In-silico Approach of Fusion Glycoprotein (F) Targeted Multi-Epitope Vaccine Against Human Respiratory Syncytial Virus (HRSV)

By

Al Saba Sejan
ID: 18346029

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
February 2023

©2023 Brac University
All rights reserved.

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:

_____

AL SABA SEJAN

Student ID: 18346029

Approval

The thesis titled “In-silico Approach of Fusion Glycoprotein (F) Targeted Multi-Epitope Vaccine Against Human Respiratory Syncytial Virus (HRSV)” submitted by Al Saba Sejan (ID: 18346029), of Spring 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
Brac University

Approved By:

Assistant Dean and
Program Director:

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
School of Pharmacy
Brac University

Dean:

Professor Dr. Eva Rahman Kabir
Professor and Dean
School of Pharmacy
Brac University

Ethics Statement

All ethical standards were maintained in the completion of this thesis. Conducting the research with integrity and following the ethical principles was given utmost importance.

Acknowledgments

I would like to begin by thanking Almighty Allah for keeping me in good health and blessing me with the capability, strength and assistance needed to complete this project. However, this research paper would not have been completed without the support of several individuals and I would like to express my gratitude to all of them.

I would like to express my gratitude to my supervisor Mohammad Kawsar Sharif Siam (Senior Lecturer, School of Pharmacy, Brac University) for helping me complete the thesis work. His continuous guidance, instructions, support and motivation throughout the project work, made it possible to complete the project work. I am really thankful for him valuable inputs and knowledge which made my work better.

I would like to acknowledge my family and friends who motivate me in every sector of life. Undoubtedly without their support, motivation, prayers and affection, I would not have come this far. Lastly, I would like to thank all the people who helped me whenever I needed and always kept supporting me.

Al Saba Sejan
ID: 18146029
School of Pharmacy
Brac University

Abstract

This work created a multi-epitope respiratory syncytial virus vaccine in-silico and simulated its biochemical effectiveness. In-silico investigation selected fusion glycoprotein (F) from envelope proteins of RSV. F protein causes virion-target cell membrane fusion. However, in-silico methodology used several servers, databases, and software such as Vaxijen v2.0 for antigenicity, NetCTL-1.2, NetMHCIIpan 4.0, and Bepipred servers which were utilized to find epitopes recognized by cytotoxic T lymphocytes (CTLs), helper T lymphocytes (HTLs), and B cells. IFNepitope, IL-4Pred, and IL-10Pred were used to identify epitopes expressed by HTLs. Linkers were used to connect epitopes. Positive findings were found in biochemical analysis of the final proposed vaccine. Positive outcomes were predicted for a range of indicators, including instability index (38.99 as stable), GRAVY score (-0.180), molecular weight (91876.78 dalton), toxicity, antigenicity (0.5810), and allergenicity. Furthermore, acceptable z-score (-9.89) and Ramachandran plot (94.95%) were obtained for the final proposed vaccine via ProSAweb and SWISS-MODEL respectively. The required outcome was also achieved by homology modeling, molecular docking, and immune response simulation via responsible servers. However, in-silico vaccine discovery for RSV and other diseases may be accelerated.

Keywords: Respiratory syncytial virus, multi-epitope vaccine, in-silico, Fusion glycoprotein, biochemical analysis, Immune simulation.

Table of contents

List of Figures

List of Tables

List of Acronyms

HRSV: Human Respiratory Syncytial Virus

LRTI: Lower Respiratory Tract Infection

CTL: Cytotoxic T-lymphocyte

HTL: Helper T-lymphocyte

hMPV: Human Metapneumovirus

hPIV: Human Para-influenza viruses

URI: Upper Respiratory Infections

IVIG: Intravenous Immunoglobulin

NS: Nonstructural proteins (NS-1 and NS-2)

IFN: Interferon

TLR: Toll-like receptor

MHC: Major histocompatibility complex

IL4: Interleukin 4

IL10: Interleukin 10

GRAVY: Grand average of hydropathicity index.

Chapter 1
Introduction

The Respiratory syncytial virus (RSV) is a significant virus responsible for respiratory tract infections in vulnerable populations such as infants, elderly individuals, and those with compromised immune systems. The virus is a major contributor to lower respiratory tract infection (LRTI), which can cause severe clinical issues. The RSV virus is transmitted by coming into contact with droplets from the nose and throat of infected individuals when they sneeze or cough. Dried respiratory secretions on bedclothes and other items can also spread RSV. The respiratory syncytial virus (RSV) can live for a long time on hard surfaces but only for a short time on skin (Munoz & Englund, 2019).

This infectious agent is still one of the top viral killers of infants in their first year. Moreover, RSV results in a significant global impact on public health every year. In 2015, it was estimated that 33.1 million cases of acute lower respiratory infections that occurred globally were linked to RSV. Out of the identified cases, roughly 10% resulted in hospitalization and 0.4% led to death (Pandaya et al., 2019). In the United States, RSV infection leads to a high number of hospitalizations and deaths each year, with a total of 125,000 cases among children under 5 years (Mathew et al., 2021). According to the Centers for Disease Control and Prevention (CDC), respiratory syncytial virus (RSV) causes thousands of pediatric hospitalizations and hundreds of deaths annually. RSV is the leading cause of bronchiolitis and pneumonia in children younger than 1 year. According to a study of ICDDRB and Department of International Health, Johns Hopkins University, United States during the time period of 2008 to 2012, they analyzed 3,170 samples for the presence of a variety of respiratory viruses, including RSV, human metapneumovirus (hMPV), human parainfluenza viruses (hPIV) 1, 2, and 3, and adenovirus. There were 555 positive samples for RSV (17.5%), with 2.5% also showing evidence of co-infection with another virus (Hossain et al., 2022).

RSV-A and RSV-B are the two major subtypes into which this virus has initially been divided. Genotypic characterization in that study of RSV showed that RSV-A (82%) contributed more acute respiratory infections than RSV-B (18%). Clinical features were similar with RSV-A and RSV-B infections. However, children with RSV-B were more likely to have upper respiratory infections (URI) but among RSV-A cases, hospitalization was higher (Hossain et al., 2022). Unfortunately, there is currently no vaccine or treatment available for RSV, making it a

significant contributor to infection-related deaths and lower respiratory illnesses in newborns, such as bronchiolitis, pneumonia, and even the potential for wheezing and asthma in the future (Mathew et al., 2021). In 1955, RSV was first isolated from chimpanzees with respiratory diseases. Then in 1957, an Infants suffering from severe cases of lower respiratory illness were used to isolate the virus which was the first isolation of RSV from human body (Shang et al., 2021). For a long time after its isolation in 1955, researchers knew very little about respiratory syncytial virus (RSV) due to its pleomorphic and cell-associated nature, physical instability, and relatively inefficient growth in cell culture. The first steps toward a complete characterization were taken in 1981, with the molecular cloning and sequencing of RSV RNA (Collins et al., 2013). Attempts to create an effective vaccine against respiratory syncytial virus failed miserably in the 1960s. A clinical vaccine (formalin inactivated vaccine) trial conducted in the Washington, D.C. in 1966 was unsuccessful. Two infants died due to their RSV infections after receiving the shot made with inactivated virus by scientists (Lindsey et al., 2022). In 1998 United States approved the first monoclonal antibody (mAb) palivizumab for the prevention of RSV and in 2008 ALN-RSV01 was approved as an antiviral siRNA for clinical trial (Shang et al., 2021). In addition, Antibodies are important prophylactic treatments for patients at risk of serious RSV infection.

The RSV virus particles are comprised of three surface proteins, fusion (F), attachment (G) and small hydrophobic (SH) proteins with the F protein being the most preserved among various RSV strains and the primary focus of neutralizing antibodies that provide protection. The initial prevention method for RSV was RespiGam produced by Medimmune, consisting of a mixture of human intravenous immunoglobulin (IVIG) with high levels of RSV-neutralizing antibodies. It effectively prevents severe lower respiratory tract infections caused by RSV in vulnerable infants (Shang et al., 2021; Yeo et al., 2021).

In the last 10 years, due to the significant advancements in mAb screening technology, many human antibodies against the RSV F proteins have been found and identified. Out of these, many RSV F-targeting antibodies have proven to be more efficient than palivizumab (Kwakkenbos et al., 2010). Recently the development of a RSV vaccine is identified as a priority for WHO's vaccine product and delivery research unit. In the meantime, Moderna's mRNA vaccine mRNA-1345 is tested in phase-1 clinical trial (Shang et al., 2021). However, there is no properly established treatment or vaccine for RSV virus though recent clinical research and trials are optimistic and may prove fruitful results in near future.

1.1 Structural and genomic features of Human Respiratory Syncytial Virus

RSV is a non-segmented negative-sense single-stranded enveloped RNA virus which belong to the *Orthopneumovirus* genus in the *Pneumoviridae* family and *Mononegavirales* order (Shang et al., 2021). There are two antigenic subgroups of this virus, designated A and B, and they have different sequences across their entire genomes. The other members of this genus are bovine RSV (BRSV), pneumonia virus of mice (PVM) and ovine RSV (ORSV) (Collins et al., 2013). The RSV virion can take either a spherical or filamentous shape and has a diameter of around 150 nm (Bachi, 1988). The virion is made up of a lipid bilayer derived from the host plasma membrane and a helical nucleocapsid with protruding spikes of glycoprotein (F and G protein). The M protein connects the viral envelope to the nucleocapsid, and it likely forms a protein shell around the nucleocapsid (Ghildyal et al., 2006). Along with all the protein a structure of Respiratory Syncytial Virus is shown in figure 1.

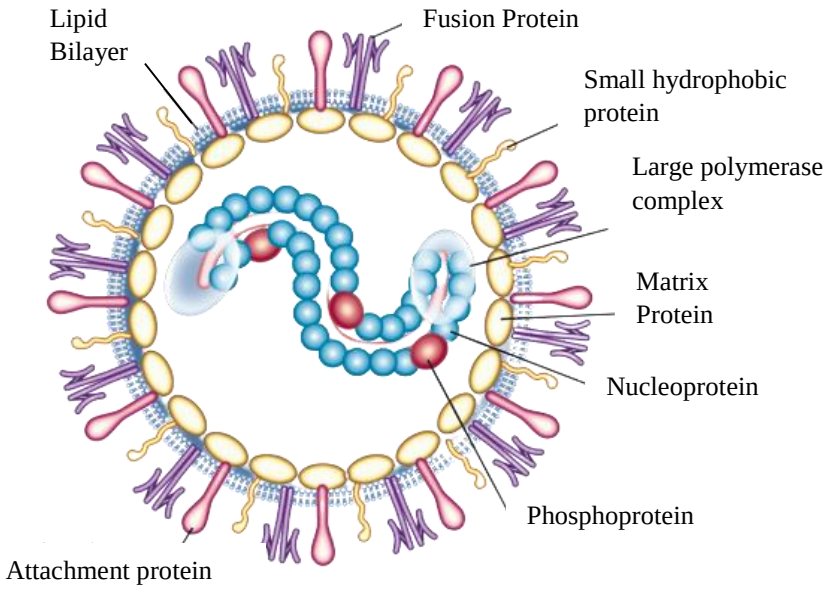


Figure 1 Structure of respiratory syncytial virus (Munoz & Englund, 2019).

Its 11 proteins are encoded by 10 genes. The viral envelope of RSV virus is composed of three proteins known as G glycoprotein, F glycoprotein, and SH protein. The G protein helps in attaching to host cells, the F protein is responsible for the fusion and entry into cells, while the SH protein is not necessary for these processes. Additionally, the virus contains five structural proteins, namely large (L) protein, nucleocapsid (N), phosphoprotein(P), matrix(M) , M2-1, and two non-structural proteins, NS1 and NS2 (Borchers et al., 2013). In the figure 2 genetic structure of RSV is shown.

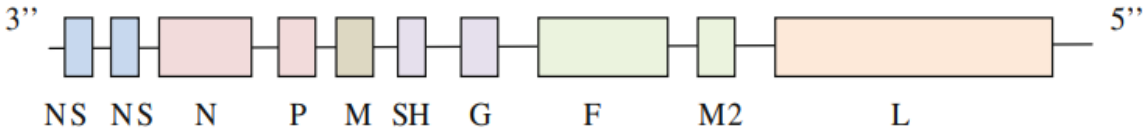


Figure 2: The genetic structure of RSV (Borchers et al., 2013)

NS = Nonstructural proteins N = Nucleoprotein P = Phosphoprotein M = Matrix protein SH = Small hydrophobic protein G = Glycoprotein F = Fusion protein M2 = Matrix protein L = Large protein.

Assembly and stability of virion structures are dependent on the viral matrix (M) protein. The nucleoprotein (N), phosphoprotein (P), and RNA-dependent RNA polymerase (L) are all components of the virion that work together to protect the viral genome and facilitate replication (Pandaya et al., 2019).

1.2 Pathogenesis of Human Respiratory Syncytial Virus

According to an article of Centers for Disease Control and Prevention RSV is a single-stranded RNA virus that primarily infects the respiratory tract. The exact pathogenesis of RSV is not fully understood, but it is believed to involve a complex interplay between the virus and the host immune response. After the virus enters the body, it binds to and infects epithelial cells in the respiratory tract, leading to the release of cytokines and chemokines that attract immune cells to the site of infection. This results in inflammation and swelling of the airways, which can cause difficulty breathing and coughing (CDC, 2021).

In severe cases, the virus can lead to the formation of multinucleated giant cells, also known as syncytia, which can obstruct the airways and lead to respiratory distress. The virus can also cause damage to the epithelial cells and impair the function of cilia, which are the tiny hair-like structures that help clear mucus from the airways. HRSV spreading from cell to cell may involve the entire respiratory tree, reaching bronchioles 1–3 days after the onset of rhinorrhea. Bronchiolar epithelium replication results in ciliated cell necrosis, syncytia formation, peribronchiolar inflammation with abundant lymphocytes and macrophages, impaired secretion clearance, small airway obstruction, and lung hyper aeration. HRSV's NS1 and NS2 non-structural proteins contribute to pathogenesis by inhibiting type-1 interferon (IFN) production. Human rhinovirus (HRSV) may also contribute to the severity of disease by modulating surfactant expression in human pulmonary epithelial cells (Proença-Módena et al., 2011). In addition to local respiratory effects, RSV can also trigger a systemic immune

response, leading to fever, decreased appetite, and fatigue. In some cases, the virus can also cause secondary bacterial infections, which can exacerbate the severity of the illness (Collins & Graham, 2008).

1.3 Viral replication and immune response of Human Respiratory Syncytial Virus

RSV replication begins with the attachment of a virus particle to a host cell, typically a ciliated epithelial cell in the nasal epithelium, mediated by G (Figure 3). When F is present, RSV is able to fuse with the plasma membrane and enter the cell (Srinivasakumar et al., 1991). The nucleocapsid is released into the cytoplasm after the viral envelope has been integrated into the cell membrane. Viruses rely on their nucleocapsids for transcription and replication (Ghildyal et al., 2006). RSV transcription and replication shares general characteristics with other paramyxoviruses. The N, P, and L proteins form a polymerase complex that is responsible for both viral replication and transcription. Gene start (GS) and termination (GE) signals (Kuo et al., 1997), both short conserved sequences that flank each coding unit of an mRNA, and the M2-1 anti-termination protein direct transcription (Fearns et al.,1994).

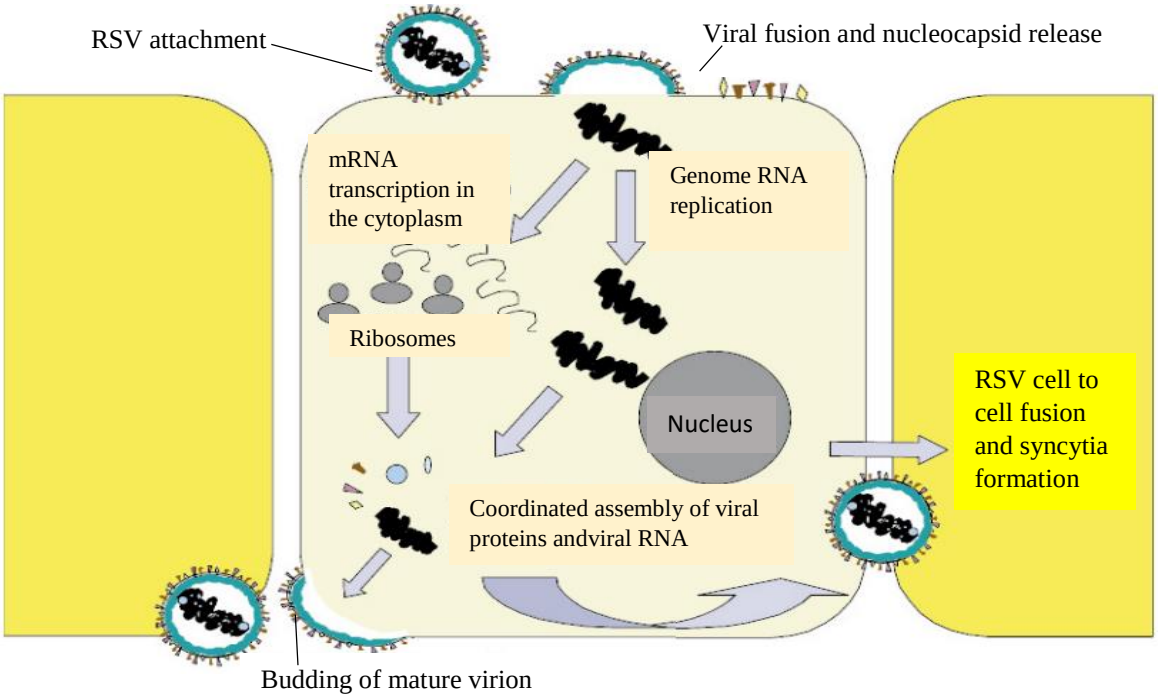


Figure 3: Schematic representation of the RSV life cycle (Hacking & Hull, 2002).

When the polymerase enters a read-through mode, where the transcription signals are ignored,

RNA replication takes place. So, a positive-sense replicative intermediate, also known as an anti-genome, is synthesized and used as a blueprint for the construction of subsequent negative-sense genomes. The M2-2 and perhaps NS1 proteins control RNA synthesis, facilitating transcription, replication, and the generation of genomic RNA (Bermingham & Collins, 1999). Human infections with RSV are characterized by epithelial damage and cellular desquamation, both of which are triggered by the immune response to RSV infection in epithelial cells (Spann et al., 2004). Glycoprotein G of the RSV envelope attaches to epithelial cells of the respiratory tract via glucosamine glycan expressed on the cell surface, and F interacts with antigen-presenting cells (macrophages, dendritic cells) via the Toll-like receptor 4 (TLR4) protein. This triggers the production and release of antiviral interferons (IFN-α, IFN-β, IFN-γ) and a cascade of pro inflammatory cytokines and chemokines (Munoz & Englund, 2019).

