE	Inducer	Inducer
SSWVSLDTAEKKARQ	SVM	POSITIVE	Inducer	Inducer
AYRSVVKDVRAYNNN	SVM	POSITIVE	Inducer	Non-inducer
YRSVVKDVRAYNNNI	SVM	POSITIVE	Inducer	Non-inducer
RSSWVSLDTAEKKAR	SVM	POSITIVE	Inducer	Non-inducer
LTAGYNINRSDQAPR	SVM	POSITIVE	Inducer	Non-inducer
QRDAAFEAVTSSRSA	SVM	POSITIVE	Inducer	Inducer
DAAFEAVTSSRSALL	SVM	POSITIVE	Inducer	Inducer

			Result	
Sequence	**Method**	**IFN**	**IL4**	**IL10**
VVKDVRAYNNNNINAS	SVM	POSITIVE	Inducer	Non-inducer
SVVKDVRAYNNNNINA	SVM	POSITIVE	Inducer	Non-inducer
KDVRAYNNNNINASIG	SVM	POSITIVE	Inducer	Non-inducer
EVGLSAITDVHDAQA	SVM	POSITIVE	Inducer	Non-inducer
VGLSAITDVHDAQAQ	SVM	POSITIVE	Inducer	Non-inducer
EQAVISAKSALEATE	SVM	POSITIVE	Inducer	Inducer
QAEFAYVAASQDLEA	SVM	POSITIVE	Inducer	Non-inducer
KQAEFAYVAASQDLE	SVM	POSITIVE	Inducer	Inducer
YEQAVISAKSALEAT	SVM	POSITIVE	Inducer	Non-inducer
AEFAYVAASQDLEAA	SVM	POSITIVE	Inducer	Non-inducer
DGGYNYGNNSNDNAK	SVM	POSITIVE	Inducer	Non-inducer
DPQLLSVAAQRDAAF	SVM	POSITIVE	Inducer	Non-inducer
ADSQYAATQQGLILR	SVM	POSITIVE	Inducer	Non-inducer
DSQYAATQQGLILRV	SVM	POSITIVE	Inducer	Inducer
AKAYFEVLRAQDNLE	SVM	POSITIVE	Inducer	Inducer
RVAKAYFEVLRAQDN	SVM	POSITIVE	Inducer	Inducer
VAKAYFEVLRAQDNL	SVM	POSITIVE	Inducer	Inducer
KAYFEVLRAQDNLEF	SVM	POSITIVE	Inducer	Inducer
RDAAFEAVTSSRSAL	SVM	POSITIVE	Inducer	Non-inducer
AYFEVLRAQDNLEFV	SVM	POSITIVE	Inducer	Inducer
TTSLTKQAEFAYVAA	SVM	POSITIVE	Inducer	Non-inducer
NTTSLTKQAEFAYVA	SVM	POSITIVE	Inducer	Non-inducer
QDVMDVNAGLKVAKK	SVM	POSITIVE	Inducer	Non-inducer
EQDVMDVNAGLKVAK	SVM	POSITIVE	Inducer	Non-inducer
DVRAYNNNINASIGA	SVM	POSITIVE	Inducer	Non-inducer
SQDLEAAYRSVVKDV	SVM	POSITIVE	Inducer	Non-inducer
ASQDLEAAYRSVVKD	SVM	POSITIVE	Inducer	Non-inducer
AASQDLEAAYRSVVK	SVM	POSITIVE	Inducer	Non-inducer

			Result	
Sequence	**Method**	**IFN**	**IL4**	**IL10**
QQGLILRVAKAYFEV	SVM	POSITIVE	Inducer	Non-inducer
QDLEAAYRSVVKDVR	SVM	POSITIVE	Inducer	Non-inducer
TQQGLILRVAKAYFE	SVM	POSITIVE	Inducer	Inducer
EAAYRSVVKDVRAYN	SVM	POSITIVE	Inducer	Non-inducer
IGALRAYEQAVISAK	SVM	POSITIVE	Inducer	Non-inducer
SIGALRAYEQAVISA	SVM	POSITIVE	Inducer	Non-inducer
GALRAYEQAVISAKS	SVM	POSITIVE	Inducer	Non-inducer
PSLTLDGGYNYGNNS	SVM	POSITIVE	Inducer	Non-inducer
VRAYNNNINASIGAL	SVM	POSITIVE	Inducer	Non-inducer
RAYNNNINASIGALR	SVM	POSITIVE	Inducer	Non-inducer
GYNINRSDQAPRESD	SVM	POSITIVE	Inducer	Non-inducer
QRSSWVSLDTAEKKA	SVM	POSITIVE	Inducer	Non-inducer
YQRSSWVSLDTAEKK	SVM	POSITIVE	Inducer	Inducer
NNNINASIGALRAYE	SVM	POSITIVE	Inducer	Non-inducer
YNNNINASIGALRAY	SVM	POSITIVE	Inducer	Non-inducer
TRTIVDVLDATRRLY	SVM	POSITIVE	Inducer	Non-inducer
AYNNNINASIGALRA	SVM	POSITIVE	Inducer	Non-inducer

There, 22 different helper T-lymphocytes were retrieved as they were IFN positive, IL4 inducer & IL inducer. These huge amounts of epitopes might add weight only to the final vaccine. For this regard, these 22 epitopes were input to the AlgPred to check the allergenicity of each of the epitopes and also input to Vaxijen v2.0 to check their antigenicity. From that server, the interpretation would be collecting the non-allergen epitopes from AlgPred and antigenic epitopes from Vaxijen v2.0.

Table 7: Allergenicity and antigenicity prediction of HTL epitopes

Sequence	**Allergenicity**	**Antigenicity**
LEFVRAEKAAVGRQL	Allergen	Non-Antigen
NLEFVRAEKAAVGRQ	Allergen	Non-Antigen
QDNLEFVRAEKAAVG	Non-Allergen	Non-Antigen
DNLEFVRAEKAAVGR	Non-Allergen	Non-Antigen
AKAYFEVLRAQDNLE	Allergen	Non-Antigen
KAYFEVLRAQDNLEF	Allergen	Non-Antigen
RAQDNLEFVRAEKAA	Non-Allergen	Non-Antigen
AQDNLEFVRAEKAAV	Non-Allergen	Non-Antigen
QQNLSLLAARISQDV	Allergen	Non-Antigen
AYFEVLRAQDNLEFV	Allergen	Non-Antigen
QQQNLSLLAARISQD	Allergen	Non-Antigen
RVAKAYFEVLRAQDN	Allergen	Non-Antigen
VAKAYFEVLRAQDNL	Allergen	Non-Antigen
SSWVSLDTAEKKARQ	Non-Allergen	Antigen
DSQYAATQQGLILRV	Allergen	Non-Antigen
TQQGLILRVAKAYFE	Allergen	Non-Antigen
QAVISAKSALEATEA	Allergen	Non-Antigen
KQAEFAYVAASQDLE	Allergen	Non-Antigen
EQAVISAKSALEATE	Allergen	Non-Antigen
YQRSSWVSLDTAEKK	Non-Allergen	Antigen
DAAFEAVTSSRSALL	Allergen	Non-Antigen
QRDAAFEAVTSSRSA	Allergen	Non-Antigen

3.4: B-cell epitope prediction:

The BepiPred linear epitope identification 2.0 was utilized to find the probable linear B-cell epitopes, 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 table 8 given below:

The anticipated B cell epitopes were plotted with the epitopes' residue scores on a graph acquired from the server. According to the article, B-Cell epitopes were found and collected, which were more than 20mer in length and which had no allergenicity shown according to the AlgPred 2.0 server.