The production of the initial innate cytokine response is one of the earliest responses of the host to an infection. Type I IFNs, of which IFN-α and IFN-β are prominent members, are among these cytokines, as are IFN-λ1, IFN-λ2, and IFN-λ3 (Spann et al., 2004). The expression of proinflammatory cytokines and chemokines is part of the innate immune response to HRSV, which is triggered by recognition via Toll-like receptor (TLR)4, TLR3, TLR2, and RIG-I (Proença-Módena et al., 2011). Nonstructural (NS) proteins NS1 and NS2 are expressed by RSV and inhibit the production of type I interferons (IFNs) such as IFN-λ1, IFN-λ2, and IFN-λ3. Virus replication is stifled in cultured cells that generate type I IFNs when the NS1 and NS2 genes are removed from the RSV genome. These findings provide further evidence that the ability to regulate viral replication is tied to the NS1 and NS2 inhibition of the host cytokine response (Peebles & Graham, 2005).

Chapter 2

Methodology

The step-by-step methods which were utilized in this in-silico vaccine design for the RSV are illustrated in Figure 4:

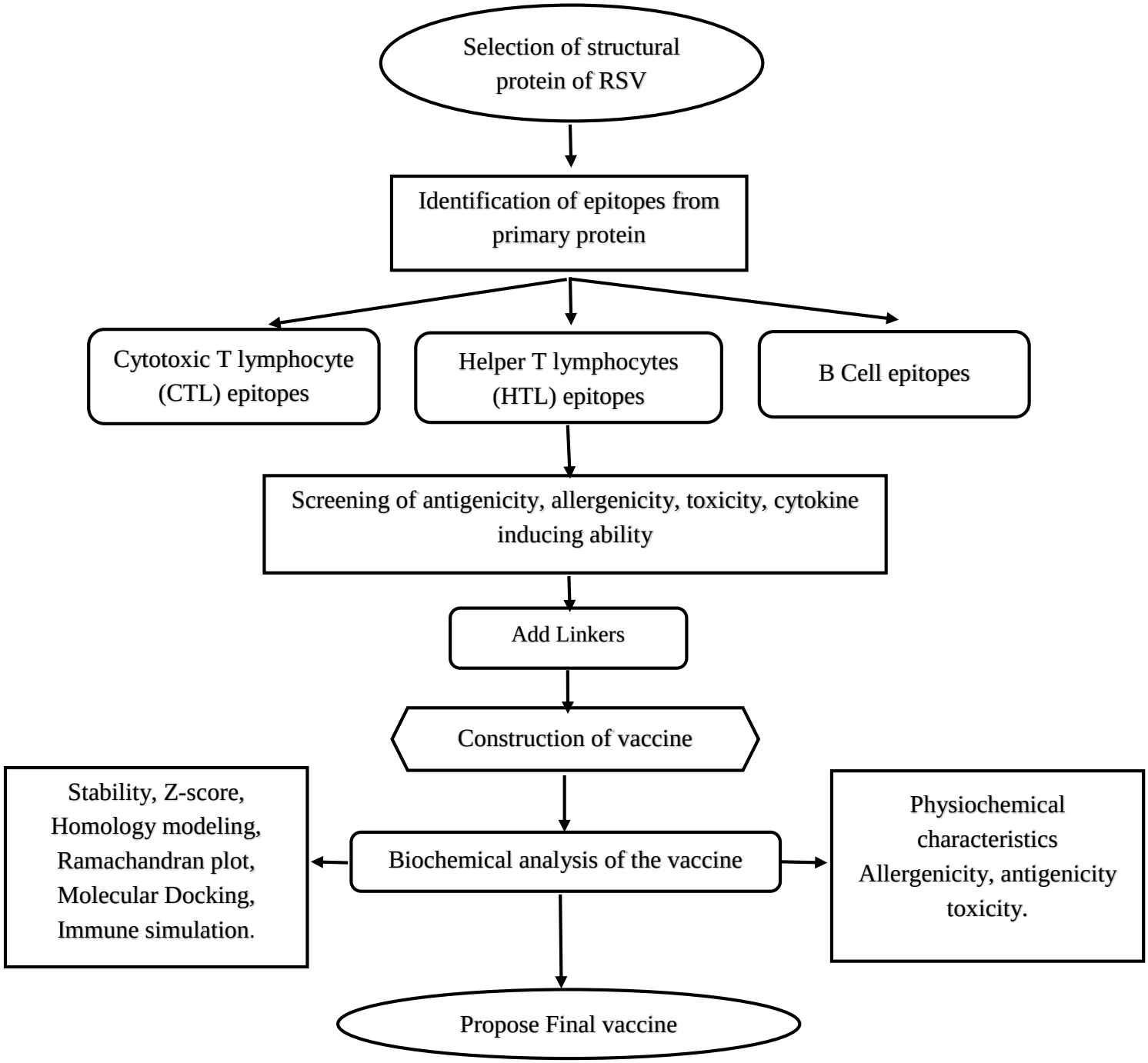


Figure 4: Step-by-step methods used in in-silico vaccine design for RSV.

In-silico methods include databases, quantitative structure-activity relationships, pharmacophores, homology modeling and other molecular modeling approaches, data mining, machine learning, network analysis tools and data analysis tools that use a computer. The model is developed and tested primarily through the use of in-silico techniques, with in vitro data generation serving as a supporting tool (Ekins et al., 2007)

2.1 Selection and Collection of Envelope Protein Sequence of HRSV

The Universal Protein Resource or UniProt (https://www.uniprot.org/) was utilized for assembling numerous protein sequences of RSV. The UniProt database contains a plethora of information, including protein sequences and their extensive annotation. Over 60 million sequences are stored in the database, with over half a million of those sequences having been manually curated by experts who examine experimental and predicted data for each protein. The rest are annotated mechanically using rule systems that are fed by the expert-curated knowledge. There has been a significant increase in the number of reference proteomes since last update in 2014, from 2,061 to 5,631. This has allowed for a more comprehensive coverage of taxonomic diversity (Bateman et al., 2017). According to uniprot website The UniProt databases are consisted with the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc).

From the Uniprot server search bar 'Respiratory syncytial virus' was searched and the organism was selected 'human'. In the search result 30,417 different protein was viewed. Among all the protein only fusion protein was targeted. The reason behind targeting this protein was fusion (F) glycoprotein control the initial phases of infection and F causes the virion membrane to fuse with a target cell membrane. The F protein is the major target for antiviral drug development (Jason S. McLellan1, William C. Ray2, 2014). In this search criteria there were four individual fusion protein and according to Uniprot unique and stable entry identifier they are P03420, O36634,P13843, P11209. Strain B proteins O36634 and P13843 were eliminated from consideration alongside the other two. P03420 and P11209, both protein of strain A, were the primary candidates for selection. In the end, P11209 prevailed over P03420. The reason behind it P11209 was updated protein than P03420. According to Uniprot P03420 was last updated in 1986-07-21 and P11209 was last updated in 1996-02-01. Thus, P11209 was the final candidate for in-silico vaccine approach. Recommended name for this protein Fusion glycoprotein (F0), gene name F and organism Human respiratory syncytial virus A (strain RSS-2). Its amino acid length was 574 and mass was 63334 Dalton or 63.33 kDa (Uniprot).

Taxonomy lineage for this protein:
Viruses > Riboviria > Orthornavirae > Negarnaviricota > Haploviricotina > Monjiviricetes > Mononegavirales (negative-sense genome single-stranded RNA viruses) > Pneumoviridae > Orthopneumovirus > Human respiratory syncytial virus > Human respiratory syncytial virus A (NCBI).
The sequences of this protein was collected from the server in fasta format. The collected protein sequence's antigenicity was predicted using Vaxijen 2.0. Protective antigen prediction in the Vaxijen server does not rely on sequence alignment (Doytchinova & Flower, 2007). The 'virus' was chosen as the organism to study on the server. The threshold point was established as 0.5. It was determined whether or not the proteins were likely antigen or non-antigen. Antigenicity prediction was taken into account either it meets the threshold or not.

2.2 Identifying Cytotoxic T Lymphocyte (CTL) Epitopes

NetCTL1.2 (https://services.healthtech.dtu.dk/service.php?NetCTL-1.2) was used to identify 9-mer CTL epitopes in the preferred vaccine candidate and using a default threshold value of 0.75. Effectiveness of TAP transport, binding of peptides to MHC class 1 proteasomal C-terminus cleavage are the three factors that influence the accuracy of NetCTL 1.2 predictions (Khalid et al., 2022). This server's predictions for CTL epitopes are limited to the 12 MHC (major histocompatibility complex) class I super types. In order to anticipate MHC class I binding and proteasomal cleavage, this server employs artificial neural network (Larsen et al., 2007). The selected protein's primary sequence was provided in FASTA format. The 'sort by score' option was used to arrange the CTL epitopes in descending order of prediction score, and the 'Combined score' option was selected. T-cell epitope and MHC-binding peptide identification play critical roles in vaccine development (Lin et al., 2008).

Then, toxicity prediction was performed because predicting the toxicity of therapeutic peptides/proteins before they are synthesized is crucial for reducing the time and resources needed to develop peptide/protein-based drugs (Gupta et al., 2013). After getting the prediction result from the NetCTL1.2 server predicted epitopes were screened for toxicity by Toxinpred (http://crdd.osdd.net/raghava/toxinpred/) server either they are toxic or non-toxic. Both potentially toxic and non-toxic peptides can be predicted and designed with the help of ToxinPred (Gupta et al., 2013). Besides, allergenicity (AllerTOP v.2.0) and antigenicity (Vaaxijen 2.0) were also checked for further filtration.

2.3 Determination of MHC class I Alleles

The Major histocompatibility complex (MHC) is a fundamental cell surface protein of the cellular immune system of vertebrates. MHC molecules are primarily responsible for displaying peptides (small protein fragments) to the intercellular space after binding to them. An immune response can be triggered and the compromised cell can be lysed if T cells recognize and bind to a peptide-MHC complex (Reynisson et al., 2021). This most recent release of NetMHCpan is version 4.1. The NetMHCpan-4.1 server can predict peptide binding to any MHC molecule based on its sequence using ANNs (Buckner et al., 2016).

Prediction of the CTL epitope-specific MHC I alleles was performed using the NetMHCpan-4.1 server (https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1). NetMHCpan-4.1 server predicts binding of peptides to any MHC molecule of known sequence using artificial neural networks (ANNs). Now, in the server 'peptide length' input given was 9 mer peptides. From the 'select allele(s)' list, all MHC molecules were selected. There was a 0.5 threshold for strong binders and a 2 threshold for weak binders by default. A binding affinity (BA) was included, and the server was set to sort results by prediction scores. A list of 'alleles and corresponding peptides' was included in the output. From the result page only the strong binder (sb) were selected.

2.4 Identifying Helper T Lymphocyte (HTL) Epitopes

Helper T lymphocyte (HTL) epitopes were identified with the help of the server NetMHCIIpan 4.0 (https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0). This server employs neural network technology to make predictions about how a given peptide sequence will bind to the MHC II molecule (Reynisson et al., 2020). The candidate protein was placed in FASTA format and in the 'select alleles' section maximum 20 alleles were allowed to select in the first entry and later rest of the alleles will be allowed as this process will repeat again. Threshold for strong binders it was 1 and for the weak binders it was 5. Then binding affinity was marked and entered for the output. From the output page only the strong binders were selected primarily and overlapping epitopes were taken in count. After that, all the SBs from two operations were collected and repeated strong binding HTL were excluded from the list.

2.5 Cytokine Stimulating Ability of Strong Binding HTL Epitopes

After predicting the helper T lymphocytes, they are needed to go through a further screening because HTLs secrete cytokines that can activate immune cells, such as interferon-gamma

(IFN- γ), interleukin 4 (IL-4), and interleukin 10 (IL-10). Furthermore, HTL-secreted cytokines can outlast inflammatory responses and protect tissues from damage (Bhuiyan et al., 2021). Vaccine research relies heavily on the discovery of HTL epitopes that can trigger the production of cytokines. Therefore, here IFNepitope server was used to incorporate the epitopes that induce IFN-γ or not (Dhanda, Vir, et al., 2013). A comparison of IFN gamma and non-IFN gamma was used for this server. In order to effectively instigate IFN gamma-inducing T-helper cells through in-silico subunit vaccine design, it is crucial to identify MHC class II binders that can accomplish this. In addition, IL4pred and IL10pred servers were used to determine whether or not the peptides induced IL-4 and IL-10 (Dhanda, Gupta, et al., 2013; Nagpal et al., 2017). The HTL epitopes' potential to induce interleukin-4 (IL-4) production was then predicted using the IL-4pred server. In the server, a hybrid model (SVM+motif) was selected. It is important to think about the inducibility of HTL epitopes by IL-4 because of the many roles that IL-4 plays, including the production of IgE, the switching of antibody isotypes, and the proliferation of antigen-presenting cells. To determine if the HTL epitopes were interleukin-10 (IL-10) inducing, the IL-10Pred server was used (Mathew et al., 2021; Nagpal et al., 2017).

2.6 Determination of B-cell Epitopes

Distinguishing linear B lymphocyte (LBL) epitopes is an additional critical step in developing epitope-based vaccine constructs, as B cell activation is essential for the activation of the humoral immune response and the generation of plasma cells against a specific antigen (Manavalan et al., 2018). When a protein's epitopes are recognized by B cells, they can activate a humoral immune response (Chen & Jensen, 2008). Here, IEDB Analysis Resource tool (http://tools.iedb.org/main/bcell/) was used to predict the B cell epitopes. ‘Bepipred linear epitope prediction 2.0’ was chosen as the method and plain format of the protein sequence was submitted. In the result page 22 different peptides with different length was found with a reciprocal graph of it.

2.7 Assembly of Vaccine Candidates

First of all, to construct the very first multi epitope vaccine combination all the CTL, HTL and B-cell epitopes which were screened added with the peptide linkers. All of the vaccine constructs included four peptide linkers: EAAAK, AAY, GPGPG, or KK. The first CTL epitope was linked to the primary protein sequence via the EAAAK linker. Connecting pairs

of CTL, HTL, and B cell epitopes were linkers of AAY, GPGPG, and KK respectively (Bhuiyan et al., 2021). The unit of the linker can repeat several times with respect of epitope count. The progression went as follows:

Protein sequence > EAAAK > CTL epitope 1 >AAY > CTL epitopes 2 ……> GPGPG > HTL epitope 1 > GPGPG > HTL epitopes 2…… > KK > B-Cell epitopes 1 > KK > B-cell epitopes 2 >KK> B-cell epitopes 3……

Furthermore, several vaccine combinations were constructed by assembling all the epitopes in different orientation. Additionally, some of the combinations were constructed by omitting few epitopes based on allergenicity, toxicity, binding affinity or sequence length.

2.8 In-silico Biochemical Analysis of Candidate Vaccines

To analyze the biochemical properties of all the constructed vaccines ProtParam tool (https://web.expasy.org/protparam/) was used. The ProtParam program was used to determine several chemical and physical characteristics of a protein based on its sequence. There is no need to acquire any further protein-related data (Gasteiger et al., 2003). Amino acid composition, atomic composition, molecular weight, theoretical pI, extinction coefficient, estimated half-life, aliphatic index, grand average of hydrophobicity (GRAVY), and instability index were all predicted for the vaccine constructs using Expasy Protparam (Rahman et al., 2020). At the very first, simply the sequence of the vaccine is needed to be submitted and this server will compute all the biochemical attributes.

2.9 Assessment of Toxicity and Allergenecity

Toxin and Toxin Target Database or in short T3DB (http://www.t3db.ca/), was used to determine whether or not the designed vaccines contained any toxins. T3DB is a one-of-a-kind bioinformatics resource as it gathers data on all known toxins and the organisms that are susceptible to them (Lim et al., 2009).In the server, the vaccine sequence was submitted in FASTA format and the BLAST parameters were stayed default. In the search result it shows if it has any target toxin with the name and types.

The allergenOnline server (http://www.allergenonline.org/) has been used for forecasting the allergenicity of the designed vaccines. It's software that can tell, from an amino acid sequence or a predicted protein structure, whether or not a protein has the potential to cause an allergic

reaction. This server uses bioinformatics techniques, such as the FASTA or BLASTP local alignment tools, to compare user-submitted protein sequences to databases of known allergens. (Goodman et al., 2016). In the server vaccine sequence was given in FASTA format and 'Show Z score' was selected while searching. For further clarification an another server AllerTopv2.0 (https://www.ddg-pharmfac.net/AllerTOP/) was used for predicting if any allergen entity presents in designed vaccine. In-silico allergen prediction using physicochemical properties of protein sequences is a relatively new field, and AllerTOP v2.0 is the first server to do so without requiring proper alignment (Dimitrov et al., 2014).

2.10 Three-Dimensional Model Creation via Homology Modeling

The 3D structures of the manufactured vaccines were visualized with the help of a Phyre2 server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). The Phyre 2 server uses a remote homology detection strategy to build 3D models and locate ligand binding sites. There are many separate software components that work together to form the Phyre2 system. The service's most popular feature is its ability to predict the three-dimensional structure of a single protein sequence (Kelley et al., 2016).

This server illustrates several 3D model of a single sequence and shows alignment coverage, confidence and the quality of the 3D models that it provides. In the server the vaccine sequence was given in the normal modeling mode with an email address. Phyre2 server will provide a PDB file via that e-mail with the necessary information after sometimes. For smooth visualization and inspection of the 3D models another software 'BIOVIA Discovery Studio Visualizer' was used.

2.11 Generating Ramachandran Plots and Z-score Evaluation of Vaccine

The stereo chemical accuracy and validation of homology model can be evaluated with the help of the Ramachandran plot. It is a two-dimensional plot of the torsional angles of amino acids φ (phi) and ψ (psi) in a protein sequence. The conformation angles phi (φ) and psi (ψ) also known as the Ramachandran angles, define the backbone of the polypeptide chain in a protein molecule. The Ramachandran Plot compares these orientations (P. Kumar & Arya, 2018; Sheik et al., 2002). The constructed vaccines' Ramachandran plots and other data were generated in the structure assessment portal of the SWISS-MODEL server (https://swissmodel.expasy.org/assess/). Here, the PDB file obtained previously from the

phyre2 server was uploaded. After that, Z score value was determined from the ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php) or protein structure analysis web. The z-score designates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations (Wiederstein & Sippl, 2007). In the ProSA webpage the PDB file was submitted and in the result page Z-score value was displayed incorporate with two graph. First graph was number of residues vs z-score and the second one was sequence position vs knowledge-based energy

2.12 Molecular Docking of Vaccine with desired Human Receptors

Molecular-docking algorithm PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php) was used to assure receptor binding affinity with the designed vaccine. The PatchDock algorithm is a geometric approach of molecular docking (Schneidman-Duhovny et al., 2005). In the input fields of the server 'Receptor molecule' was filled with PDB id '3W3G' which indicates the human toll like receptor 8 (TLR8) and 'Ligand molecule' was filled with the PDB file of the constructed vaccine. Additionally, an e-mail address was inputted in the required field. Rest of the field remained default. With the result a pdb file was received via e-mail which contained solution no., score, area, 3D transformation, PDB file of the complex.

2.13 Immune Response Simulations

C-ImmSim (https://kraken.iac.rm.cnr.it/C-IMMSIM/) is an immune simulation and vaccine expression analysis tool which used to predict whether or not a vaccine would produce an adequate immune response if given under realistic conditions (Khalid et al., 2022). The cellular and humoral responses of the mammalian immune system to an invading antigen can be obtained through this agent-based model implementation (Rahman et al., 2020). This server is used for immune simulation, which results in multiple graphs predicting antigen, immunoglobulin, lymphocyte, and other immune cell levels over time after vaccination doses (Rapin et al., 2010). The vaccination schedule was entered into the C-ImmSim server, along with three doses. Injections 1, 2, and 3 each had a time step of 84, 168, and 168, respectively. Hours of time are represented by each discrete time interval. An injection schedule of three shots spaced 28 days apart was decided upon for the vaccination.

2.14 Remarks regarding Methodology

Numerous in-silico screening and computational analyses have been made possible by the massive amounts of genomic data obtained as a result of sequencing projects. Today, computational analysis is the foundation for vaccine design due to its efficiency and low cost. This study used an in-silico method to carry out each stage of this study of vaccine development and the subsequent biochemical analysis. Thus, numerous number of online server and software were utilized throughout the vaccine construction process and also a plethora of research articles were reviewed in order to verify and validate the in-silico approach of Fusion Glycoprotein Targeted Multi-Epitope Vaccine against Human Respiratory Syncytial Virus.

Chapter 3

Results

3.1 Antigenicity of Selected Protein

At the very first, P03420, O36634, P13843, and P11209 are the unique and stable identifiers for the four separate fusion proteins that were found in the uniprot server. In addition to the other two, the strain B proteins O36634 and P13843 were ruled out of consideration because strain A is more responsible for disease development. Primary candidates included strain A proteins P03420 and P11209. Finally, P11209 (strain RSS-2) was chosen over P03420 because it was the most updated protein sequence. The sequence of the P11209 protein is given below-

MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSKGYLSALRTGWYTSVITIE
LSNIKENKCNGTDAKVKLIKQELDKYKSAVTELQLLMQSTPATNNRARRELPRFMN
YTLNNTKNTNVTLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVNKIKSALLST
NKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIVNKQSCSISNIETVIEFQQKNNRLLEI
TREFSVNAGVTTPVSTYMLTNSELLSLINDMPITNDQKKLMSNNVQIVRQQSYSIMSII
KEEVLAYVVQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTDRGWYCDNAGS
VSFFPLAETCKVQSNRVFCDTMNSLTLPSEVNLCNIDIFNPKYDCKIMTSKTDVSSSVI
TSLGAIVSCYGKTKCTASNKDRGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQE
GKSLYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDELLHNVNAGKST
TNIMITTIIIVIIVILLSLIAVGLLLYCKARSTPVTLSKDQLSGINNIAFSN

In the vaxijen 2.0 server antigenicity was checked either it is probable antigen or not. The antigenicity of this sequence was predicted to be 0.5679 (Probable ANTIGEN) in vaxijen 2.0 (Figure:5) (Doytchinova & Flower, 2007).

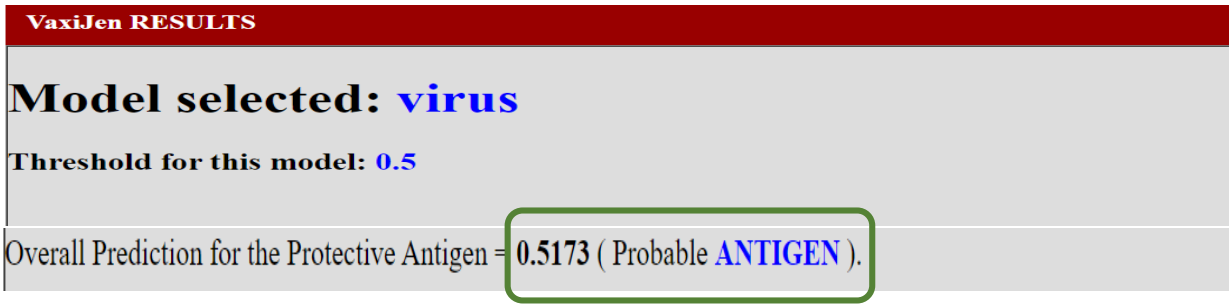


Figure 5: Antigenicity of the Fusion protein (strain RSS-2) of Human Respiratory Syncytial Virus (Doytchinova & Flower, 2007).

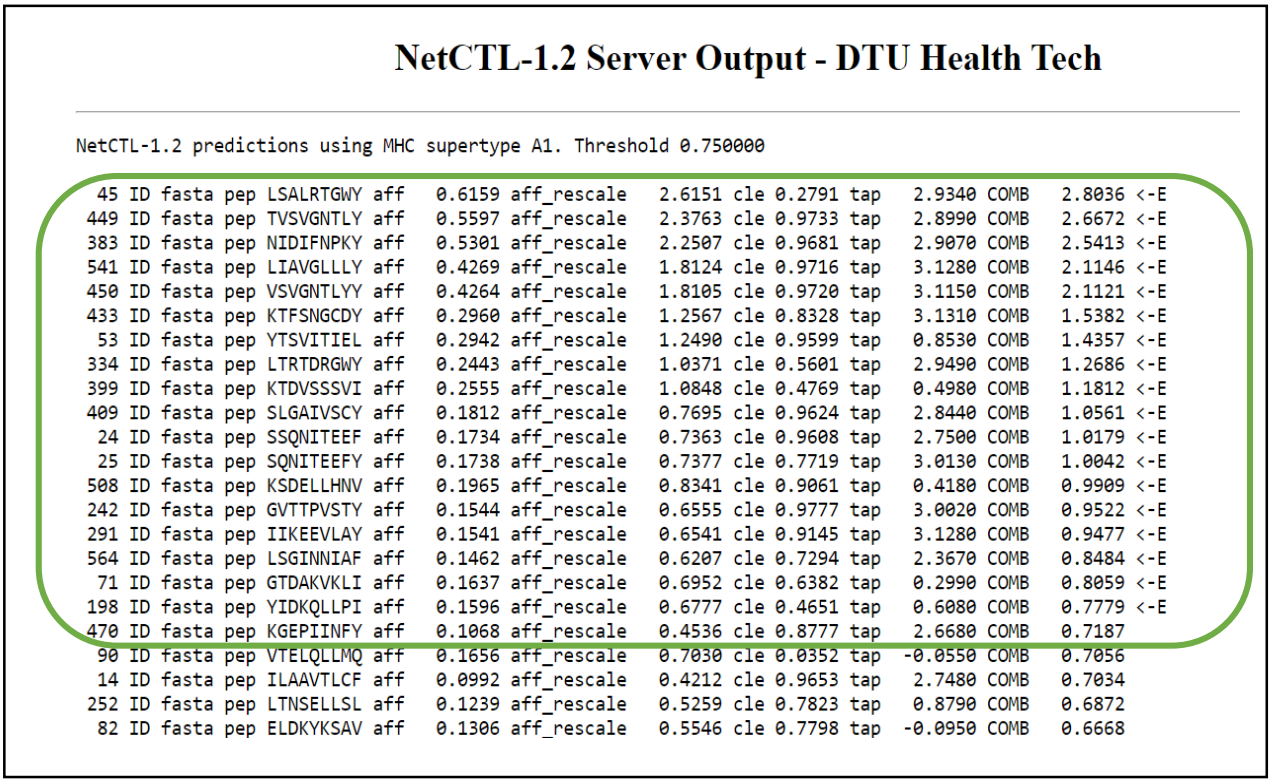
As this fusion protein shows the required antigenicity it will be the base protein sequence for the further study.

3.2 Identifying CTL Epitopes

From the NetCTL1.2 servers, using MHC super type A1 and a threshold of 0.75, 18 CTL epitopes were identified in the fusion protein sequence of RSV. CTL epitopes with their corresponding combined scores are shown in Table 1.

Table 1: CTL epitopes with their corresponding combined scores.

CTL Epitope (NetCTL1.2)	**Combined Score (NetCTL1.2)**
LSALRTGWY	2.8036
TVSVGNTLY	2.6672
NIDIFNPKY	2.5413
LIAVGLLLY	2.1146
VSVGNTLYY	2.1121
KTFSNGCDY	1.5382
YTSVITIEL	1.4357
LTRTDRGWY	1.2686
KTDVSSSVI	1.1812
SLGAIVSCY	1.0561
SSQNITEEF	1.0179
SQNITEEFY	1.0042
KSDELLHNV	0.9909
GVTTPVSTY	0.9522
IIKEEVLAY	0.9477
LSGINNIAF	0.8484
GTDAKVKLI	0.8059
YIDKQLLPI	0.7779



NetCTL-1.2 Server Output - DTU Health Tech

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

45	ID	fasta	pep	LSALRTGWY	aff	0.6159	aff_rescale	2.6151	cle 0.2791	tap	2.9340 COMB 2.8036 <-E
449	ID	fasta	pep	TVSVGNTLY	aff	0.5597	aff_rescale	2.3763	cle 0.9733	tap	2.8990 COMB 2.6672 <-E
383	ID	fasta	pep	NIDIFNPKY	aff	0.5301	aff_rescale	2.2507	cle 0.9681	tap	2.9070 COMB 2.5413 <-E
541	ID	fasta	pep	LIAVGLLLY	aff	0.4269	aff_rescale	1.8124	cle 0.9716	tap	3.1280 COMB 2.1146 <-E
450	ID	fasta	pep	VSVGNTLYY	aff	0.4264	aff_rescale	1.8105	cle 0.9720	tap	3.1150 COMB 2.1121 <-E
433	ID	fasta	pep	KTFSNGCDY	aff	0.2960	aff_rescale	1.2567	cle 0.8328	tap	3.1310 COMB 1.5382 <-E
53	ID	fasta	pep	YTSVITIEL	aff	0.2942	aff_rescale	1.2490	cle 0.9599	tap	0.8530 COMB 1.4357 <-E
334	ID	fasta	pep	LTRTDRGWY	aff	0.2443	aff_rescale	1.0371	cle 0.5601	tap	2.9490 COMB 1.2686 <-E
399	ID	fasta	pep	KTDVSSSVI	aff	0.2555	aff_rescale	1.0848	cle 0.4769	tap	0.4980 COMB 1.1812 <-E
409	ID	fasta	pep	SLGAIVSCY	aff	0.1812	aff_rescale	0.7695	cle 0.9624	tap	2.8440 COMB 1.0561 <-E
24	ID	fasta	pep	SSQNITEEF	aff	0.1734	aff_rescale	0.7363	cle 0.9608	tap	2.7500 COMB 1.0179 <-E
25	ID	fasta	pep	SQNITEEFY	aff	0.1738	aff_rescale	0.7377	cle 0.7719	tap	3.0130 COMB 1.0042 <-E
508	ID	fasta	pep	KSDELLHNV	aff	0.1965	aff_rescale	0.8341	cle 0.9061	tap	0.4180 COMB 0.9909 <-E
242	ID	fasta	pep	GVTTPVSTY	aff	0.1544	aff_rescale	0.6555	cle 0.9777	tap	3.0020 COMB 0.9522 <-E
291	ID	fasta	pep	IIKEEVLAY	aff	0.1541	aff_rescale	0.6541	cle 0.9145	tap	3.1280 COMB 0.9477 <-E
564	ID	fasta	pep	LSGINNIAF	aff	0.1462	aff_rescale	0.6207	cle 0.7294	tap	2.3670 COMB 0.8484 <-E
71	ID	fasta	pep	GTDAKVKLI	aff	0.1637	aff_rescale	0.6952	cle 0.6382	tap	0.2990 COMB 0.8059 <-E
198	ID	fasta	pep	YIDKQLLPI	aff	0.1596	aff_rescale	0.6777	cle 0.4651	tap	0.6080 COMB 0.7779 <-E
470	ID	fasta	pep	KGEPIINFY	aff	0.1068	aff_rescale	0.4536	cle 0.8777	tap	2.6680 COMB 0.7187
90	ID	fasta	pep	VTELQLLMQ	aff	0.1656	aff_rescale	0.7030	cle 0.0352	tap	-0.0550 COMB 0.7056
14	ID	fasta	pep	ILAAVTLCF	aff	0.0992	aff_rescale	0.4212	cle 0.9653	tap	2.7480 COMB 0.7034
252	ID	fasta	pep	LTNSELLSL	aff	0.1239	aff_rescale	0.5259	cle 0.7823	tap	0.8790 COMB 0.6872
82	ID	fasta	pep	ELDKYKSAV	aff	0.1306	aff_rescale	0.5546	cle 0.7798	tap	-0.0950 COMB 0.6668

Figure 6: CTL epitopes were found in NetCTL-1.2 server (Larsen et al., 2007).

3.3 Determination of MHC Class I Alleles Specific to CTL Epitopes

NetMHCpan 4.1was used to identify specific MHC 1 alleles to the CTL epitopes that were found in NetCTL 1.2. Based on the peptide sequence, the NetMHCpan-4.1 server can predict which MHC molecules the peptide will bind to (Buckner et al., 2016). Among all the 18 CTL epitopes 11 alleles were found where peptide length was 9mer, the threshold for strong binders was 0.5 and for weak binders was 2. With the repetition total 17 strong binding alleles were listed. They are:

Table 2: Strong binding MHC I alleles specific to CTL epitopes

Allele	Peptide	%Rank_EL	%Rank_BA	Aff(nM)
1 HLA-A*01:01	LSALRTGWY	0.461	0.146	195.69
1 HLA-A*01:01	TVSVGNTLY	0.143	0.228	361.90
1 HLA-A*01:01	NIDIFNPKY	0.065	0.118	153.11
1 HLA-A*01:01	VSVGNTLYY	0.098	0.089	104.97
1 HLA-A*01:01	KTFSNGCDY	0.495	0.475	1033.21

Allele	Peptide	%Rank_EL	%Rank_BA	Aff(nM)
1 HLA-A*02:01	KSDELLHNV	0.256	1.834	235.11
1 HLA-A*26:01	TVSVGNTLY	0.154	0.325	937.18
1 HLA-A*26:01	VSVGNTLYY	0.330	0.474	1441.55
1 HLA-A*26:01	GVTTPVSTY	0.117	0.810	2688.18
1 HLA-A*26:01	IIKEEVLAY	0.120	0.545	1700.70
1 HLA-B*58:01	VSVGNTLYY	0.259	0.340	60.65
1 HLA-B*58:01	SSQNITEEF	0.399	1.216	411.84
1 HLA-B*15:01	TVSVGNTLY	0.480	1.526	421.58
1 HLA-B*15:01	SLGAIVSCY	0.393	0.660	123.53
1 HLA-B*15:01	SQNITEEFY	0.086	0.759	146.78
1 HLA-B*15:01	GVTTPVSTY	0.087	1.270	322.36
1 HLA-B*15:01	IIKEEVLAY	0.009	0.211	38.93

3.4 In-silico Prediction of Antigenicity, Allergenicity, and Toxicity of CTL Epitopes

Toxinpred was used to predict the potential toxicity of the CTLs discovered in NetCTL1.2. All CTL epitopes were classified as non-toxins by the toxinpred server. Shown in Figure 7.

Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq_1	LSALRTGWY	-0.68	Non-Toxin	-0.04	0.09	-0.77	1.00	1066.35
seq_1	TVSVGNTLY	-0.98	Non-Toxin	0.06	0.53	-0.82	0.00	953.19
seq_1	NIDIFNPKY	-1.03	Non-Toxin	-0.12	-0.61	-0.22	0.00	1123.40
seq_1	LIAVGLLLY	-1.05	Non-Toxin	0.42	2.67	-1.48	0.00	974.40
seq_1	VSVGNTLYY	-0.94	Non-Toxin	0.08	0.47	-1.03	0.00	1015.26
seq_1	KTFSNGCDY	-0.63	Non-Toxin	-0.23	-0.98	0.03	0.00	1034.22
seq_1	YTSVITIEL	-1.19	Non-Toxin	0.15	1.11	-0.74	-1.00	1038.34
seq_1	LTRTDRGWY	-0.69	Non-Toxin	-0.39	-1.41	0.08	1.00	1167.41
seq_1	KTDVSSSVI	-0.46	Non-Toxin	-0.11	0.27	0.19	0.00	935.16
seq_1	SLGAIVSCY	-0.80	Non-Toxin	0.19	1.50	-0.92	0.00	912.19
seq_1	SSQNITEEF	-1.08	Non-Toxin	-0.21	-1.00	0.26	-2.00	1054.20
seq_1	SQNITEEFY	-1.37	Non-Toxin	-0.18	-1.06	-0.03	-2.00	1130.30
seq_1	KSDELLHNV	-1.03	Non-Toxin	-0.24	-0.73	0.43	-0.50	1054.30
seq_1	GVTTPVSTY	-1.10	Non-Toxin	0.04	0.24	-0.69	0.00	924.14
seq_1	IIKEEVLAY	-0.92	Non-Toxin	0.05	0.73	-0.08	-1.00	1077.42
seq_1	LSGINNIAF	-1.16	Non-Toxin	0.16	1.02	-0.86	0.00	948.22
seq_1	GTDAKVKLI	-0.87	Non-Toxin	-0.10	0.21	0.33	1.00	944.27
seq_1	YIDKQLLPI	-1.12	Non-Toxin	-0.00	0.31	-0.37	0.00	1102.48

1/1 100

Figure 7: ToxinPred results for CTL epitopes (Gupta et al., 2013).

The vaxijen2.0 server was used to test for antigenicity, and it found that only four CTL were likely to be an antigen (Table -3). In addition, AllerTOPv.2.0 was utilized to ensure that there were no allergens present, and with the exception of four CTL, there were none (Table -3).