Table 8: Allergenicity test of B-Cell epitopes

Sequence	**Allergenicity**
SQELYQRSSWVSLDTAEKKARQADSQYAAT	Non-Allergen
KAAVGRQLEQTKQRFEVGLSAITDVHDAQA	Allergen
GQEYSKLAVLDTKRFAASRTTESSEA	Non-Allergen
GGYNYGNNSNDNAKNTSGEEYND	Allergen
VPLYTGGNTTSLTKQAEFAYVAASQD	Potential Allergen

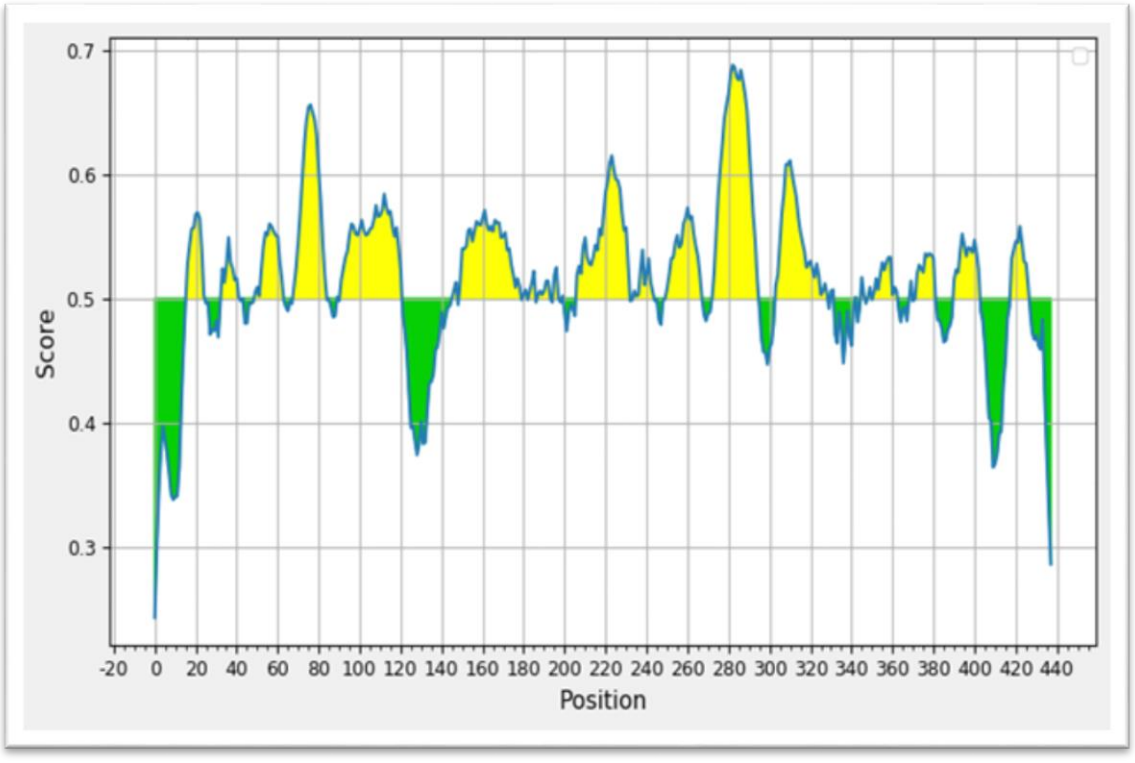


Figure 8: Score vs position graph of B-cell epitope (Jespersen et al., 2017).

3.5: Construction of Final vaccine:

As it was a multiepitope-based vaccine preparation approach, the candidacy of the vaccine depended on choosing the best credible epitopes, which were 3 CTL epitopes, 2 HTL epitopes, and 2 Linear B cell epitopes, attached using linkers between the adjuvant-sequence linking regions. At first, the outer membrane protein sequence of Vibrio Cholerae was considered the adjuvant portion of the vaccine. This adjuvant was linked with CTL epitopes using the "EAAAK" linker. Then CTL epitopes were linked with "AAY" linkers and used "GPGPG" with HTL epitopes. Lastly, HTL epitopes were linked to B-cell epitopes using "KK" linkers and B-Cell epitopes were also linked using that linker (figure 9).

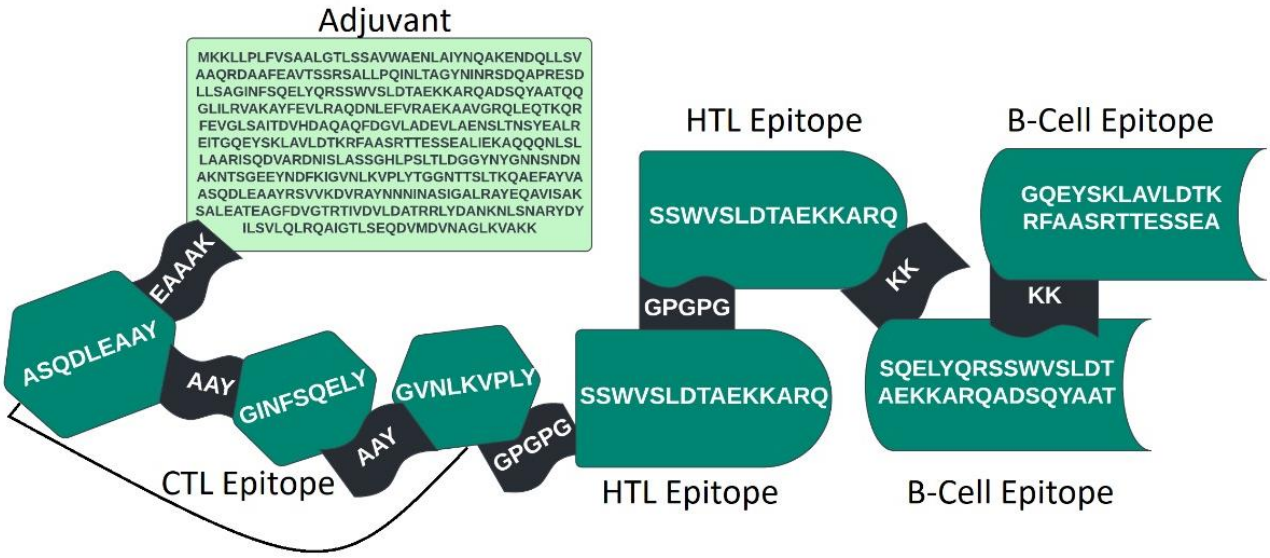


Figure 9: Constructed vaccine using linkers

The final constructed vaccine is as follows:

MKKLLPLFVSAALGTLSSAVWAENLAEIYNQAKENDPQLLSVAAQRDAAFEAVTSSRS
ALLPQINLTAGYNINRSDQAPRESDLLSAGINFSQELYQRSSWVSLDTAEKKARQADSQ
YAATQQGLILRVAKAYFEVLRAQDNLEFVRAEKAAVGRQLEQTKQRFEVGLSAITDVH
DAQAQFDGVLADEVLAENSLTNSYEALREITGQEYSKLAVLDTKRFAASRTTESSEALIE
KAQQQNLSLLAARISQDVARDNISLASSGHLPSLTLDGGYNYGNNSNDNAKNTSGEEY
NDFKIGVNLKVPLYTGGNTTSLTKQAEFAYVAASQDLEAAYRSVVKDVRAYNNNINAS
IGALRAYEQAVISAKSALEATEAGFDVGTRTIVDVLDATRRLYDANKNLSNARYDYILS