Table 3: Antigenicity, toxicity and allergenicity of the CTL epitopes selected from NetMHCpan 4.1

CTL Epitope (NetCTL1.2)	Antigenicity (Vaxijen 2.0)	Toxicity (Toxin-pred)	Alllergenicity (AllerTOP v.2.0)
LSALRTGWY	1.11 (Probable ANTIGEN)	Non toxin	Probable ALLERGEN
TVSVGNTLY	0.8829(Probable ANTIGEN)	Non toxin	Probable ALLERGEN
NIDIFNPKY	0.782 (Probable ANTIGEN)	Non toxin	Probable ALLERGEN
VSVGNTLYY	1.05 (Probable ANTIGEN)	Non toxin	Probable ALLERGEN
KTFSNGCDY	0.08 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen
SLGAIVSCY	0.16 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen
IIKEEVLAY	0.06 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen
KSDELLHNV	0.1685 (Probable NON-ANTIGEN).	Non toxin	Probable Non-allergen
SSQNITEEF	0.17 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen
SQNITEEFY	0.16 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen
GVTTPVSTY	0.42 (Probable NON ANTIGEN)	Non toxin	Probable Non-allergen

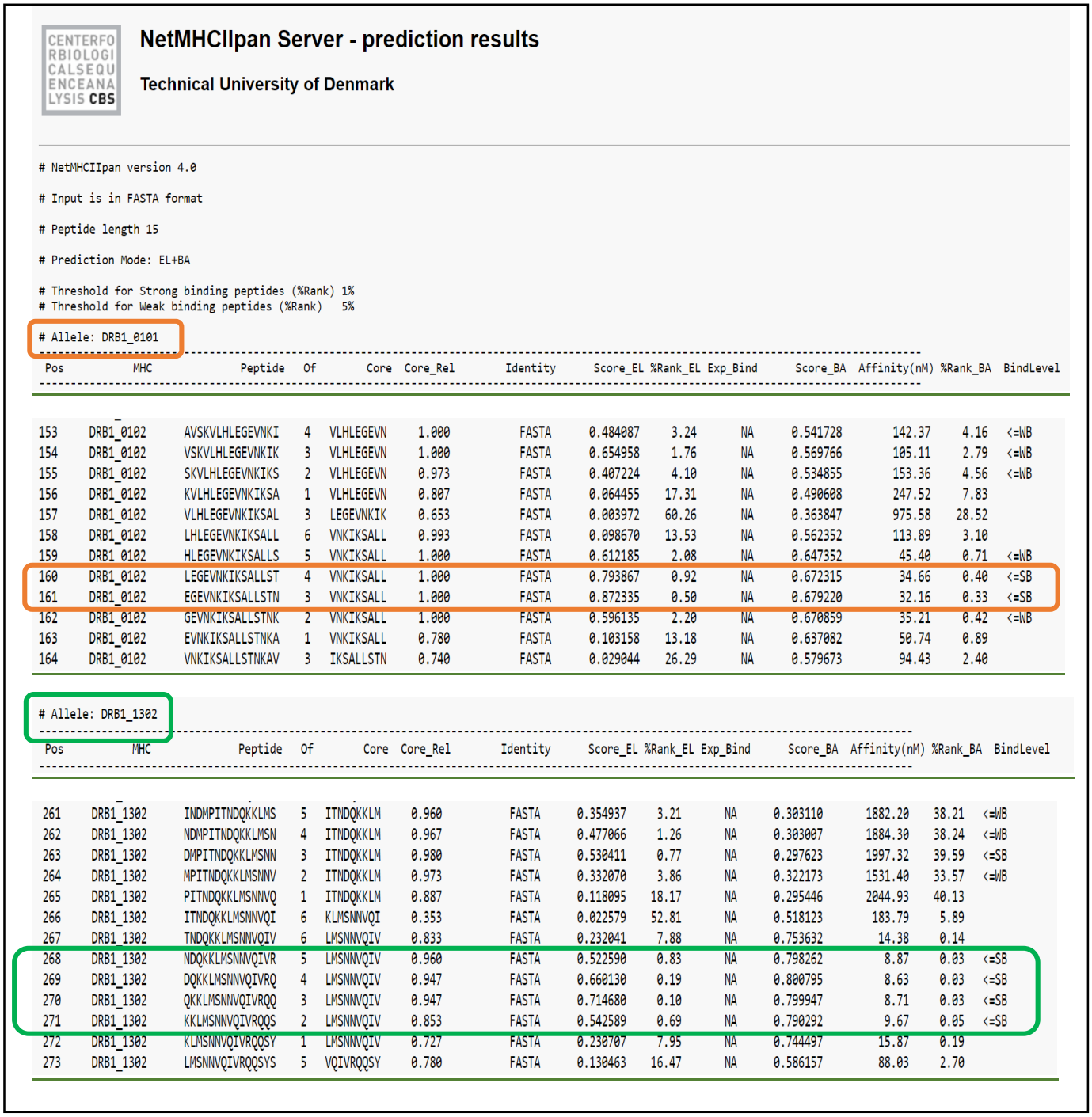
3.5 Identification of Strong Binding HTL Epitopes

From the NetMHCIIpan version 4.0 server HTL epitopes were identified. In that server protein sequence was inputted in fasta format, peptide length was 15mer, prediction mode was EL+BA. Right after submitting the work in the result page it shows peptide sequence, core of the peptide, percentile rank, affinity, binding level the corresponding MHC alleles. They are:

DRB1_0101,DRB1_0102,DRB1_0103,DRB1_0301,DRB1_0305,DRB1_0401,DRB1_040,DRB1_0403,DRB1_0404,DRB1_0405,DRB1_0408,DRB1_0701,DRB1_0801,DRB1_0803,DRB1_0901,DRB1_1001,DRB1_1101,DRB1_1104,DRB1_1201,DRB1_1301,DRB1_1302,D

RB1_1303,DRB1_1401,DRB1_1402,DRB1_1454,DRB1_1501,DRB1_1503,DRB1_1601,D
RB3_0101,DRB3_0202,DRB4_0101,DRB4_0103,DRB5_0101,DRB5_0202

In the Figure 8, Strong and weak binding peptides for allele DRB1_0101(first page) and DRB1_0302(first page) in NetMHCIIpan 4.0 Server is placed.



NetMHCIIpan Server - prediction results

Technical University of Denmark

```
# NetMHCIIpan version 4.0

# Input is in FASTA format

# Peptide length 15

# Prediction Mode: EL+BA

# Threshold for Strong binding peptides (%Rank) 1%
# Threshold for Weak binding peptides (%Rank)   5%

# Allele: DRB1_0101
```

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
153	DRB1_0102	AVSKVLHLEGEVNKI	4	VLHLEGEVN	1.000	FASTA	0.484087	3.24	NA	0.541728	142.37	4.16	<=WB
154	DRB1_0102	VSKVLHLEGEVNKIK	3	VLHLEGEVN	1.000	FASTA	0.654958	1.76	NA	0.569766	105.11	2.79	<=WB
155	DRB1_0102	SKVLHLEGEVNKIKS	2	VLHLEGEVN	0.973	FASTA	0.407224	4.10	NA	0.534855	153.36	4.56	<=WB
156	DRB1_0102	KVLHLEGEVNKIKSA	1	VLHLEGEVN	0.807	FASTA	0.064455	17.31	NA	0.490608	247.52	7.83	
157	DRB1_0102	VLHLEGEVNKIKSAL	3	LEGEVNKIK	0.653	FASTA	0.003972	60.26	NA	0.363847	975.58	28.52	
158	DRB1_0102	LHLEGEVNKIKSALL	6	VNKIKSALL	0.993	FASTA	0.098670	13.53	NA	0.562352	113.89	3.10	
159	DRB1_0102	HLEGEVNKIKSALLS	5	VNKIKSALL	1.000	FASTA	0.612185	2.08	NA	0.647352	45.40	0.71	<=WB
160	DRB1_0102	LEGEVNKIKSALLST	4	VNKIKSALL	1.000	FASTA	0.793867	0.92	NA	0.672315	34.66	0.40	<=SB
161	DRB1_0102	EGEVNKIKSALLSTN	3	VNKIKSALL	1.000	FASTA	0.872335	0.50	NA	0.679220	32.16	0.33	<=SB
162	DRB1_0102	GEVNKIKSALLSTNK	2	VNKIKSALL	1.000	FASTA	0.596135	2.20	NA	0.670859	35.21	0.42	<=WB
163	DRB1_0102	EVNKIKSALLSTNKA	1	VNKIKSALL	0.780	FASTA	0.103158	13.18	NA	0.637082	50.74	0.89	
164	DRB1_0102	VNKIKSALLSTNKAV	3	IKSALLSTN	0.740	FASTA	0.029044	26.29	NA	0.579673	94.43	2.40	

```
# Allele: DRB1_1302
```

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
261	DRB1_1302	INDMPITNDQKKLMS	5	ITNDQKKLM	0.960	FASTA	0.354937	3.21	NA	0.303110	1882.20	38.21	<=WB
262	DRB1_1302	NDMPITNDQKKLMSN	4	ITNDQKKLM	0.967	FASTA	0.477066	1.26	NA	0.303007	1884.30	38.24	<=WB
263	DRB1_1302	DMPITNDQKKLMSNN	3	ITNDQKKLM	0.980	FASTA	0.530411	0.77	NA	0.297623	1997.32	39.59	<=SB
264	DRB1_1302	MPITNDQKKLMSNNV	2	ITNDQKKLM	0.973	FASTA	0.332070	3.86	NA	0.322173	1531.40	33.57	<=WB
265	DRB1_1302	PITNDQKKLMSNNVQ	1	ITNDQKKLM	0.887	FASTA	0.118095	18.17	NA	0.295446	2044.93	40.13	
266	DRB1_1302	ITNDQKKLMSNNVQI	6	KLMSNNVQI	0.353	FASTA	0.022579	52.81	NA	0.518123	183.79	5.89	
267	DRB1_1302	TNDQKKLMSNNVQIV	6	LMSNNVQIV	0.833	FASTA	0.232041	7.88	NA	0.753632	14.38	0.14	
268	DRB1_1302	NDQKKLMSNNVQIVR	5	LMSNNVQIV	0.960	FASTA	0.522590	0.83	NA	0.798262	8.87	0.03	<=SB
269	DRB1_1302	DQKKLMSNNVQIVRQ	4	LMSNNVQIV	0.947	FASTA	0.660130	0.19	NA	0.800795	8.63	0.03	<=SB
270	DRB1_1302	QKKLMSNNVQIVRQQ	3	LMSNNVQIV	0.947	FASTA	0.714680	0.10	NA	0.799947	8.71	0.03	<=SB
271	DRB1_1302	KKLMSNNVQIVRQQS	2	LMSNNVQIV	0.853	FASTA	0.542589	0.69	NA	0.790292	9.67	0.05	<=SB
272	DRB1_1302	KLMSNNVQIVRQQSY	1	LMSNNVQIV	0.727	FASTA	0.230707	7.95	NA	0.744497	15.87	0.19	
273	DRB1_1302	LMSNNVQIVRQQSYS	5	VQIVRQQSY	0.780	FASTA	0.130463	16.47	NA	0.586157	88.03	2.70	

Figure 8: Strong and weak binding peptides for allele DRB1_0101(first page) and DRB1_0302(first page) in NetMHCIIpan 4.0 Server (Reynisson et al., 2021).

Total 223 Strong binding epitopes were found initially including overlapping core and repetition. Repeated HTL peptides were excluded and rest of the HTL epitopes were checked for their cytokine Stimulating Ability.

3.6 Cytokine Stimulating Ability of Obtained HTL Epitopes

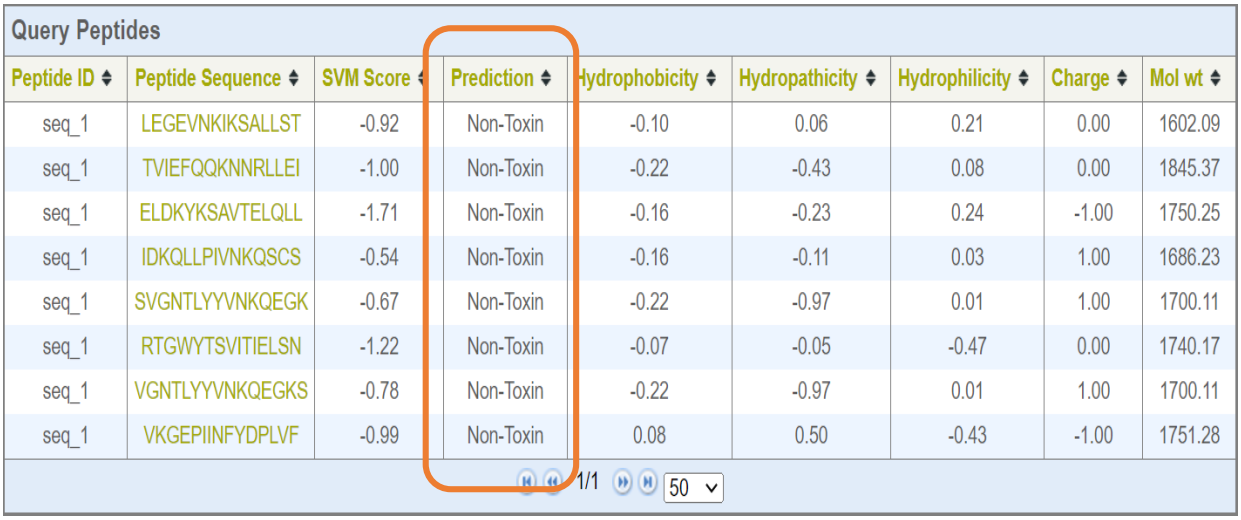
All strong binding HTL epitopes which were found and filtered in NetMHCIIpan4.0 were screened through IFNepitope, il4pred, il10pred servers for analyzing their cytokine stimulating ability if they are IL4, IL10 inducer and positive to IFN-gamma or not. In the table 4 only IL-4, IL-10, and IFN-gamma inducers HTL epitopes were placed.

Table 4: IL-4, IL-10, and IFN-gamma inducers HTL epitopes.

HTL Sequence (NetMHCIIpan-4.0)	**IFN (ifnepitope)**	**IL4 (il4pred)**	**IL10 (il10pred)**
LEGEVNKIKSALLST	Positive	Inducer	Inducer
TVIEFQQKNNRLLEI	Positive	Inducer	Inducer
ELDKYKSAVTELQLL	Positive	Inducer	Inducer
IDKQLLPIVNKQSCS	Positive	Inducer	Inducer
SVGNTLYYVNKQEGK	Positive	Inducer	Inducer
RTGWYTSVITIELSN	Positive	Inducer	Inducer
VGNTLYYVNKQEGKS	Positive	Inducer	Inducer
VKGEPIINFYDPLVF	Positive	Inducer	Inducer

3.7 In-silico Prediction of Antigenicity, Allergenicity, and Toxicity of HTL Epitopes

Total 8 HTL epitopes were found which met all the criteria for IL4, IL10 and IFN- γ. These 8 HTL epitopes then screened for antigenicity, toxicity and allergenicity with the servers Vaxijen 2.0, Toxinpred and AllerTOP v.2.0 respectively. In figure 9 toxinpred results were shown, all the HTL epitopes were non-toxic. For better effectiveness and to produce a desired or intended result these screenings are important.



Query Peptides

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
seq_1	LEGEVNKIKSALLST	-0.92	Non-Toxin	-0.10	0.06	0.21	0.00	1602.09
seq_1	TVIEFQQKNNRLLEI	-1.00	Non-Toxin	-0.22	-0.43	0.08	0.00	1845.37
seq_1	ELDKYKSAVTELQLL	-1.71	Non-Toxin	-0.16	-0.23	0.24	-1.00	1750.25
seq_1	IDKQLLPIVNKQSCS	-0.54	Non-Toxin	-0.16	-0.11	0.03	1.00	1686.23
seq_1	SVGNTLYYVNKQEGK	-0.67	Non-Toxin	-0.22	-0.97	0.01	1.00	1700.11
seq_1	RTGWYTSVITIELSN	-1.22	Non-Toxin	-0.07	-0.05	-0.47	0.00	1740.17
seq_1	VGNTLYYVNKQEGKS	-0.78	Non-Toxin	-0.22	-0.97	0.01	1.00	1700.11
seq_1	VKGEPIINFYDPLVF	-0.99	Non-Toxin	0.08	0.50	-0.43	-1.00	1751.28

1/1 50

Figure 9: ToxinPred results for HTL epitopes (Gupta et al., 2013).

In the table 5 HTL epitopes and their corresponding antigenicity, toxicity and allergenicity were tabulated.

Table 5: Antigenicity, toxicity and allergenicity of the HTL epitopes

HTL Sequence	Antigenicity (Vaxijen 2.0)	Toxicity (Toxin-pred)	Allergenicity (AllerTOP v.2.0)
LEGEVNKIKSALLST	0.1009 (Probable Non-Antigen)	Non toxic	Probable NON-ALLERGEN
TVIEFQQKNNRLLEI	1.2807 (Probable Antigen)	Non toxic	Probable ALLERGEN
ELDKYKSAVTELQLL	0.0309 (Probable Non-Antigen)	Non toxic	Probable ALLERGEN
IDKQLLPIVNKQSCS	0.6068 (Probable Antigen)	Non toxic	Probable ALLERGEN
SVGNTLYYVNKQEGK	0.6938 (Probable Antigen)	Non toxic	Probable ALLERGEN
RTGWYTSVITIELSN	0.7815 (Probable Antigen)	Non toxic	Probable ALLERGEN
VGNTLYYVNKQEGKS	0.7002 (Probable Antigen)	Non toxic	Probable ALLERGEN
VKGEPIINFYDPLVF	0.6139 (Probable Antigen)	Non toxic	Probable ALLERGEN

3.8 Determination of B-cell Epitopes

The B cell epitope prediction was made using the IEDB Analysis Resource tool. The protein sequence was submitted in plain format, and the program "Bepipred linear epitope prediction 2.0" was chosen as the method. There was 22 different peptide sequence with different length Shown in result page (Table 6).

Table 6: B cell epitopes prediction result (Bepipred Linear Epitope Prediction 2.0)

No.	Start	End	Peptide	Length
1	26	35	QNITEEFYQS	10
2	64	74	IKENKCNGTDA	11
3	76	76	V	1
4	81	88	QELDKYKS	8
5	94	134	QLLMQSTPATNNRARRELPRFMNY TLNNTKNTNVTLSKKRK	41
6	200	202	DKQ	3
7	205	205	P	1
8	207	210	VNKQ	4
9	212	215	CSIS	4
10	238	250	SVNAGVTTPVSTY	13
11	262	283	NDMPITNDQKKLMSNNVQIVRQ	22
12	328	329	EG	2
13	338	341	DRGW	4
14	349	353	VSFFP	5
15	367	378	CDTMNSLTLPSE	12
16	383	393	NIDIFNPKYDC	11
17	396	402	MTSKTDV	7
18	429	437	RGIIKTFSN	9
19	460	490	NKQEGKSLYVKGEPIINFYDPLVFPSDEFDA	31
20	496	508	NEKINQSLAFIRK	13
21	511	520	ELLHNVNAGK	10
22	554	571	STPVTLSKDQLSGINNIA	18

Along with the peptide sequence table a graph of position vs score was also given in the result page. According to the graph average score of B cell epitopes was 0.463, maximum 0.648 and minimum score was 0.173 (Table 10).