VLQLRQAIGTLSEQDVMDVNAGLKVAKKEAAAKASQDLEAAYAAYGINFSQELYAAY
GVNLKVPLYGPGPGSSWVSLDTAEKKARQGPGPGYQRSSWVSLDTAEKKKKSQELYQ
RSSWVSLDTAEKKARQADSQYAATKKGQEYSKLAVLDTKRFAASRTTESSEA

3.6: Biochemical Analysis of the Constructed Vaccine:

The PROTPARAM tool on the Expasy server was used to perform biochemical analyses to evaluate the proposed vaccine. The results are based on a molecular formula, molar mass, instability index, aliphatic index, theoretical PI, GRAVY, and other characteristics (Fig 13-15). The vaccine's instability index value was 37.58, which, according to the server, indicates that it is stable since it has an importance of less than 40. The vaccine is hydrophilic, as shown by the -0.435 Grand Average of Hydropathicity (GRAVY) rating. Hydrophilic vaccinations are preferred because hydrophobic vaccines are more prone to contamination and loss of functioning.

Atomic composition:

Carbon = 2749

Hydrogen = 4376

Nitrogen = 778

Oxygen = 898

Sulfur = 2

Formula: C2749H4376N778O898S2

Total number of atoms: 8803

Extinction coefficients:

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

Ext. coefficient = 67730

Abs 0.1% (=1 g/l) 1.079

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).

The **instability index (II)** was computed to be 37.58, which classified the protein as stable.

Grand average of hydropathicity (GRAVY) was measured as -0.435, which is an excellent outcome to indicate the proposed vaccine as a well-prepared vaccine.

Number of Amino Acid: 576 Molecular Weight: 62.757 kDa Theoretical PI: 5.57

Figure 10: "The amino acid content, molecular weight, and theoretical PI of the constructed vaccine", (Gasteiger et al., 2005).

3.7: Constructed Vaccine's Allergenicity and Toxicity Evaluation:

Using a proposed methodology, the Allergen web server validated the vaccine's allergic sensitivity, allergenicity. A value of 0.5 was set based on z-score analysis, to efficiently quantify the protein's effectiveness, and the Full FASTA 36 approach was employed in this case. Additionally, the T3DB server was used to observe whether the proposed vaccine candidate was showing toxicity or not, and the result revealed that there was nothing to predict as toxic, which empowers the vaccine towards its credibility.

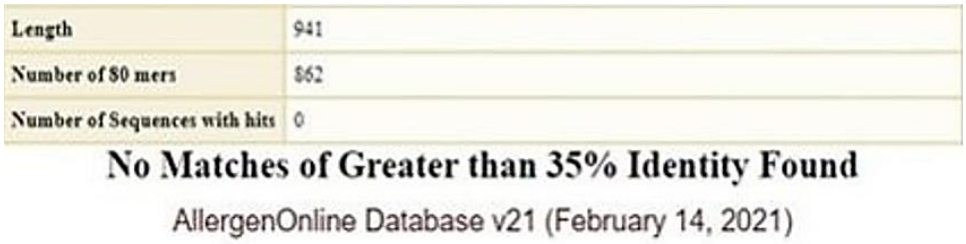


Figure 11: "Allergenicity of the developed vaccine as examined by the Allergen Online server", (Gendel & Jenkins, 2006).

Figure 12: "The designed vaccine's toxicity prediction came no result", (Wishart et al., 2015).

3.8: Homology modeling of vaccine:

It is critical as well as crucial to obtain a 3D structure of prepared In-silico vaccine in order to continue the relevance study. It could be constructed as a 3D design while it is converted in the form of a PDB file using the in-silico technique, Phyre2 server. This PDB file was created using the homology modeling approach, and the highest-ranking template, c1tqqC, was used to model 418 resides (73% of the sequence) with a 100% confidence (figure 13).

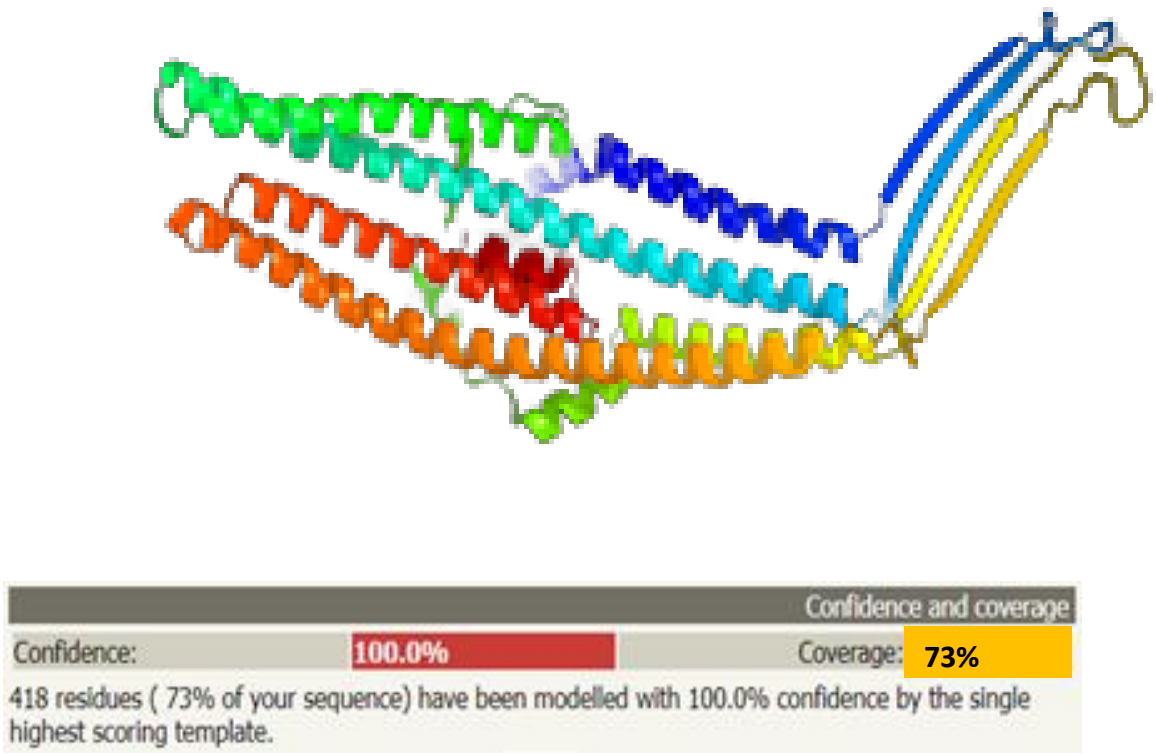


Figure 13: "A 3D dynamic model of the immunization was constructed using the Phyre2 server.", ((Mezulis et al., 2015)

3.9: Analysis of Homology modeling of vaccine:

The PDB structure of the proposed vaccine, which was retrieved from phyre2 was studied further using the swiss model service of expasy server. The Ramachandran plot analysis was performed using the SWISS PDB plotter, where the "Favored region" was found 94.00% and "Outliers" were 2.16%, which is predicted as a successful outcome. Alongside, a Z-score against residue analysis curve was generated using the PROSA web server, to help observe the potential role-playing structure was occupied as the value came negative (-7.98) and it was preferred to be stable than expected, as the black dot was in the gray zone (figure 14).

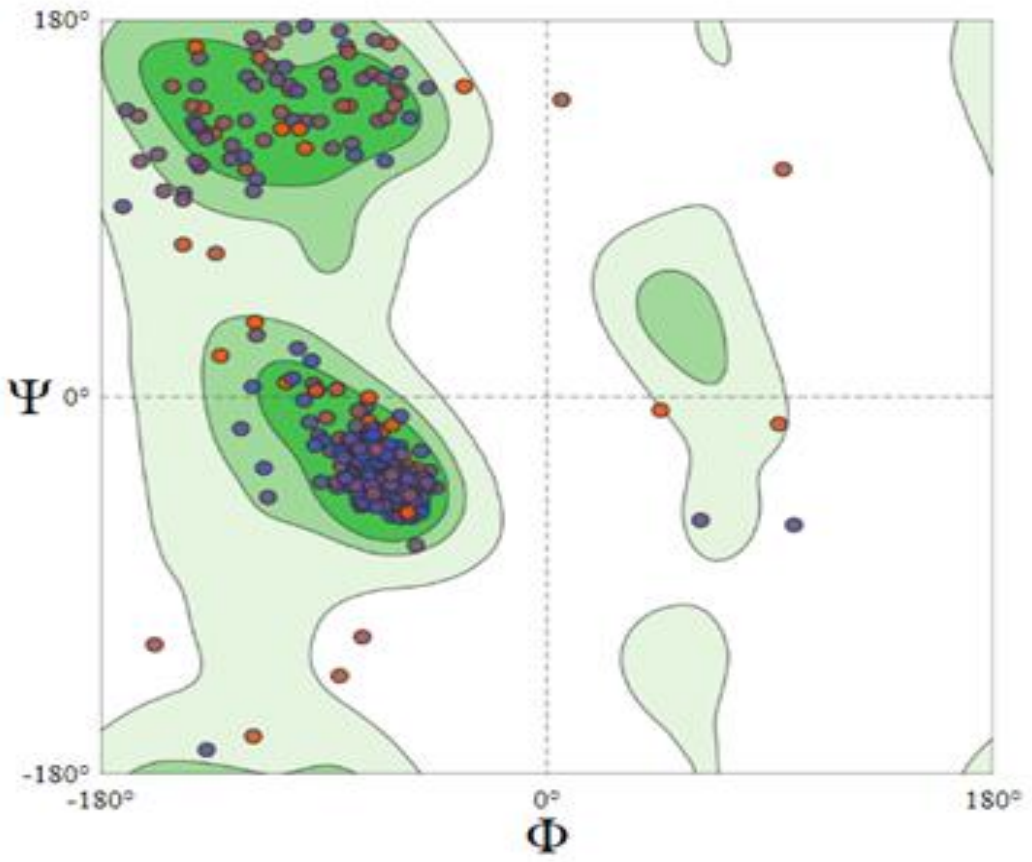


Figure 14: "SWISS PDB plotter for localizing Ramachandran plot", (Waterhouse et al., 2018)et al., 2018)].

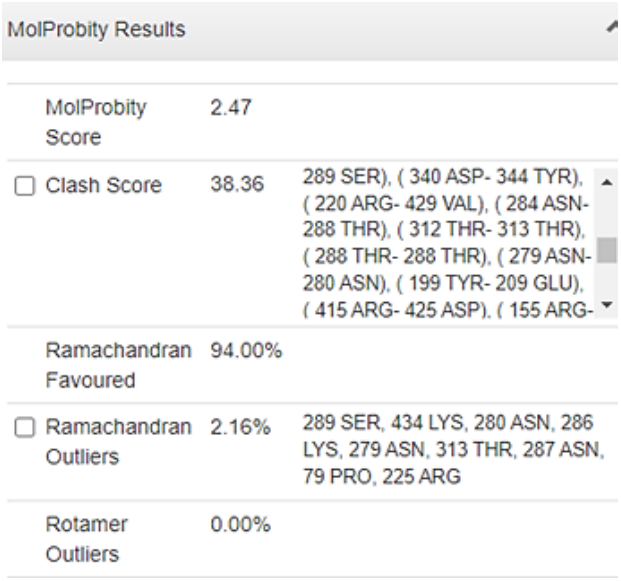


Figure 15: "MolProbity results (Waterhouse et al., 2018).

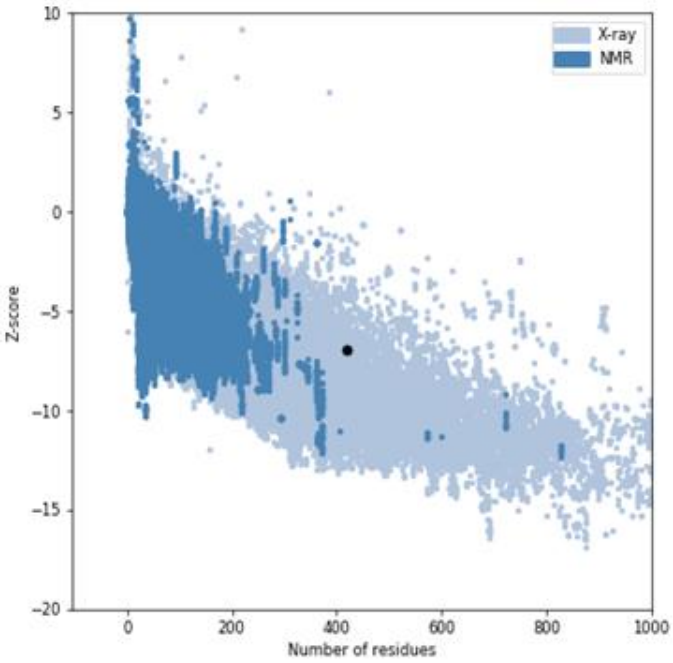


Figure 16: "Overall model quality: Z-score analysis", (Waterhouse et al., 2018).

The result found that for QMEANDisCo's Global score was 0.74.

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

Molecular docking was used to assess the binding affinity between the proposed vaccine design and the relevant human "Toll-Like Receptor (TLR4)", which is part of the toll-like receptor family (TLR). The TLR4 receptor was collected from protein data bank of Europe, which was retrieved as the most X-ray diffracted top resolution resultant of 3.6Å, with genome ID "3ULA" "Crystal structure of the TV3 mutant F63W-MD-2-Eritoran complex" (Han et al., 2012). This receptor family, which includes protein-rich receptors, activates the innate immune system, because they are single-pass membrane-spanning receptors, they are usually found in cells that are primarily responsible for removing pathogens that have infiltrated the body. TLRs are distinguished by numbers 1 to 13 on their labeling. Amongst, TLR4 was used as receptor protein here. The ligand (retrieved from the phyre2 service as a PDB file) and TLR4 were employed in our study. The best combination between TLR8 and our suggested vaccine got the maximum score of 19366 with a transformation of (-0.59 -1.48 1.09 85.67 -14.26 205.06) 371.46 KJmol-1, covering an area of

3200.20 square angstroms. The PDB structure of the created protein-ligand combination may be seen using Discovery Studio 2016's 64-bit client version.

For checking the binding ability with human TLR4 receptor, the Cluspro V2.0 docking server was utilized, (Nezafat et al., 2016), which rotated receptor TLR4 with antigenic ligand proposed vaccine for 1000 times and could observe all of them to screen out the best clusters. The chosen clusters were ranked by the Cluspro V2.0 server interpreting on the criteria of conveying high cluster number as well as lower energy (Comeau et al., 2004). There, 23 clusters had been chosen to show as result from where, the "0" ranked cluster was chosen as this cluster of receptor and ligand molecule shows highest cluster member (79) along with lowest energy (-1123.2) for which this docked cluster was retrieved for further use (Table 9).