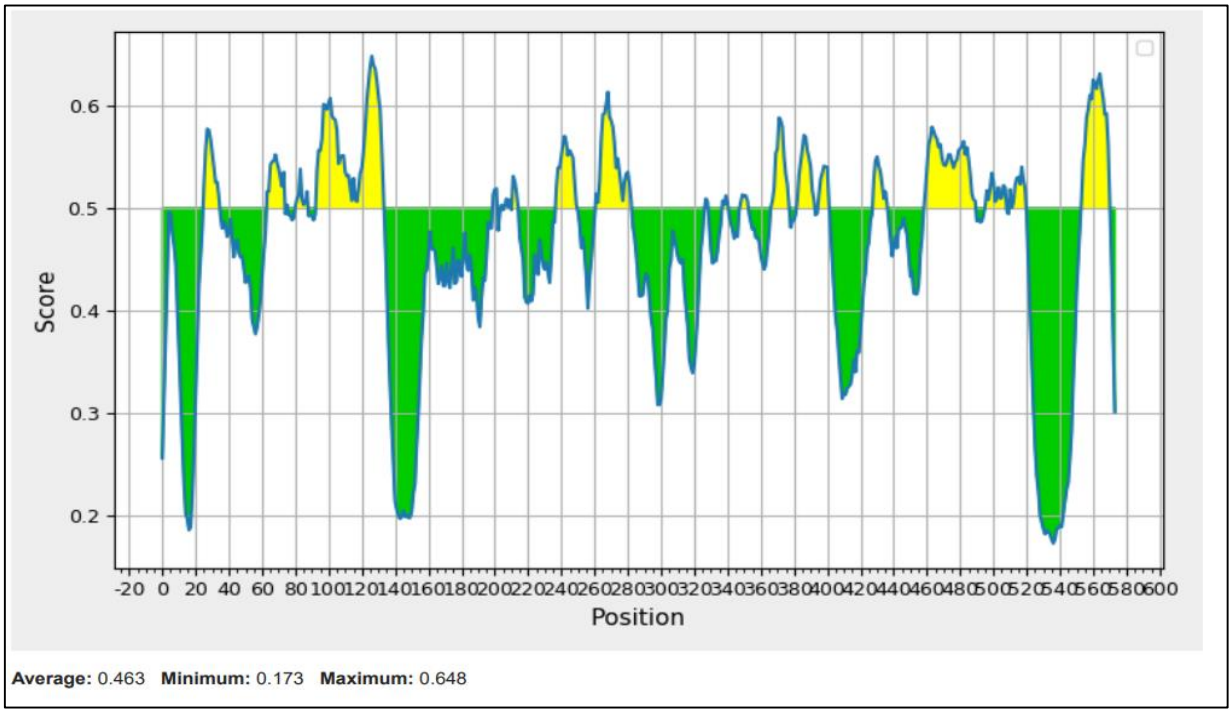


Figure 10: Score vs Position graph of B cell epitopes from IEDB Analysis Resource tool (Haste Andersen et al., 2006).

3.9 Assembly of Vaccine Candidates

After all of the epitopes have been carefully chosen, now they can be combined into a multi-epitope vaccine. The sequences of amino acids can be put in different orders, but all CTL, HTL, and B-cell sequences should be put together. Linkers, which are short chains of amino acids, are used to separate the different sequences. Although they serve different functions, the linkers KK, AAY, GPGPG, and EAAAK are all often utilized. The most frequent linkers utilized to connect CTL, HTL, and B cell epitopes are KK, AAY, and GPGPG (Martinelli, 2022). The EAAAK linker was used to join the CTL peptides to the protein while still allowing for sufficient space between them to allow for efficient target engagement. The selection of this linker was based on its ability to unify two distinct protein functions. AAY linker was used to divide each CTL epitope, and a GPGPG linker was used to divide each HTL epitope. Separating the individual B-cell epitopes using KK linkers allowed for their immunogenic potential to be maintained (Bhuiyan et al., 2021). To design multi-epitope vaccine, primarily all the CTL, HTL and B-cell epitopes that assessed was used for the very first constructed

vaccine. Further, different vaccine combination was prepared arranging the epitopes in different manners. Initially, 4 CTL epitopes (Table 7), 6 HTLs (Table 8) were selected based on their antigenicity and toxicity for further vaccine combinations.

Table 7: CTL epitopes selected for final vaccine combination.

CTL Epitope (NetCTL1.2)	Antigenicity (Vaxijen 2.0)	Toxicity (Toxin-pred)
LSALRTGWY	1.11 (Probable ANTIGEN)	Non toxin
TVSVGNTLY	0.8829(Probable ANTIGEN)	Non toxin
NIDIFNPKY	0.782 (Probable ANTIGEN)	Non toxin
VSVGNTLYY	1.05 (Probable ANTIGEN)	Non toxin

Table 8:HTL epitopes selected for final vaccine combination.

HTL Sequence	Antigenicity (Vaxijen 2.0)	Toxicity (Toxin-pred)
TVIEFQQKNNRLLEI	1.2807 (Probable Antigen)	Non toxic
IDKQLLPIVNKQSCS	0.6068 (Probable Antigen)	Non toxic
SVGNTLYYVNKQEGK	0.6938 (Probable Antigen)	Non toxic
RTGWYTSVITIELSN	0.7815 (Probable Antigen)	Non toxic
VGNTLYYVNKQEGKS	0.7002 (Probable Antigen)	Non toxic
VKGEPIINFYDPLVF	0.6139 (Probable Antigen)	Non toxic

All B-cell epitopes were employed in the initial vaccine combination. Only antigenic B-cells were incorporated in some combinations, however in order to make prediction, B-cell epitopes with fewer than 9mer were excluded. In the table 9, B-cell epitopes which are antigenic has been shown.

Table 9: B-cell epitopes selected for final vaccine combination.

No. (according to Table-6)	B-cell epitopes	Length	Antigenicity (Vaxijen 2.0)
5	QLLMQSTPATNNRARRELPRFMNY TLNNTKNTNVTLSKKRK	41	0.6080 (Probable ANTIGEN).
16	NIDIFNPKYDC	11	0.9724 (Probable ANTIGEN).
17	MTSKTDV	7	0.5145 (Probable ANTIGEN)
20	NEKINQSLAFIRK	13	0.5726 (Probable ANTIGEN).
22	STPVTLSKDQLSGINNIA	18	0.5611 (Probable ANTIGEN).

There were a total of seven proposed vaccine candidates, based on the varied ways in which the epitopes were arranged. General characteristics and methods of vaccine manufacturing are outlined below for clarity:

1) Vaccine candidate 1- constructed with all the antigenic, non-antigenic and non-toxic epitopes that were initially found. There were 12 CTL epitopes, 8 HTL epitopes and 22 B-cell epitopes.
2) Vaccine candidate 2- constructed with only antigenic epitopes that were screened by Vaxijen 2.0 and non-toxic epitopes that were screened by Toxinpred.
3) Vaccine candidate 3- created by replacing position of CTL and HTL epitopes and removing B-cell epitope that are less than 9mer and highly antigenic epitopes that are more than 1 according to Vaxijen 2.0.
4) Vaccine candidate 4- was created by altering the position of all antigenic CTL, HTL and B-Cell epitopes.
5) Vaccine candidate 5- generated by shifting the position of antigenic CTL, HTL, and B-Cell epitopes.
6) Vaccine candidate 6- was created by ordering the epitopes ascendingly in terms of antigenicity.

7) Vaccine candidate 7- The vaccine was developed by ranking the epitopes from most antigenic to least antigenic.

All the candidate vaccines with the corresponding sequences are tabulated in the Table 10.

Table 10: All the candidate vaccine with the corresponding sequences

No.	Vaccine construction criteria	Sequence of the constructed vaccine
1	Vaccine Combination 1 constructed with all epitopes	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSKGY LSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQELDKYK SAVTELQLLMQSTPATNNRARRELPRFMNYTLNNTKNTNV TLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVNKIKSAL LSTNKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIVNKQSCSIS NIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLTNSELL SLINDMPITNDQKKLMSNNVQIVRQQSYSIMSIIKEEVLAYV VQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTDRGWY CDNAGSVSFFPLAETCKVQSNRVFCDTMNSLTLPSEVNLCNI DIFNPKYDCKIMTSKTDVSSSVITSLGAIVSCYGKTKCTASNKD RGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQEGKSLYVK GEPIINFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDELLH NVNAGKSTTNIMITTIIIVIIVILLSLIAVGLLLYCKARSTPVTLSK DQLSGINNIAFSNEAAAKLSALRTGWYAAYNIDIFNPKYAAYK TFSNGCDYAAYTVSVGNTLYAAYVSVGNTLYYAAYKSDELLH NVAAYGVTTPVSTYAAYIIKEEVLAYAAYSSQNITEEFAAYSQ NITEEFYAAYGVTTPVSTYAAYSLGAIVSCYGPGPGLEGEVNK IKSALLSTYGPGPGTVIEFQQKNNRLLEIGPGPGELDKYKSAVTE LQLLGPGPGIDKQLLPIVNKQSCSGPGPGSVGNTLYYVNKQEG KGPGPGRTGWYTSVITIELSNGPGPGVGNTLYYVNKQEGKSGP GPGVKGEPIINFYDPLVFKKQNITEEFYQSKKIKENKCNGTDAK KVKKQELDKYKSKKQLLMQSTPATNNRARRELPRFMNYTLN NTKNTNVTLSKKRKKKDKQKKPKKVNKQKKCSISKKSVNAGV TTPVSTYKKNDMPITNDQKKLMSNNVQIVRQKKEKKDRGW KKVSFFPKKDTMNSLTLPSEKKNIDIFNPKYDCKKMTSKTDVKK RGIIKTFSNKKNKQEGKSLYVKGEPIINFYDPLVFPSDEFDAKK NEKINQSLAFIRKKELLHNVNAGKKSTPVTLSKDQLSGINNIA
2	Vaccine combination 2 constructed with only non-toxic and antigenic epitopes	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSKG YLSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQELDKY KSAVTELQLLMQSTPATNNRARRELPRFMNYTLNNTKNTN VTLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVNKIKS ALLSTNKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIVNKQSC SISNIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLTN SELLSLINDMPITNDQKKLMSNNVQIVRQQSYSIMSIIKEE VLAYVVQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTD RGWYCDNAGSVSFFPLAETCKVQSNRVFCDTMNSLTLPSE VNLCNIDIFNPKYDCKIMTSKTDVSSSVITSLGAIVSCYGKTK CTASNKDRGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQ EGKSLYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQSLAFI RKSDELLHNVNAGKSTTNIMITTIIIVIIVILLSLIAVGLLLYCKAR

		STPVTLSKDQLSGINNIAFSNEAAAKLSALRTGWYAAYNIDIF NPKYAAYTVSVGNTLYAAYVSVGNTLYYAAYGPGPGTVIEF QQKNNRLLEIGPGPGIDKQLLPIVNKQSCSGPGPGSVGNTL YYVNKQEGKGPGPGRTGWYTSVITIELSNGPGPGVGNTL YYVNKQEGKSGPGPGVKGEPIINFYDPLVFKKQLLMQSTPAT NNRARRELPRFMNYTLNNTKNTNVTLSKKRKKNIDIFNPKY DCKKMTSKTDVKKNEKINQSLAFIRKKSTPVTLSK DQLSGINNIA
No.	Vaccine construction criteria	Sequence of the constructed vaccine
3	Vaccine combination 3- created by replacing position of CTL and HTL epitopes and removing B-cell epitope that are less than 9mer and highly antigenic epitopes that are more than 1	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSKGY LSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQELDKYKS AVTELQLLMQSTPATNNRARRELPRFMNYTLNNTKNTNVT LSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVNKIKSALL STNKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIVNKQSCSISN IETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLTNSELLSLI NDMPITNDQKKLMSNNVQIVRQQSYSIMSIIKEEVLAYVVQL PLYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTDRGWYCDN AGSVSFFPLAETCKVQSNRVFCDTMNSLTLPSEVNLCNIDIFN PKYDCKIMTSKTDVSSSVITSLGAIVSCYGKTKCTASNKDRGIIK TFSNGCDYVSNKGVDTVSVGNTLYYVNKQEGKSLYVKGEPIIN FYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDELLHNVNAGK STTNIMITTIIIVIIVILLSLIAVGLLLYCKARSTPVTLSKDQLSGIN NIAFSNEAAAKTVSVGNTLYAAYNIDIFNPKYGPGPGVGNTL YYVNKQEGKSGPGPGSVGNTLYYVNKQEGKGPGPGIDKQL LPIVNKQSCSGPGPGRTGWYTSVITIELSNGPGPGVKGEPIIN FYDPLVFKKSTPVTLSKDQLSGINNIAKKNIDIFNPKYDCKKNE KINQSLAFIRKKQLLMQSTPATNNRARRELPRFMNYTLNN TKNTNVTLSKKRK
4	Vaccine Combination 4- was created by altering the position of all antigenic CTL, HTL and B-Cell epitopes.	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSK GYLSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQEL DKYKSAVTELQLLMQSTPATNNRARRELPRFMNYTLNN TKNTNVTLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEG EVNKIKSALLSTNKAVVSLSNGVSVLTSKVLDLKNYIDKQL LPIVNKQSCSISNIETVIEFQQKNNRLLEITREFSVNAGVTT PVSTYMLTNSELLSLINDMPITNDQKKLMSNNVQIVRQQ SYSIMSIIKEEVLAYVVQLPLYGVIDTPCWKLHTSPLCTTNT KEGSNICLTRTDRGWYCDNAGSVSFFPLAETCKVQSNRV FCDTMNSLTLPSEVNLCNIDIFNPKYDCKIMTSKTDVSSSV ITSLGAIVSCYGKTKCTASNKDRGIIKTFSNGCDYVSNKGV DTVSVGNTLYYVNKQEGKSLYVKGEPIINFYDPLVFPSDEF DASISQVNEKINQSLAFIRKSDELLHNVNAGKSTTNIMITT IIIVIIVILLSLIAVGLLLYCKARSTPVTLSKDQLSGINNIAFSN EAAAKLSALRTGWYAAYNIDIFNPKYAAYTVSVGNTLYAAY VSVGNTLYYGPGPGTVIEFQQKNNRLLEIGPGPGIDKQLLP IVNKQSCSGPGPGSVGNTLYYVNKQEGKGPGPGRTGWY TSVITIELSNGPGPGVGNTLYYVNKQEGKSGPGPGVKGEP IINFYDPLVFKKQLLMQSTPATNNRARRELPRFMNYTLNN TKNTNVTLSKKRKKNIDIFNPKYDCKKNEKINQSLAFIRKK STPVTLSKDQLSGINNIA

No.	Vaccine construction criteria	Sequence of the constructed vaccine
5	Vaccine Combination 5 generated by shifting the position of antigenic CTL, HTL, and B-Cell epitopes.	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSKG YLSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQELDKY KSAVTELQLLMQSTPATNNRARRELPRFMNYTLNNTKNTN VTLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVNKIKSA LLSTNKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIVNKQSCS ISNIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLTNSE LLSLINDMPITNDQKKLMSNNVQIVRQQSYSIMSIIKEEVLA YVVQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTDRG WYCDNAGSVSFFPLAETCKVQSNRVFCDTMNSLTLPSEVN LCNIDIFNPKYDCKIMTSKTDVSSSVITSLGAIVSCYGKTKCTA SNKDRGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQEGK SLYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQSLAFIRKS DELLHNVNAGKSTTNIMITTIIIVIIVILLSLIAVGLLLYCKARST PVTLSKDQLSGINNIAFSNEAAAKNIDIFNPKYAAYVSVGNT LYYAAYLSALRTGWYAAYTVSVGNTLYGPGPGRTGWYTSV ITIELSNGPGPGVGNTLYYVNKQEGKSGPGPGSVGNTLYYV NKQEGKGPGPGTVIEFQQKNNRLLEIGPGPGIDKQLLPIVNK QSCSGPGPGVKGEPIINFYDPLVFKKQLLMQSTPATNNRARRE LPRFMNYTLNNTKNTNVTLSKKRKKKSTPVTLSKDQLSGINN IAKKNEKINQSLAFIRKKKNIDIFNPKYDC
6	Vaccine Combination 6- was created by ordering the epitopes ascendingly in terms of antigenicity.	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSAVSK GYLSALRTGWYTSVITIELSNIKENKCNGTDAKVKLIKQELD KYKSAVTELQLLMQSTPATNNRARRELPRFMNYTLNNTK NTNVTLSKKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEV NKIKSALLSTNKAVVSLSNGVSVLTSKVLDLKNYIDKQLLPIV NKQSCSISNIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTY MLTNSELLSLINDMPITNDQKKLMSNNVQIVRQQSYSIMSI IKEEVLAYVVQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNICL TRTDRGWYCDNAGSVSFFPLAETCKVQSNRVFCDTMNSL TLPSEVNLCNIDIFNPKYDCKIMTSKTDVSSSVITSLGAIVSCY GKTKCTASNKDRGIIKTFSNGCDYVSNKGVDTVSVGNTLYY VNKQEGKSLYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQ SLAFIRKSDELLHNVNAGKSTTNIMITTIIIVIIVILLSLIAVGLLL YCKARSTPVTLSKDQLSGINNIAFSNEAAAKLSALRTGWYAAY VSVGNTLYYAAYTVSVGNTLYAAYNIDIFNPKYGPGPGTVIEF QQKNNRLLEIGPGPGRTGWYTSVITIELSNGPGPGVGNTL YYVNKQEGKSGPGPGSVGNTLYYVNKQEGKGPGPGVKGEP IINFYDPLVFGPGPGIDKQLLPIVNKQSCSKKNIDIFNPKYDCKK QLLMQSTPATNNRARRELPRFMNYTLNNTKNTNVTLSKKRK KKNEKINQSLAFIRKKSTPVTLSKDQLSGINNIA

No.	Vaccine construction criteria	Sequence of the constructed vaccine
7	Vaccine Combination 7- was developed by ranking the epitopes from most antigenic to least antigenic	MELPILKTNAITAILAAVTLCFASSQNITEEFYQSTCSA VSKGYLSALRTGWYTSVITIELSNIKENKCNGTDAKVK LIKQELDKYKSAVTELQLLMQSTPATNNRARRELPRF MNYTLNNTKNTNVTLSKKRKRRFLGFLLGVGSAIASG IAVSKVLHLEGEVNKIKSALLSTNKAVVSLSNGVSVLT SKVLDLKNYIDKQLLPIVNKQSCSISNIETVIEFQQKNN RLLEITREFSVNAGVTTPVSTYMLTNSELLSLINDMPIT NDQKKLMSNNVQIVRQQSYSIMSIIKEEVLAYVVQLP LYGVIDTPCWKLHTSPLCTTNTKEGSNICLTRTDRGW YCDNAGSVSFFPLAETCKVQSNRVFCDTMNSLTLPSE VNLCNIDIFNPKYDCKIMTSKTDVSSSVITSLGAIVSCY GKTKCTASNKDRGIIKTFSNGCDYVSNKGVDTVSVG NTLYYVNKQEGKSLYVKGEPIINFYDPLVFPSDEFDAS ISQVNEKINQSLAFIRKSDELLHNVNAGKSTTNIMITTIII VIIVILLSLIAVGLLLYCKARSTPVTLSKDQLSGINNIAFS NEAAAKNIDIFNPKYAAYTVSVGNTLYAAYVSVGNTL YYAAYLSALRTGWYGPGPGIDKQLLPIVNKQSCSGPG PGVKGEPIINFYDPLVFGPGPGSVGNTLYYVNKQEGK GPGPGVGNTLYYVNKQEGKSGPGPGRTGWYTSVITI ELSNGPGPGTVIEFQQKNNRLLEIKKSTPVTLSKDQLS GINNIAKKNEKINQSLAFIRKKKQLLMQSTPATNNRA RRELPRFMNYTLNNTKNTNVTLSKKRKKKNIDIFNPKY DC

3.10 In-silico Biochemical Analysis of primary protein sequence and Candidate Vaccines

To analyze the biochemical properties of all the constructed vaccines ProtParam tool was used. In the table 11 a brief comparison of their biochemical test is shown.