Table 9: Protein-Protein Docking best cluster Cluspro V 2.0

Cluster	Members	Representative	Weighted Score
0	79	Center	-1019.7
		Lowest Energy	-1123.2
1	76	Center	-981.0
		Lowest Energy	-1138.9
2	49	Center	-922.3
		Lowest Energy	-1032.8
3	41	Center	-925.7
		Lowest Energy	-1040.1
4	36	Center	-981.9
		Lowest Energy	-1153.8
5	35	Center	-913.8
		Lowest Energy	-1034.6

According to the preferable PDB file viewer, Discovery Studio Visualizer was intended to provide the view of chosen best resulting docked file.

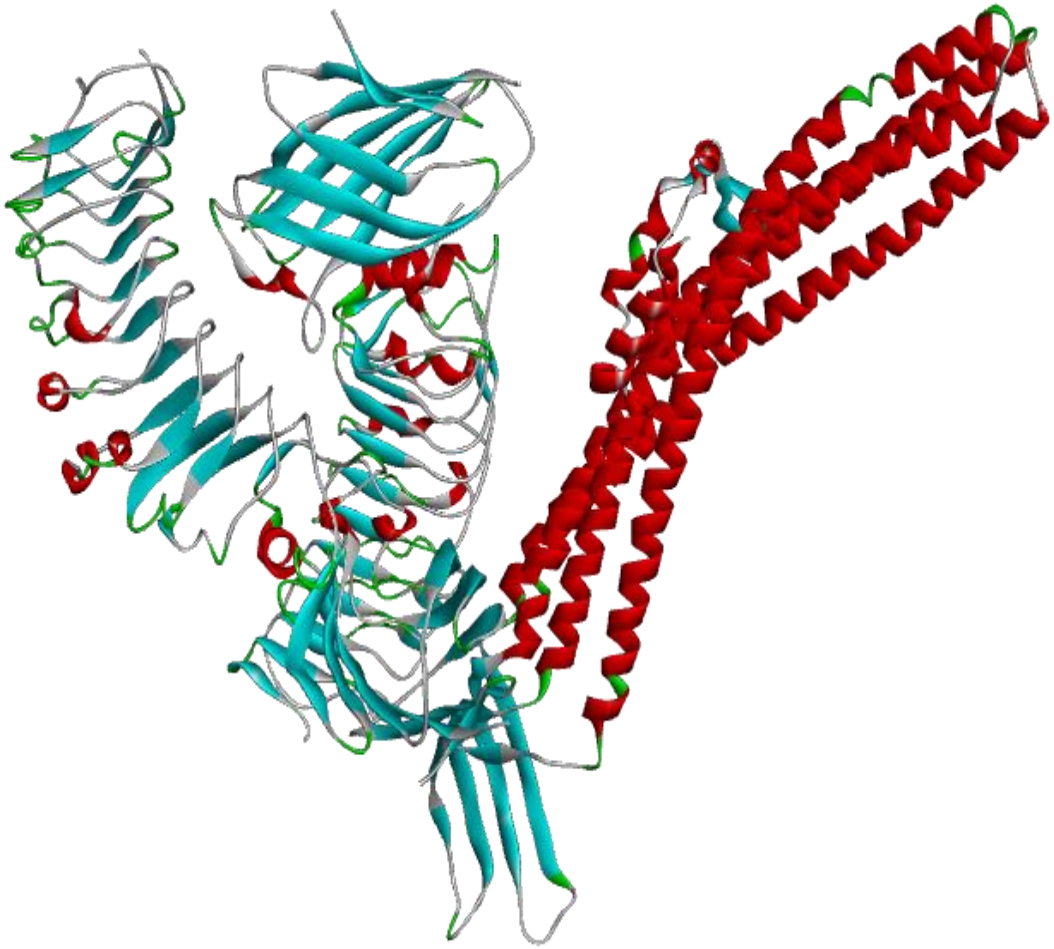
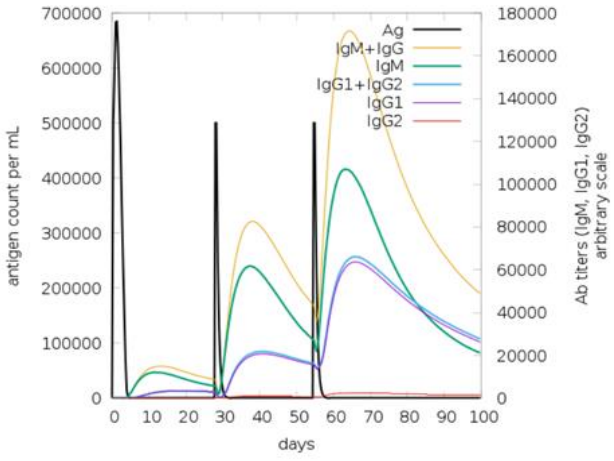


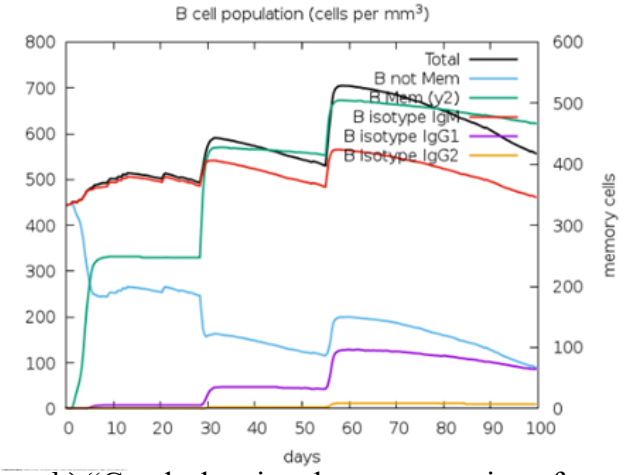
Figure 17: “The docked complex between the TLR4 receptor and the proposed vaccination in 3D using Discover Studio”, (Han et al., 2012).

3.11: Immune Simulation in silico for the immune response:

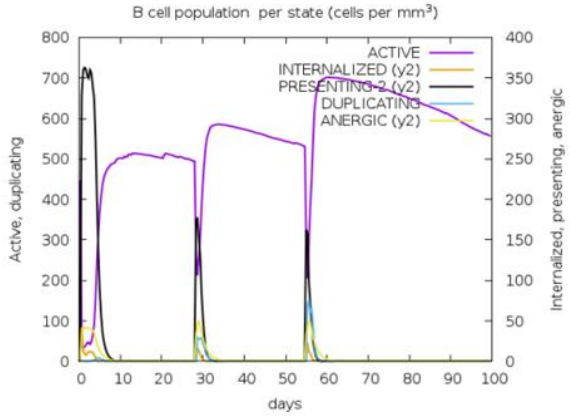
Susceptible immune response was recorded for the final vaccination, using the C-ImmSim online service, which resulted in immunological profiles for the desired vaccine as shown in figures 18 (a-m). IgG1 + IgG2 and IgM were employed to determine secondary and tertiary immune response proliferation, as well as a reduction in antigen count (IgG + IgM), suggesting immune response proliferation (Banagozar et al., 2019).