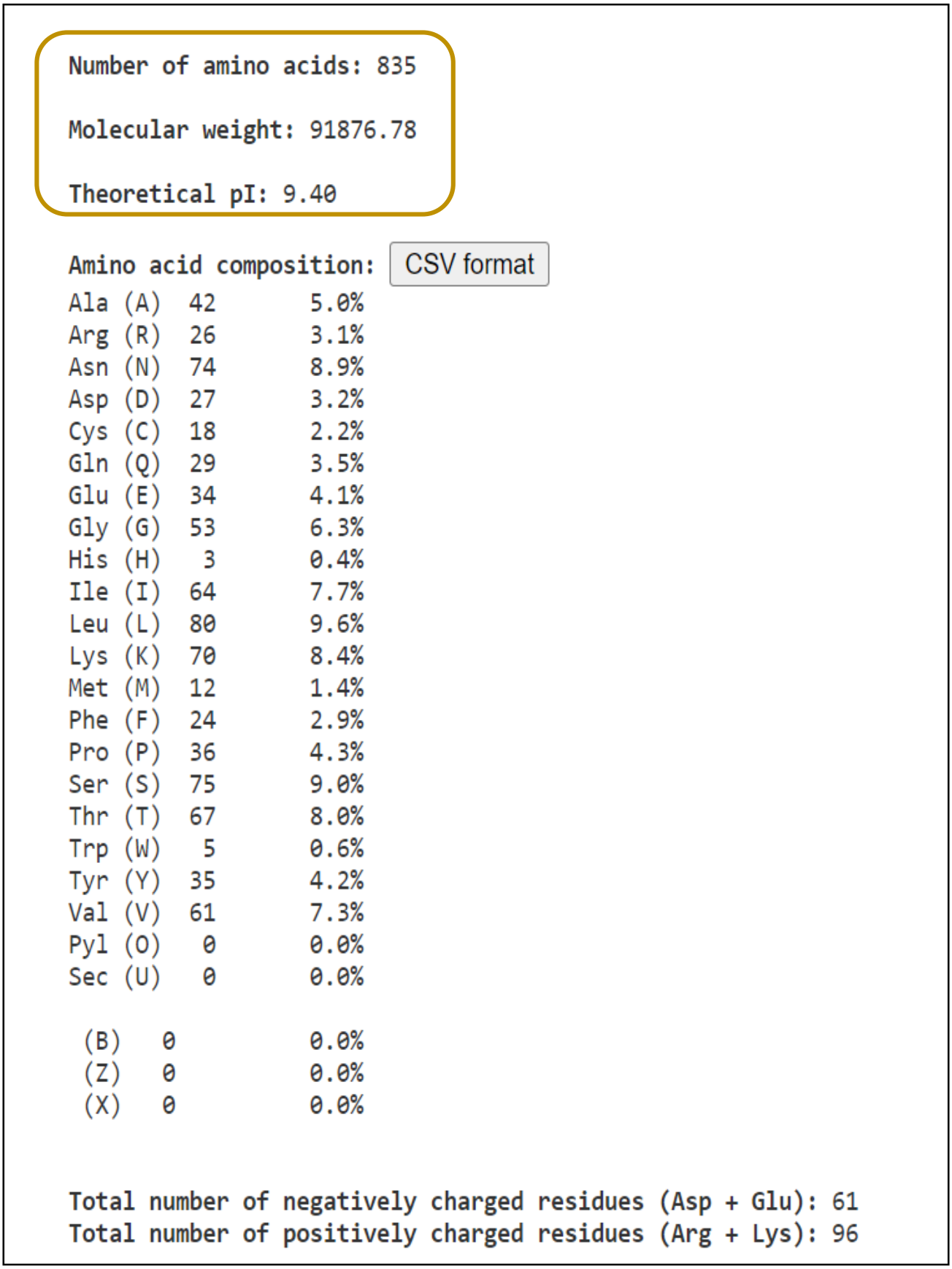
Table 11: In-depth biochemical evaluation of primary protein and all 7 vaccine candidates.

Biochemical parameters	Primary Protein	Vaccine candidate 1	Vaccine candidate 2	Vaccine candidate 3	Vaccine candidate 4	Vaccine candidate 5	Vaccine candidate 6	Vaccine candidate 7
Stability, GRAVY, Molecular Weight(M.W) (Protparam)	Instability index-40.42(unstable), GRAVY: -0.007, M.W- 63334.16 dalton	Instability index-40.68 (unstable), GRAVY:-0.348 M.W-128599.86 dalton	Instability index-39.18 (stable), GRAVY:-0.188 M.W- 93201.33 dalton	Instability index-39.16 (stable), GRAVY:-0.189 M.W- 87028.28 dalton	Instability index-38.99 (stable), GRAVY:-0.180 M.W- 91876.78 dalton	Instability index-39.09 (stable), GRAVY:-0.180 M.W- 91876.78 dalton	Instability index-39.17 (stable), GRAVY:-0.180 M.W- 91876.78 dalton	Instability index-39.17 (stable), GRAVY:-0.180 M.W- 91876.78 dalton
Antigenicity (vaxijen 2.0)	0.5173 (Probable ANTIGEN).	0.5073(Probable ANTIGEN).	0.5732 (Probable ANTIGEN).	0.5578 (Probable ANTIGEN).	= 0.5810 (Probable ANTIGEN).	= 0.5808 (Probable ANTIGEN).	0.5677 (Probable ANTIGEN).	0.5713 (Probable ANTIGEN).
Allergenicity (Allergenonline, AllerTop v2.0)	Non- Allergen	Non- Allergen	Non- Allergen	Non- Allergen	Non- Allergen	Non- Allergen	Non- Allergen	Non- Allergen
Ramachandran favored region (SWISS-MODEL)	94.95%	94.95%	94.95%	94.95%	94.95%	94.95%	94.95%	94.95%
Z-score (Prosa-Web)	-9.89	-9.89	-9.89	-9.89	-9.89	-9.89	-9.89	-9.89
Phyre2	Confidence-100% Coverage- 80%	Confidence-100% Coverage- 54%	Confidence-100% Coverage- 54%	Confidence-100% Coverage- 58%	Confidence-100% Coverage- 55%	Confidence-100% Coverage- 55%	Confidence-100% Coverage- 55%	Confidence-100% Coverage- 55%

A total of seven predicted vaccine candidates fared the best. All of the predicted vaccines were subjected to biochemical analysis in order to propose final vaccine candidates. Each showed different characteristics in their biochemical analysis.

From the comparison of these 7 possible vaccine candidate no. 4 was chosen over other 6 as the final proposed vaccine. It was selected due to high antigenicity than other candidates. According to vaxijen 2.0 server vaccine 4's antigenicity is 0.5810 which is highest among other vaccines. Additionally, other biochemical parameters are near to others and acceptable.

Through the use of the ProtParam server checking is performed to determine biochemical composition. Vaccine 4 has 835 amino acids, molecular weight of 91876.78 dalton (or 91.876 kda), theoretical pI (isoelectric point) of 9.40, a total of 61 negatively charged residues, and a total of 96 positively charged residues, with alanine making up roughly 5% of the amino acid composition. According to the stability index, vaccine candidate 4 is stable. Based on the results of protparam, this vaccine sequence is stable since its instability index is below 40, coming in at 38.99. The Grand average of hydropathicity index (GRAVY) of the vaccine 4 is -0.180. Hydrophobicity in proteins is represented by the GRAVY index, which is determined by summing the hydropathy values of each residue and dividing by the total number of residues in the sequence. Hydrophobic proteins tend to have GRAVY scores higher than 0 (Magdeldin et al., 2012). The corresponding chemal formula for the vaccine is $C_{4087}H_{6601}N_{1097}O_{1238}S_{30}$ and total number of atoms 13053. All the information of vaccine canditate 4 has been shown in figure 11 (a) and (b).



Number of amino acids: 835

Molecular weight: 91876.78

Theoretical pI: 9.40

Amino acid composition: CSV format

Ala (A)	42	5.0%
Arg (R)	26	3.1%
Asn (N)	74	8.9%
Asp (D)	27	3.2%
Cys (C)	18	2.2%
Gln (Q)	29	3.5%
Glu (E)	34	4.1%
Gly (G)	53	6.3%
His (H)	3	0.4%
Ile (I)	64	7.7%
Leu (L)	80	9.6%
Lys (K)	70	8.4%
Met (M)	12	1.4%
Phe (F)	24	2.9%
Pro (P)	36	4.3%
Ser (S)	75	9.0%
Thr (T)	67	8.0%
Trp (W)	5	0.6%
Tyr (Y)	35	4.2%
Val (V)	61	7.3%
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): 61
Total number of positively charged residues (Arg + Lys): 96

Figure 11(a): Biochemical analysis report of protparam (Gasteiger et al., 2003).

Atomic composition:

Carbon	C	4087
Hydrogen	H	6601
Nitrogen	N	1097
Oxygen	O	1238
Sulfur	S	30

Formula: $C_{4087}H_{6601}N_{1097}O_{1238}S_{30}$
Total number of atoms: 13053

Extinction coefficients:

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

Ext. coefficient 80775
Abs 0.1% (=1 g/l) 0.879, assuming all pairs of Cys residues form cystines

Ext. coefficient 79650
Abs 0.1% (=1 g/l) 0.867, assuming all Cys residues are reduced

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).

Instability index:

The instability index (II) is computed to be 38.99
This classifies the protein as stable.

Aliphatic index: 93.47

Grand average of hydropathicity (GRAVY): -0.180

Figure 11 (b): Biochemical analysis report of protparam (Gasteiger et al., 2003).

3.11 Allergenicity and Toxicity Prediction of Constructed Vaccine

At the outset, the AllerTOP v.2.0 server was used to evaluate potential allergenicity. Using this approach, we can determine if the entire vaccine sequence is allergenic. Additionally, the AllergenOnline server was utilized to make antigenicity predictions. Allergen entities are scanned for in the sequences submitted to this server (Goodman et al., 2016). Either from the server's perspective, the allergenicity or allergen entity sequence was absent. Result from the AllerTOP server is shown in figure 12.

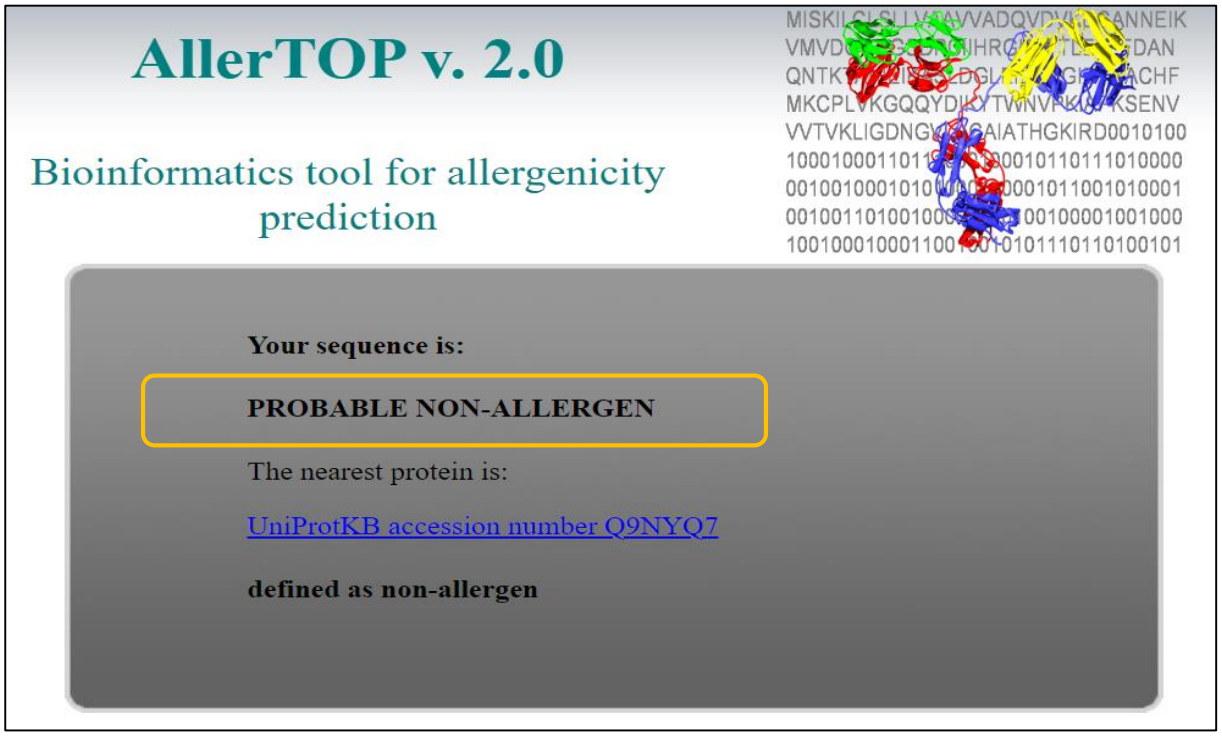


Figure 12: AllerTOP result for vaccine candidate 4 (Dimitrov et al., 2014).

T3DB (Toxin and Toxin Target Database) was approached to assess whether or not the designed vaccines contained any toxins. When it comes to bioinformatics tools, T3DB is one of a kind because it contains information on both toxic substances and the organisms that are susceptible to them (Wishart et al., 2015). The T3DB server did not return any results when the vaccination sequence was inputted into it. This suggested that the vaccine would not be toxic in the manner that was predicted. Figure 13 displays the output from the T3DB server.

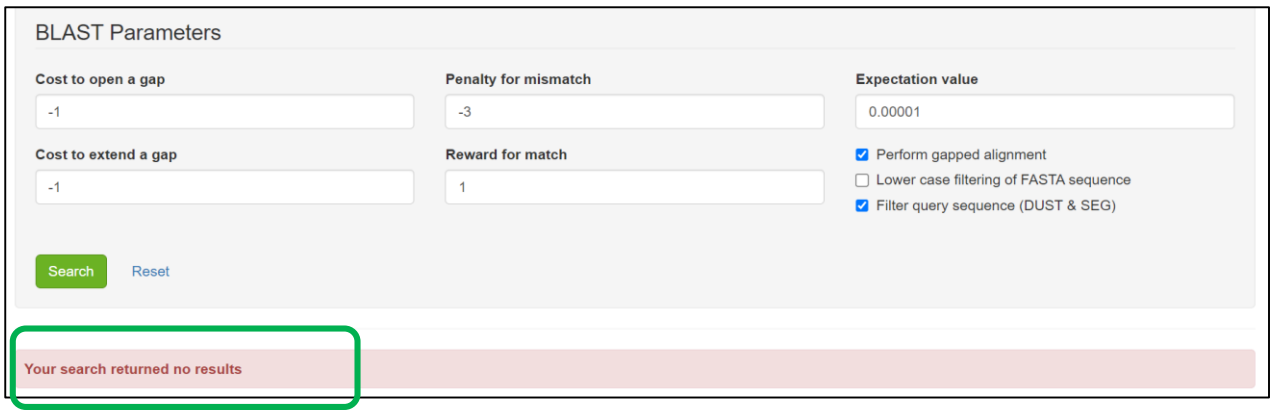


Figure 13: T3DB server's result for vaccine candidate 4 (Wishart et al., 2015).

3.12 Homology modeling

In order to acquire the homology modeling, the Phyre2 server was used. The Phyre2 framework is the result of the collaboration of numerous individual pieces of software. The ability to predict the three-dimensional structure of a single protein sequence is the service's main draw. This server displays multiple 3D models for a single sequence, along with information about the alignment's coverage, confidence, and the quality of the models themselves. The sequence of the vaccine inputted in phyre2 server and later a e-mail with restlt pdb file will be sent. Discovery studio software was used to open it. According to this server sequence of the vaccine candidate 4 generated 3D model with 100% confidence and 55% coverage. Inpretation is 459 residues of the vaccine have been modelled which is 55% of the total residues with 100 confidence.This percentage determine the accuracy of the model. This figure should be greater than 30–40% for a model with a high degree of accuracy. However, it is essential to realize that models can still be very useful at low sequence identities (15%) as long as confidence is high (Kelley et al., 2016). In the research article "In-silico analysis of ectodomain G protein of Respiratory Syncytial Virus" researchers recevied an coverage from phyre2 server around 43% (Shafat et al., 2017). These evidences suggests that vaccine candidate 4 has a relatively accurate model. In the figure 14 confidence and coverage information and in the figure 15 homology modeling of vaccine 4 is shown.

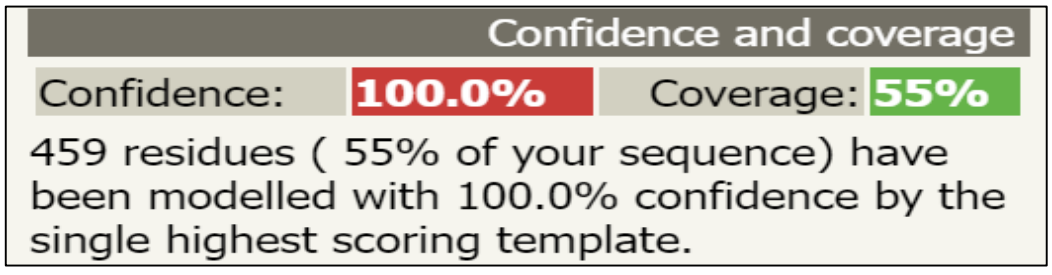


Figure 14: Confidence and coverage of vaccine 4 from Phyre2 (Kelley et al., 2016).