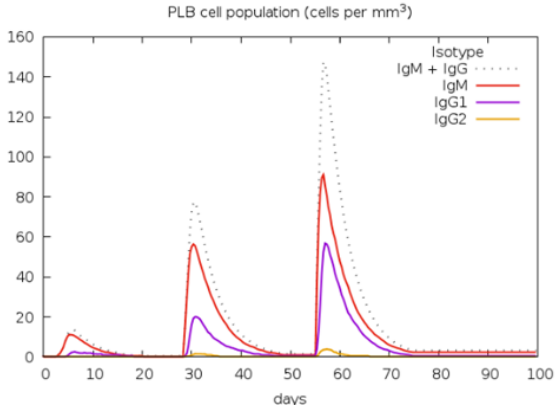
a) "The pathogen, the immunoglobulins, and the immune complexes." (Banagozar et al., 2019) following vaccine doses in C-IMMSIM



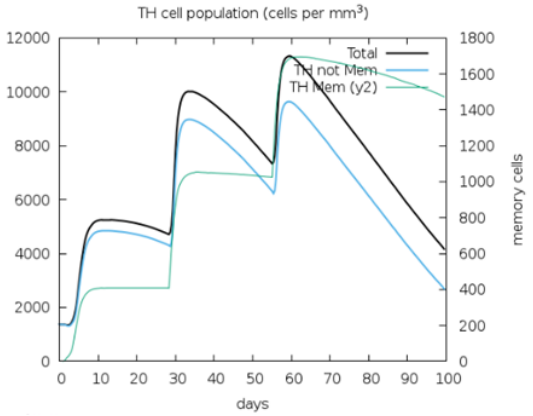
b) "Graph showing the concentration of B cells based on subtypes and administration." (Banagozar et al., 2019)



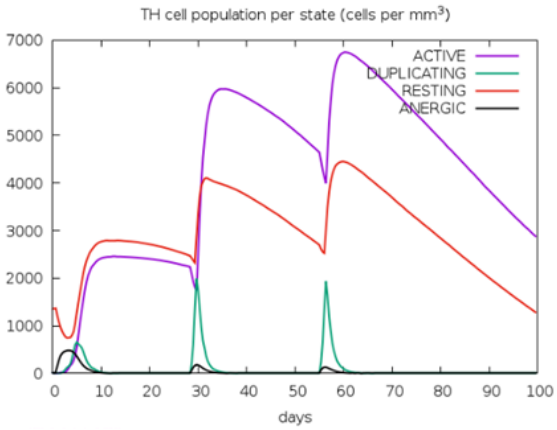
c) "Graph showing entity-state of B cells versus days after vaccine administration." (Banagozar et al., 2019).



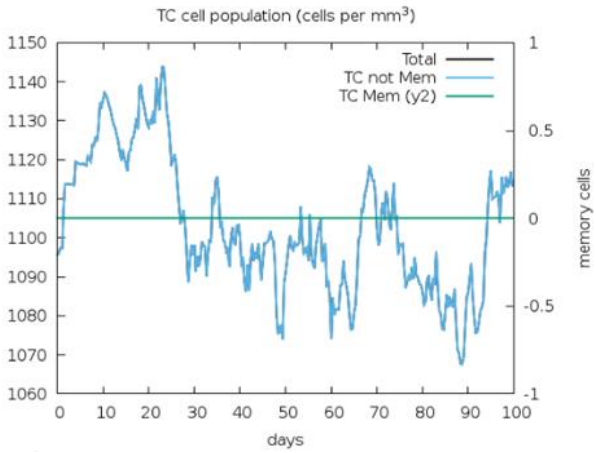
d) "B cell population expansion in plasma vs. vaccination treatment day" (Banagozar et al., 2019)



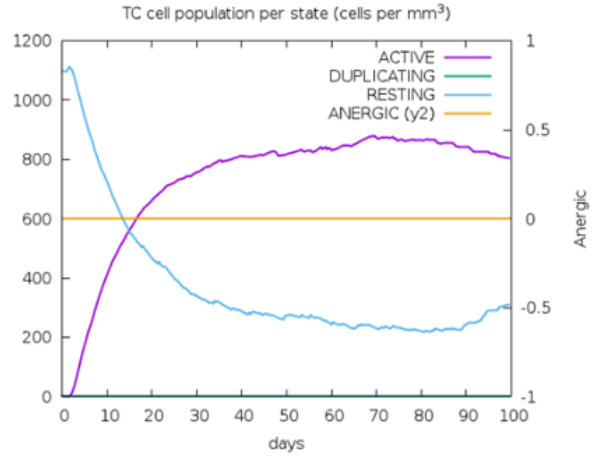
e) "CD-4 HTL epitopes count. The plot shows the total and memory count." (Banagozar et al., 2019)



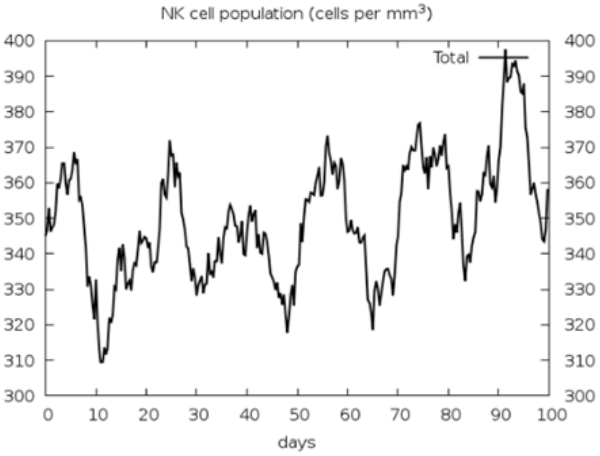
f) "CD-4 HTL epitopes count subdivided into per entity-state" (Banagozar et al., 2019)



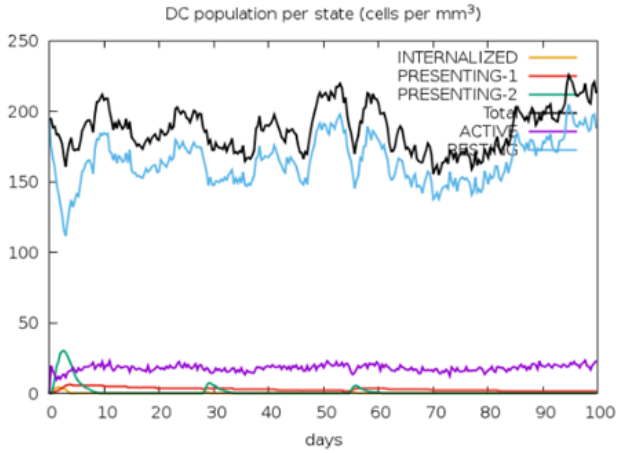
g) "CTL total count (Total and Memory)" 2022, Feb 12 (Banagozar et al., 2019).