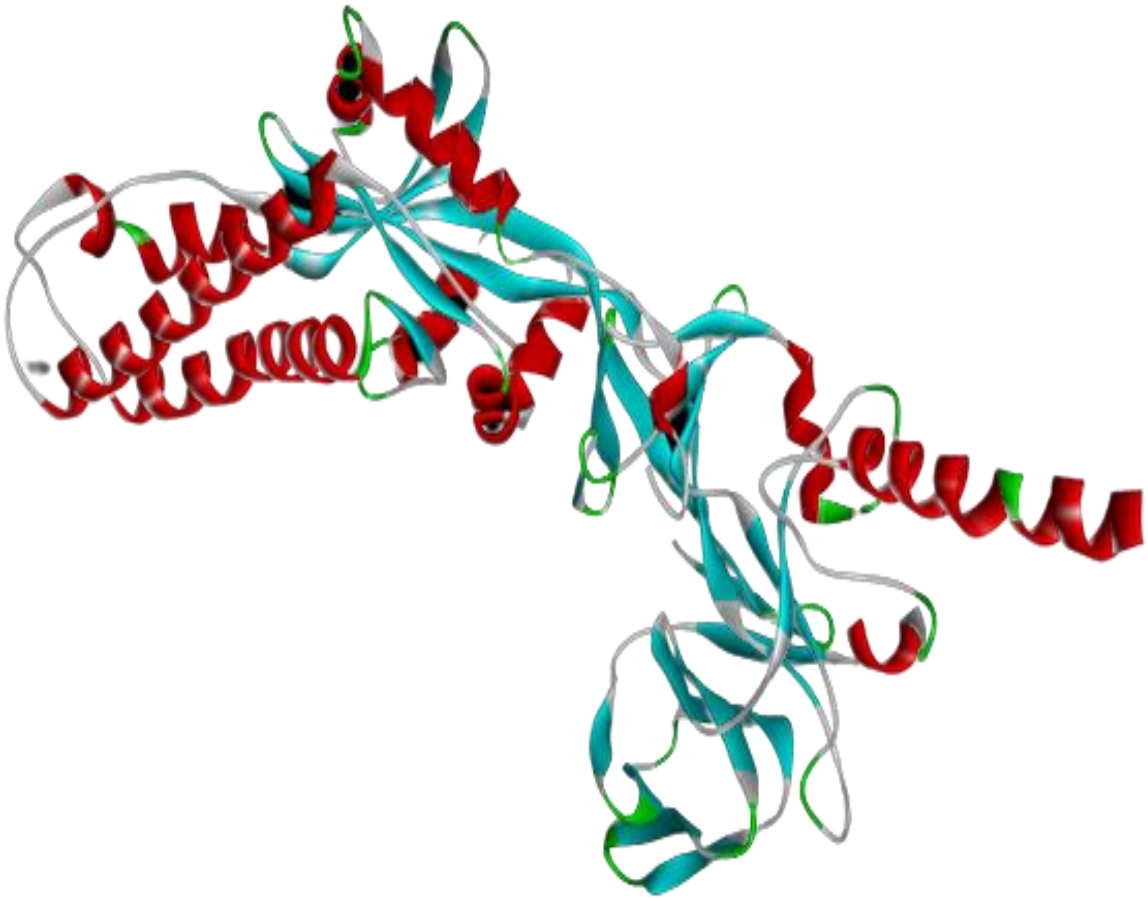
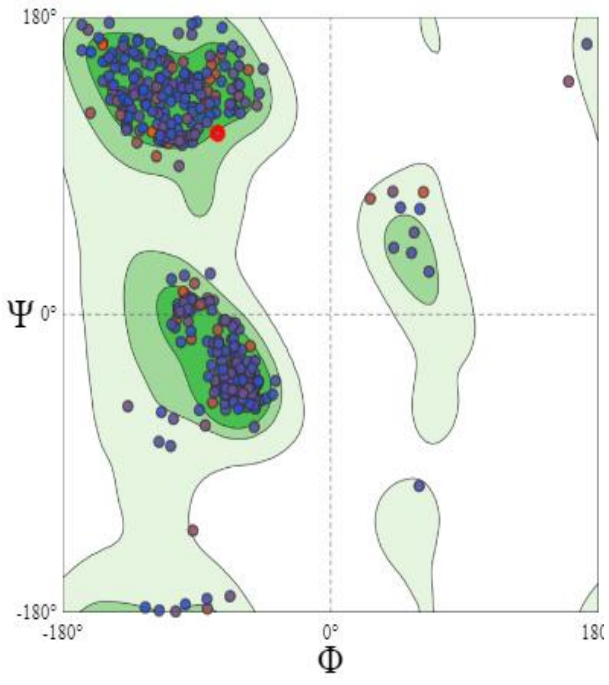


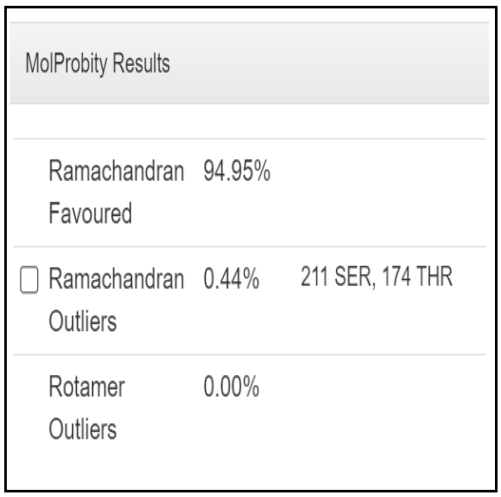
Figure 15: Homology modeling of vaccine candidate 4 from Phyre 2 server (Kelley et al., 2016).

3.13 Ramachandran Plots and Z-Score Evaluation

The Ramachandran plot is the most frequently used method for validating a homology model. The plot shows the empirical distribution of data points observed within a single structure. By utilizing SWISS-MODEL server Ramachandran plot was generated. A Ramachandran plot diagram, the outcomes of MolProbity, the estimates of Quality, and the measurements of Residue Quality are all included on the results page. Findings from MolProbity suggest that, Ramachandran Favoured region 94.95%, Ramachandran Outliers 0.44%, Rotamer Outliers 0.00%. Amino acids located in the Ramachandran outlier represent unfavorable regions of the plot. The results from this server suggest that the homology model would therefore likely be impactful. In the figure 16(a) and 16(b) Ramachandran plot and MolProbity Results are shown.



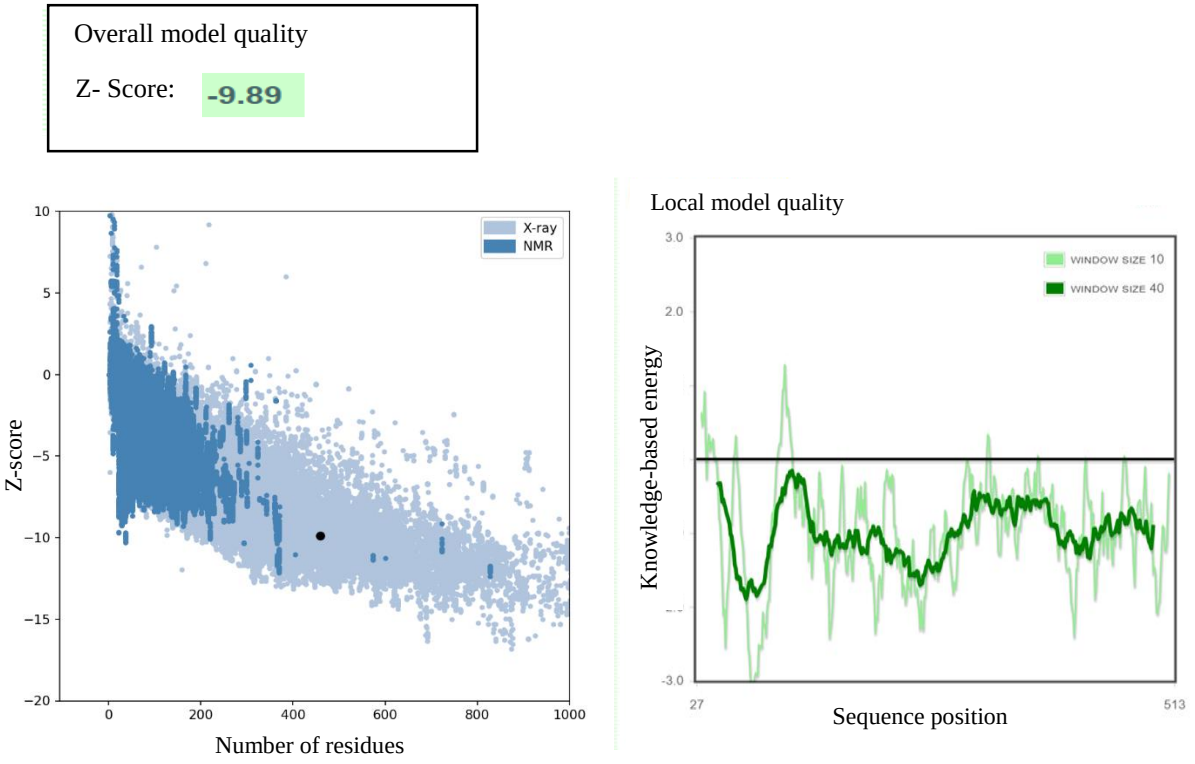
16(a)



16(b)

Figure 16 (a): Ramachandran plots of vaccine 4 from SWISS-MODEL (Sheik et al., 2002) and 16 (b): MolProbity results in SWISS-MODEL (Sheik et al., 2002).

Z score value was determined from the ProSA-web. The z-score designates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. After uploading the PDB file to the ProSA website, the Z-score value and two accompanying graphs were displayed on the results page. Number of residues vs. z-score was the first graph (Overall model quality), and sequence position vs. knowledge-based energy (Local model quality) was the second. According to the server z-score is -9.89. A z-score of −9.89 was obtained, which showed that the modeled protein shows a better protein model. A more negative z-score implies a better protein model (Kwofie et al., 2018). The local model quality of the protein was also generated. Amino acids residues with more negative energy levels have a high tendency of contributing to the overall quality (Kwofie et al., 2018). In figure 17(a) and 17(b) overall model quality graph along with the z-score and Local model quality graph is shown respectively.



17 (a) *17 (b)*

Figure 17 (a): Z-score vs. number of residues graph and Z-score in ProSAweb server (Wiederstein & Sippl, 2007) and 17 (b): knowledge-based energy with sequence position graph (Wiederstein & Sippl, 2007).

3.14 Molecular Docking of Final Vaccine with Relevant Human Receptor

The molecular docking process was conducted on the PATCHDOCK server. How precisely the link between the vaccine and the human toll-like receptor (TLRs) form was predicted here. Recognizing pathogens in the extracellular matrix is an important part of the immune response, and TLRs, which are members of the pattern recognition receptors (PRRs) family and have been conserved throughout evolution, play a key role in this process. Toll-like receptors (TLRs) are critical for the proper regulation of inflammatory responses and the activation of innate or adaptive immune responses required for the efficient removal of infectious pathogens from the body. Moreover, Because of their ability to recognize conserved pathogen-associated molecular patterns (PAMPs), conserved structures of pathogens, or damage caused by pathogens within the host, TLRs play an important role in the first line of defense against pathogens (Sameer & Nissar, 2021). The Toll-like receptors (TLRs) are a family of type I transmembrane proteins. Human TLRs have been categorized into ten different groups (TLR1-TLR10). Some TLRs (TLR1, TLR2, TLR4, TLR5, and TLR6) are found on the cell surface, while others (TLR3, TLR7, TLR8, and TLR9) are found inside the cell. Where a particular

TLR is found depends on where the ligand it recognizes was first encountered. Cell surface TLRs are primarily responsible for detecting extracellular bacterial products, whereas endosomal TLRs detect viral and bacterial nucleic acids (Dowling & Mansell, 2016).

Human toll-like receptor or TLR-8 (3W3G) was entered as the receptor molecule in the PATCHDOCK server. As the ligand molecule, the final vaccine was entered. For the receptor-ligand complex, the server provided several possible outcomes.

Solution No: Number of the solution

Score: Geometric shape complementarity score. The solutions are sorted according to this score.

ACE: Atomic contact energy.

Area: Approximate interface area of the complex.

Transformation: 3D transformation: 3 rotational angles and 3 translational parameters. The transformation should be applied on the ligand molecule.

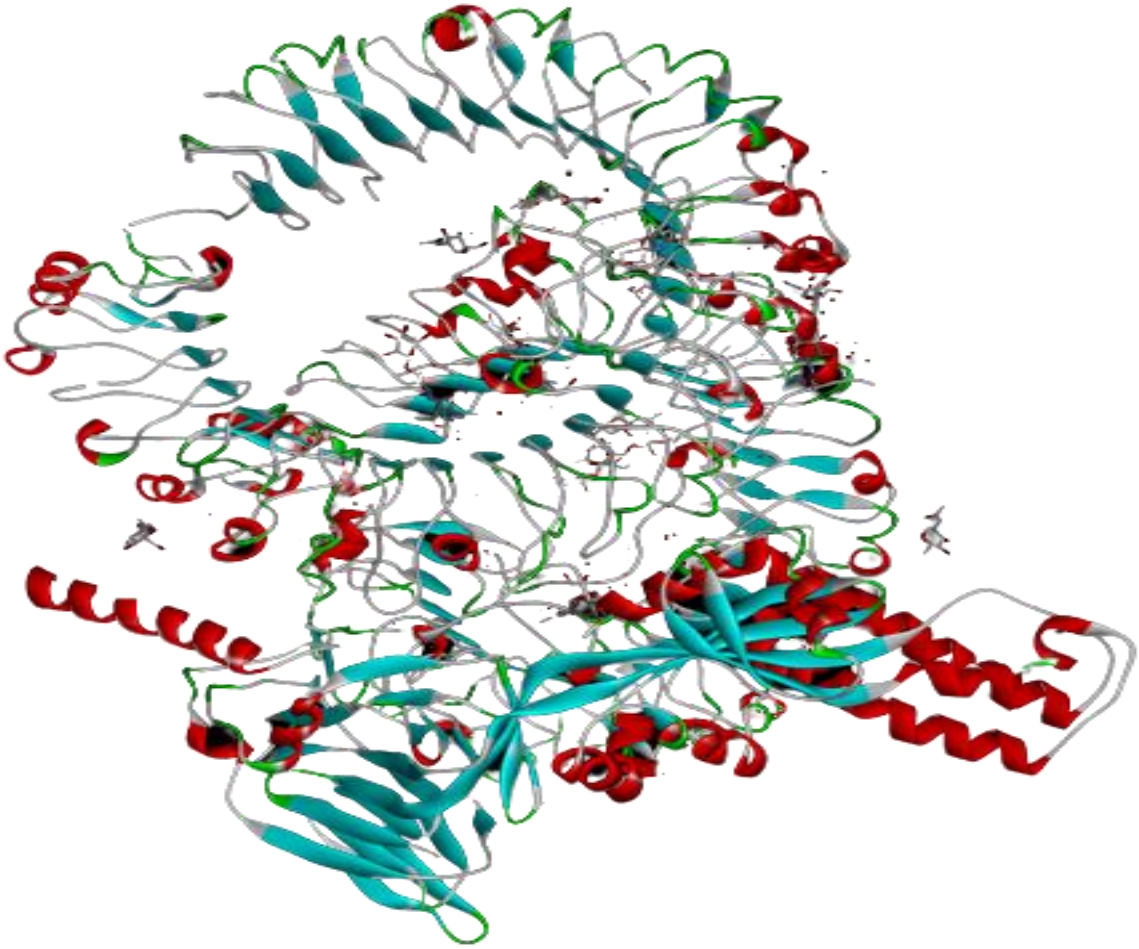
PDB file of the complex: The predicted complex structure in PDB format (Schneidman-Duhovny et al., 2005).

In terms of output scores, the best complex (solution 1) had an ACE of -321.99 , an area of 4659.80, and a score of 21240. The transformation of the vaccine 4 was -2.03 -0.95 -1.64 -1.84 50.73 -18.76. Patchdock uses the molecular surface representation to generate flat, concave, or convex patches. To generate potential transformations, the server finds pairs of patches that work well together. Subsequently, root mean square deviation (RMSD) is used to filter out superfluous solutions (Schneidman-Duhovny et al., 2005). In the figure 18(a) Highest scoring solution in PATCHDOCK server and 18(b) Molecular docking of vaccine and receptor in PATCHDOCK server is shown.

Receptor	Ligand	Complex Type	Clustering RMSD	User e-mail
3w3g	final.casp1.pdb	Default	4.0	sejanalsaba@gmail.com

Solution No	Score	Area	ACE	Transformation
1	21240	4659.80	-321.99	-2.03 -0.95 -1.64 -1.84 50.73 -18.76

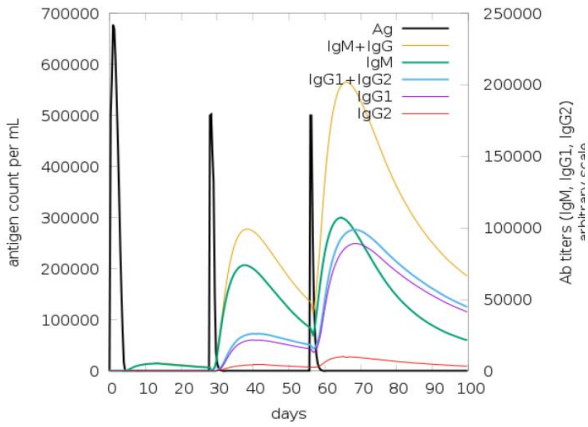
Figure 18(a): Highest scoring solution in PATCHDOCK server (Schneidman-Duhovny et al., 2005)



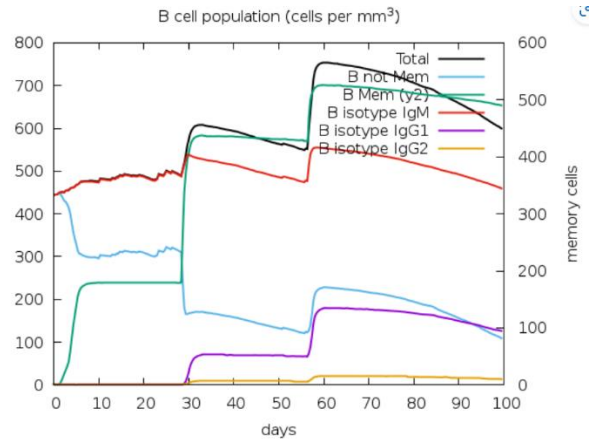
18 (b): Molecular docking of vaccine and receptor in PATCHDOCK server (Schneidman-Duhovny et al., 2005).