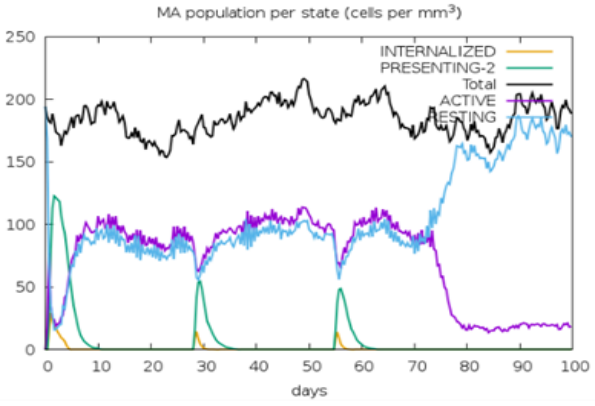
h) "CTL count per entity state" 2022, Feb 12. (Banagozar et al., 2019)



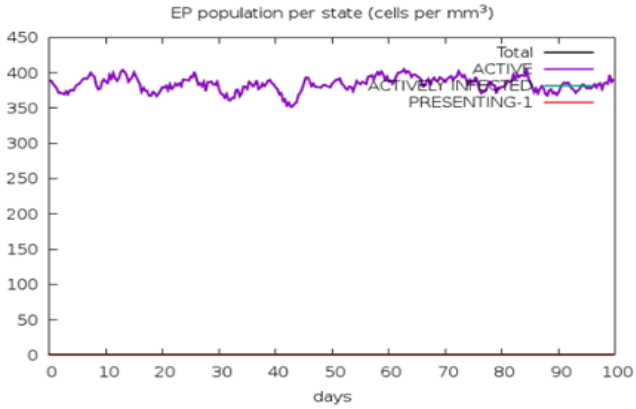
i) "Total NK cell population count after vaccine administration" (Banagozar et al., 2019)



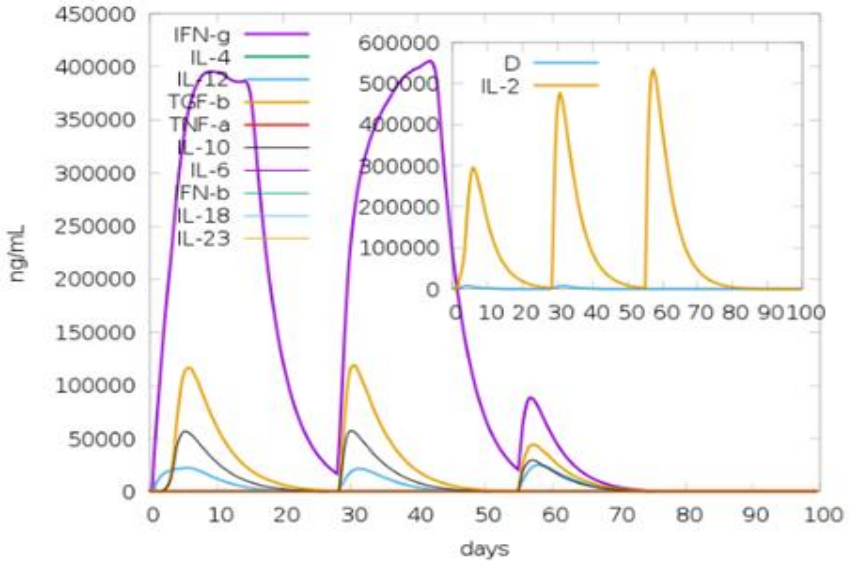
j) "Antigenic peptides can be found in both MHC I and MHC II molecules in the DC population." (Banagozar et al., 2019).



k) "Macrophages population growth per entity state" (Banagozar et al., 2019)



l) "Epithelial cells total population count per entity-state" (Banagozar et al., 2019)



m) “Concentration of cytokines and interleukins. In the inset figure, the danger signal is shown with the leukocyte growth factor IL-2.”(Banagozar et al., 2019).

Figure 18: C-Immsim immune simulations (a-m)

Chapter 4: Discussion

Disregarding over 100 years of asking about, cholera continues to bring issues and shocks. All through a significant part of the 20th hundred years, this disease being brought about by Vibrio Cholerae of the O1 serogroup, and the disorder was generally obliged to Asiatic and African countries. Anyway, the recent 10 years of the 20th century saw critical degrees of progress, twice, inside the historical backdrop of it. In 1991, a far-reaching discharge of cholera broke out in South America, to speak region unaffected by the disease for a hundred years. In 1992, a plainly unused far and wide brought about by a generally dark V. cholerae (O139) occurred in India and Bangladesh. The O139 broad has stretched out rapidly in networks remembered to be essentially protected to V. cholerae O1. The O139 episode has expanded rapidly to different nations, counting the United States, in people recalled to be for the most part impenetrable to V. cholerae O1. All these components of cholera in this write-up analyzing and counting clinical microbial science, the consideration of illness transmission, pathophysiology, and clinical properties of the ailment. Interesting complement will be put on the gigantic divulgences finished as of late in understanding the nuclear pathophysiology of this disease and developing unused ages of immunizations to hinder it.

The UniProt Knowledgebase (UniProtKB) server was used to collect the proper "Outer Membrane Protein Sequence of Vibrio Cholerae O1 Serotype", following the footpath of previously proved and established antigenic and effectual "Outer Membrane Protein" of Vibrio Cholerae "In Silico" vaccine approach. The antigenicity result of the selected "Outer Membrane Protein Sequence of Vibrio Cholerae O1 Serotype" is 0.5469, stated as "Probable ANTIGEN" by the VaxiJen V2.0. Then, three Cytotoxic T Lymphocyte (CTL) epitopes, two Helper T lymphocyte (HTL) epitopes, and three antibodies binding epitope of B-Cell epitope are generated using NetCTL - 1.2 & NetMHCpan - 4.1 (Pan-specific binding of peptides to MHC class I proteins of known sequence) for CTL epitopes, NetMHCIIpan - 4.0 (Pan-specific binding of peptides to MHC class II alleles of known sequence) for HTL epitopes, BepiPred (Predicting the location of linear B-cell epitopes) for B-cell epitopes. This proposed vaccine was then developed using linkers to be added and tested for antigenicity using Vaxijen V2.0 and the result was 0.6203. Then, the biochemical analysis in PROTPARAM showed the stability profile resulting from the vaccine's instability index (37.58) and GRAVY (-0.435). Examining the homology modeling using the Phyre2 server, generated a

PDB model of the final vaccine, which had 100 percent confidence of covering 73 percent of the vaccine model. The z-score (-6.98) was used to help consider the overall quality of the model using the ProSA-web server. After that, the molecular docking was done using Patchdock and achieved a score of 20884 in a 3101.90 square angstrom region. At last, using the C-IMMSIM server for In Silico immune response, we could observe an effective immunity generated through the server.

The C-immsim website determines the B cell lymphocyte concentration after vaccination as well. B cell epitopes are important in both humoral and cellular immunity, and IgM, IgG1, and IgG2 concentrations dictate B cell concentrations. Graphs displaying the density of B cells in each condition. Last but not least, plasma B cells have been identified. These cells have the potential to be employed as pharmaceuticals. Graphs demonstrating CTL and HTL epitope concentrations were received from the server in the same way as plots exhibiting B cell concentrations were generated. Let us look at how 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. The C-Immsim server also shows the host's WBC, DC, epithelial, NK cells, and macrophages count after vaccination. 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 were created to depict epithelial cell and cytokine development after immunization.

Chapter 5: Conclusion

The Blue Death, Cholera epidemic, history denotes it as, a deoxygenated, dehydrated death of humans due to the bacterial species named Vibrio Cholerae. When the sanitation system is extremely devastated by the purity of water and lacks the knowledge of hygiene, inhabitants make themselves vulnerable to this water and airborne microbe, Vibrio Cholerae. After getting affected by Vibrio Cholerae, a person gets desiccated due to water loss and infecting intestinal lumen (small and large intestine), if it stays untreated, the death of the person might occur in a few hours. Studying this severely spreading bacteria species and the extremity of the causing disease, an "In Silico" approach is made to prepare a vaccine against Vibrio Cholerae. This Research Work is done in two major steps. Vaccine Construction and Its Biochemical analysis. The UniProt Knowledgebase (UniProtKB) to search and explore the proteome of Vibrio Cholerae, specifically the "Outer Membrane Protein Sequence" of the pathogen, screened based on the proteomic antigenicity. The outer membrane proteins are the one severely infectable part of gram-negative bacteria like Vibrio Cholerae to its host animal, functionally, its outer membrane is more invadable towards the eukaryotic cell, making it a pickable concern to design a vaccine against its pathogen, Vibrio Cholerae. After several antigenicity tests, a suitable protein sequence was discovered, showing excellent antigenicity. With the protein antigen selection the corresponding Cytotoxic T-Lymphocyte, the multi-epitope vaccine preparation starts with achieving Cytotoxic T-Lymphocyte (CTL) epitopes of A1 supertype of major histocompatibility complex (MHC) I, using the protein fasta format of "Outer Membrane protein of Vibrio Cholerae O1 Serotype", with a threshold of 0.75, from NetCTL-1.2 Server. According to the server's combined score, there had been 14 collectively gained A1 Supertypes which were used to run and get the MHC Class I corresponding Alleles. The orientation of the epitopes were assessed predicting Z-score which is negative 0.7.25, indicating that the created 3D protein model is quite favorable and the black dot's position in bluish grey zone interprets the prepared model as predictable in terms of the database. Then the ligand-receptor binding ability was assessed by molecular docking between the TLR4 of the host cell and the ligand vaccine and the highest cluster was retained as of its lowest energy. After docking a simulation study was conducted in 3 days at a 28 days interval. Susceptible immune response was recorded for the final vaccination, using the C-ImmSim online service, which resulted in immunological profiles for the desired vaccine. IgG1 + IgG2 and IgM were employed to determine secondary and tertiary immune response proliferation, as well as a reduction in antigen count (IgG

+ IgM), suggesting immune response proliferation. The C-immsim website determines the B cell lymphocyte concentration after vaccination as well. B cell epitopes are important in both humoral and cellular immunity, and IgM, IgG1, and IgG2 concentrations dictate B cell concentrations. Graphs displaying the density of B cells in each condition. Last but not least, plasma B cells have been identified. The final constructed vaccine was stable, but the result of homology modeling through phyre2 needs to be more precise and accurately measured, alongside the Ramachandran outlier (2.16%) must be reduced to below 1%. Alongside, The in-depth allergenicity profile is missing due to newly collected strains. Therefore, vaccine allergenicity should be tested after enriching the database of allergen online.

References

Almagro-Moreno, S., & Taylor, R. K. (2013). Cholera: Environmental Reservoirs and Impact on Disease Transmission. *Microbiology Spectrum*, *1*. https://doi.org/10.1128/MICROBIOLSPEC.OH-0003-2012

Álvaro-Benito, M., Morrison, E., Abualrous, E. T., Kuropka, B., & Freund, C. (2018). Quantification of HLA-DM-dependent major histocompatibility complex of class II immunopeptidomes by the peptide landscape antigenic epitope alignment utility. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/FIMMU.2018.00872/BIBTEX

B Cell Tools. (2022). Retrieved June 13, 2022, from http://tools.iedb.org/main/bcell/

Banagozar, A., Vadivel, K., Stuijk, S., Corporaal, H., Wong, S., Lebdeh, M. A., Yu, J., & Hamdioui, S. (2019). CIM-SIM: Computation in Memory SIMuIator. *Proceedings of the 22nd International Workshop on Software and Compilers for Embedded Systems, SCOPES 2019*. https://doi.org/10.1145/3323439.3323989

Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., de Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLOS Computational Biology*, *9*. https://doi.org/10.1371/JOURNAL.PCBI.1003266

Conner, J. G., Teschler, J. K., Jones, C. J., & Yildiz, F. H. (2016). Staying alive: Vibrio cholerae's cycle of environmental survival, transmission, and dissemination. *Microbiology Spectrum*, *4*. https://doi.org/10.1128/MICROBIOLSPEC.VMBF-0015-2015

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

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

Gendel, S. M., & Jenkins, J. A. (2006). Allergen sequence databases. *Molecular Nutrition & Food Research*, *50*. https://doi.org/10.1002/MNFR.200500271

Gram Negative Bacteria Concept & Examples | What is Gram Negative Bacteria? | Study.com. (2022). Retrieved June 10, 2022, from https://study.com/learn/lesson/gram-negative-bacteria-concept-examples.html

Han, J., Kim, H. J., Lee, S. C., Hong, S., Park, K., Jeon, Y. H., Kim, D., Cheong, H. K., & Kim, H. S. (2012). Structure-based rational design of a Toll-like receptor 4 (TLR4) decoy receptor with high binding affinity for a target protein. *PLoS ONE*, *7*. https://doi.org/10.1371/journal.pone.0030929

Harding, C. v., & Unanue, E. R. (1990). Quantitation of antigen-presenting cell MHC class II/peptide complexes necessary for T-cell stimulation. *Nature*, *346* https://doi.org/10.1038/346574A0

Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, *154*. https://doi.org/10.1111/IMM.12889

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*. https://doi.org/10.1093/NAR/GKX346

Khan, M., Khan, S., Ali, A., Akbar, H., Sayaf, A. M., Khan, A., & Wei, D. Q. (2019). Immunoinformatics approaches to explore Helicobacter Pylori proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. *Scientific Reports 2019, 9*. https://doi.org/10.1038/s41598-019-49354-z

Khan, R., Petersen, F. C., & Shekhar, S. (2019). Commensal bacteria: An emerging player in defense against respiratory pathogens. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.01203

Mezulis, S., Yates, C. M., Wass, M. N., E Sternberg, M. J., & Kelley, L. A. (2015). *The Phyre2 web portal for protein modeling, prediction and analysis*. https://doi.org/10.1038/nprot.2015.053

Mulder, N. J., Kersey, P., Pruess, M., & Apweiler, R. (2007). In Silico Characterization of Proteins: UniProt, InterPro and Integr8. *Molecular Biotechnology 2007 38*. https://doi.org/10.1007/S12033-007-9003-X

Nezafat, N., Karimi, Z., Eslami, M., Mohkam, M., Zandian, S., & Ghasemi, Y. (2016). Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational Biology and Chemistry*, *62*. https://doi.org/10.1016/j.compbiolchem.2016.04.006

Non-cholera Vibrio cholerae Infections | Cholera | CDC. (2022). Retrieved June 13, 2022, from https://www.cdc.gov/cholera/non-01-0139-infections.html

Raheem, S. G., Salh, K. K., Ibrahim, K. S., & Gorji, A. E. (2021). In-Silico Designing a Multi-Peptide Vaccine: Against Vibrio Cholera. *International Journal of Peptide Research and Therapeutics 2021, 27*. https://doi.org/10.1007/S10989-021-10190-3

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*. https://doi.org/10.1093/NAR/GKAA379

Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. (2010). NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*, *62*. https://doi.org/10.1007/S00251-010-0441-4)

van der Pol, L., Stork, M., & van der Ley, P. (2015). Outer membrane vesicles as platform vaccine technology. *Biotechnology Journal*, *10*. https://doi.org/10.1002/BIOT.201400395

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*. https://doi.org/10.1093/NAR/GKY427

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2015). T3DB: the toxic exposome database. *Nucleic Acids Research*, *43*. https://doi.org/10.1093/NAR/GKU1004

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 2021, 11*. https://doi.org/10.1038/s41598-021-81749-9