3.15 Immune Simulations of the vaccine candidate

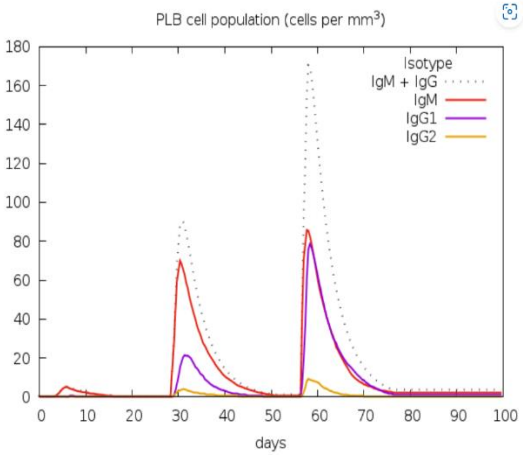
The C-IMMSIM server simulated the vaccine's immunogenicity and immune response in-silico. Using c-immsim, the suggested vaccination's immune system effects and immunological response may be predicted. The body's immunological reaction must be assessed by monitoring antibody or immunoglobulin levels in response to vaccination timing and dosage. It's also essential to monitor how immune cells increase and diminish with vaccine dosages. Figure 19(a) to 19(j) demonstrates how the C-IMMSIM server predicted and graphically showed these characteristics in response to the vaccine.



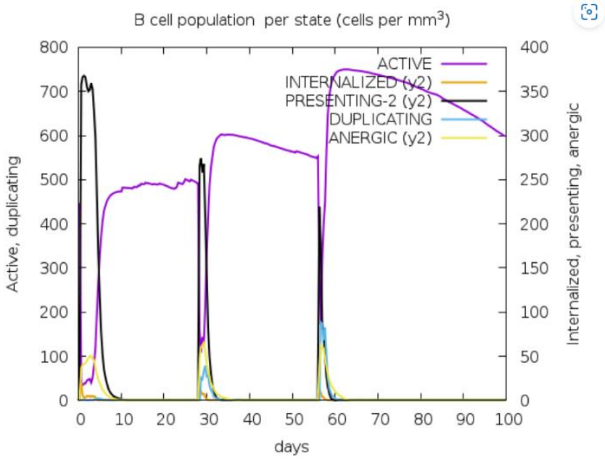
19 (a) Antigen count per mL and antibody titers (Rapin et al., 2010).



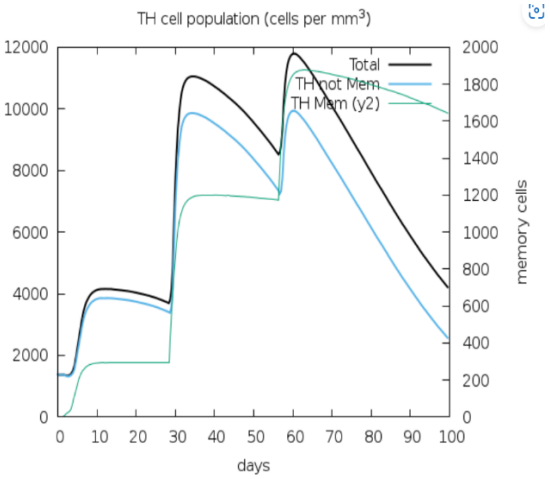
19 (b) Total count of B lymphocytes and memory cells (Rapin et al., 2010).



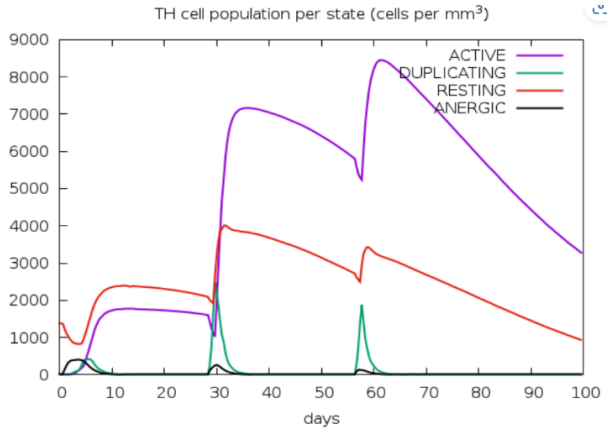
19 (c) Number of plasma B cells according to their isotypes (Rapin et al., 2010).



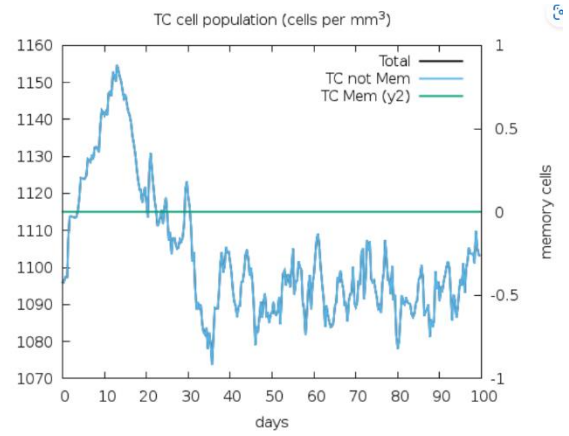
19 (d) Entity-state of B cell population (Rapin et al., 2010)



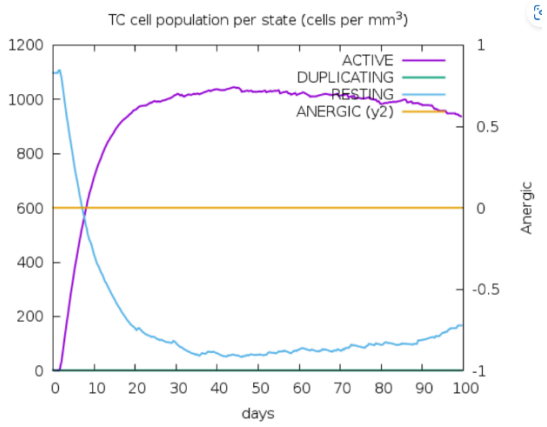
19 (e) Number of CD4 T helper lymphocytes (Rapin et al., 2010).



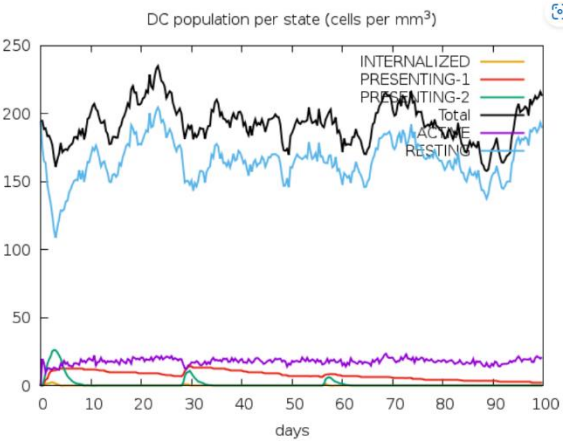
19 (f) Entity state of CD4 Helper T cells (Rapin et al., 2010).



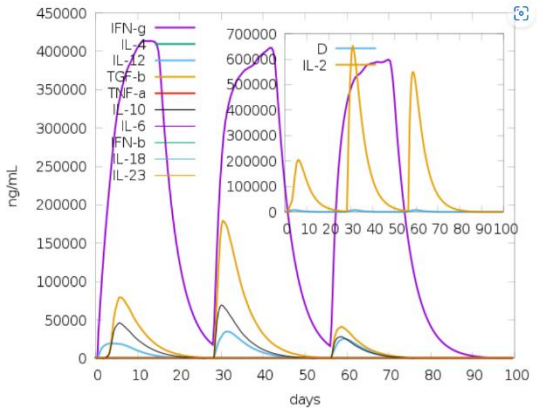
19 (g) Number of CD8 T cytotoxic (TC) cells (Rapin et al., 2010).



19 (h) Number of CD8 T cytotoxic (TC) cells per entity (Rapin et al., 2010).



19 (i) Total number of dendritic cells and their states (Rapin et al., 2010).



19 (j) Total number of dendritic cells and their states (Rapin et al., 2010).

Figure 19 (a-j): Immune simulations of vaccine via C-IMMSIM (Rapin et al., 2010).

After administering the proposed vaccine in the immune system antigen count per ml can be observed through figure 19(a). Even though IgM and IgG antibody levels increase gradually over the course of the first 28 days following vaccination, they increase dramatically after around 60 days as a result of the repeated delivery and completion of the vaccine doses. Antibody formation suggested that the vaccine had the intended effect on the immune system. In the figure 19(b), B-cell population (cell per mm^3) is shown. It is important to keep memory for the invaded antigen for further antibody production in the body for future infection. In addition to developing antibodies, B lymphocytes are responsible for the development of

memory cells in the body. The body's ability to quickly recognize the same disease later on and produce a protective immune response is greatly enhanced by the memory cells generated from B cells. Successful vaccination relies significantly on the generation of functional memory cells. After vaccination, the total number of B lymphocytes and memory cells were plotted in this figure.

In figure 19(c), the plasma b lymphocyte count is plotted against the day after vaccination, alongside the isotypes IgM + IgG, IgG1, IgG2, and IgM. This information is presented in conjunction with the figure. Additionally, in the figure 19(d) is a representation of the B cells by entity state using a diagram. It showed the total number of B cells, the proportion of B cells that incorporated the antigen, and the number of B cells that were duplicating.

In the figure 19(e) graph TH or helper T-cell population rise is shown against the day of vaccination. A subset of immune cells known as helper T-cells. The thymus gland is responsible for producing them along with a few other cell types. Whenever an infection occurs, the body's Helper T-cells will recognize. They signal to other immune cells to participate in the fight against the pathogen. Involved in the body's adaptive immune response, these T-cells are crucial.

Figure 19(f) depicts the duplicating, resting, and expected active or anergic states of the helper T cell. Figure 19(g) illustrated cytotoxic T cell population along with total and memory cell against the days of vaccine implementation. Figure 19(h) illustrated The number of cytotoxic T cells predicted to be active, to undergoing duplication, be in anergic state or be in resting state. In the figure 19(i) dendritic cell or DC population per state is shown. DC is specialized immune cell that helps stimulate immunological responses by displaying antigens on its surface to other cells of the immune system. Dendritic cells are phagocytes that also function as antigen-presenting cells (APC). In this figure DC cells presenting on MHC class I molecules are marked by the notation PRESENTING-1, while DC cells presenting on MHC class II molecules are denoted by the notation PRESENTING-2. In addition, total, active and resting state also be shown in the graph. Lastly, the concentration of cytokines and interleukins is depicted in figure 19(j). Danger signal and leukocyte growth factor IL-2 are depicted in the inset scatter graph.

Chapter 4
Discussion

Lower respiratory tract infection (LRTI) is a common medical problem, and respiratory syncytial virus (RSV) is a leading cause. Several proteins make up this virus. An essential protein in the virus's pathogenesis must be targeted in order to develop an effective vaccine. Fusion glycoprotein (F0) was focused in this study in an effort to develop an RSV vaccine. Further steps, such as CTL, HTL, and B-cell epitope prediction, are required after the primary protein for the vaccine has been selected. That particular protein need to have required antigenicity for further vaccine design proceedings. Here, threshold for the antigenicity was settled in 0.5 in the Vaxijen 2.0 server and it turned out with 0.5173 (Figure 5), which indicates this protein is probable antigen.

Cytotoxic T lymphocyte identification was next proceedings. In most instances, CTL are the first line of defense against viral infections. It is the presentation of viral peptides made inside an infected cell in the binding groove of class I MHC that allows CTL to identify the virus (Keogan et al., 2021). Adding CTL epitopes to the vaccine is a promising strategy for eliciting effective immunity to RSV. After vaccination, CTLs can also aid in the formation of immunological memory (N. Kumar et al., 2021). In this study, initially 18 CTL epitopes were found via NetCTT-1.2 server (Figure 6). These CTLs were further screened for toxicity, antigenicity, allergenicity and corresponding MHC class 1 alleles. Via NetMHCpan-4.1 major histocompatibility complex or MHC class 1 alleles were predicted. The results showed that alleles existed for 11 of the 18 CTL epitopes. The MHC class I protein is essential for the adaptive immune system to function properly. It is the role of many proteins in this family to display peptides on the cell surface, where they can be recognized by T cells (Wieczorek et al., 2017). Consequently, this research involved making predictions about the binding of MHC class I alleles.

Moreover, toxicity was checked by Toxinpred server and all the CTL epitopes were non-toxic. Additionally, antigenicity was checked by Vaxijen 2.0 server and four out of 11 epitopes were came out as probable antigen. In addition, the AllerTOP v.2.0 server classified four of the four epitopes as potential allergens. Although the final vaccine may be non-allergenic, the possibility that it contains CTL epitopes from allergens has not been ruled out. After the CTL

epitopes were identified, the NetMHCIIpan4.0 server predicted the HTL epitopes. Subsequently, 223 HTL epitopes with strong binding were identified for the F0 protein of RSV.

In order to design a feasible subunit vaccine, it is necessary to identify efficient HTL epitopes for the chosen protein sequence. Almost every type of adaptive immune response relies on the presence of helper T cells, making them among the most pivotal immune cells. They aid in the activation of cytotoxic T cells, which kill infected target cells, as well as B cells, which secrete antibodies and macrophages, which destroy ingested microbes (Dhanda, Vir, et al., 2013).

When a helper T cell is stimulated by an antigen-presenting cell, it can develop into either a TH1 or TH2 effector helper cell. Cytokines secreted by these two subclasses of effector helper T cells are characteristic of their respective functions. Interferon-γ (IFN-γ) and tumor necrosis factor-alpha (TNF-a) are secreted by TH1 cells, which then activate macrophages to kill microbes within the macrophages' phagosomes if the cell undergoes differentiation into a TH1 cell. Cytotoxic T cells, which target infectious cells, will also be activated. Additionally, If, on the other hand, the helper T cell differentiates into a TH2 cell, it will primarily protect the animal from extracellular pathogens by secreting interleukins 4, 5, 10, and 13 (IL-4, IL-5, IL-10, and IL-13). IgE and subclasses of IgG antibodies that bind to mast cells, basophils, and eosinophils can all be stimulated to be produced by a TH2 cell (Alberts B, Johnson A, Lewis J, et al., 2002). Using the IL4pred, IL10pred, and IFNepitope servers, the 223 HTL epitopes obtained in this study were screened to predict which HTLs induce the cytokines IL-4, IL-10, and IFN-γ simultaneously. Eight such HTL epitopes were discovered through screening (Table 5). After that, these epitopes were screened for antigenicity and toxicity. Only six such epitopes were found which were probable antigen and went for further vaccine design proceeding.

Afterwards, B-cell epitopes were predicted for the vaccine design. B cell epitopes are pivotal in vaccine development because B cells play a central role in antibody production and the development of memory cells that can identify pathogens in the event of a future infection (Parvizpour et al., 2020). Via Bepipred Linear Epitope Prediction 2.0 tool B-cells were predicted and total 22 B-cells were found (Table 6). Furthermore, after checking the antigenicity only 5 antigens were found as probable antigen (Table 8). Subsequently, after predicting all the epitopes seven vaccine candidates were developed with the linkers according to several articles (Table 9). Vaccine candidate 4 was chosen as the final proposed vaccine after extensive biochemical analysis was performed on all of the vaccine candidates.

Biochemical analysis included stability, molecular weight, GRAVY score, amino acid composition was predicted via protparam server. The proposed vaccine was predicted as stable (instability index- 38.99), desired molecular weight (91876.78 dalton), negative GRAVY score (-0.180) which indicates maximum solubility. The vaccine's antigenicity (0.5810) was much enhanced from the preliminary main protein (0.5173), making it well suited for inducing a robust immune response in the body. Furthermore, the proposed vaccine should not contain any allergenicity or allergic entity anticipated by the AllerTOP v2.0 (Figure 12) and Allergenonline server. In addition, the T3DB server's estimated toxicity should not be preserved in the vaccine sequence (Figure 13). These proposed vaccines had absurdly low risk profiles across the board. Phyre 2 homology modeling successfully depicted the 3D structure of the vaccine, demonstrating its reliability and range of protection (Figure 14,15). The Patchdock server was used to display the vaccine's molecular docking with the human TLR8 receptor (Figure 18 a,b). The vaccination showed potential when analyzed with a Z-score in ProSAweb (Figure 17 a,b) and a Ramachandran plot in SWISS-MODEL (Figure 16 a,b). In addition, the Ramachandran optimal region score was 94.95%, which would be good.

Al the very last, immune system simulation was simulated by the C-IMMSIM server. This server anticipated desired result for this proposed vaccine sequence and showed several graph supported it. According to this server desired immune cells were produced after vaccination with proper time frame. After about 60 days, the expected antibody response was manifested, as measured by an increase in IgG and IgM antibody levels (Figure 19 a). For preventing future infection memory cell production is very important. In the figure 19 (b) and (d) B memory cell population count and population state was shown in diagram. In figure 19 (c) Number of plasma B cells according to their isotypes was shown. Helper T cells in C-IMMSIM showed a positive predictive response, suggesting that they will contribute to spark an immunological response by activating other immune cells (Figures 19 e and f). In the Figure 19 (g) depicted the overall T cell population, memory T cell population, and cytotoxic T cell population in relation to the introduction of vaccines. Predicted numbers of active, replicating, anergic, and resting cytotoxic T cells were depicted in Figure 19 (h). Following the Figure 19 (i) and (j) was shown the population and population per state of dendritic cells.

Chapter 5
Conclusion

To sum up, vaccines are an enormous success tale, preventing millions of deaths annually. The immune system is honed to better combat harmful microbes through its use. The World Health Organization estimates that annually, vaccines prevent disease in 2.5 million people and save their lives. Developing new and effective vaccines is made easier by using a computational approach to drug discovery. To reduce the time, money, and effort required to create a vaccine or medication, in-silico analysis can be used. Many contemporary computational tools have already entered clinical trials due to their high accuracy and reliability. Thus in-silico approach was chosen for predicting RSV vaccine. There are numerous studies are ongoing against this virus although there is no licensed vaccine till the date. Respiratory syncytial virus is not crucial for adult but it can be vital for elderly and infants as it cause serious respiratory issues for them. Many cases are gathered every year globally. In the study, a multi-epitope vaccination against RSV was developed with the use of in-silico methods, with the vaccine's target being the virus's fusion glycoprotein. The vaccine was built methodically, step by step, by identifying key epitopes with the help of credible computational methods. Furthermore, in-silico biochemical study confirmed the vaccine's potential efficacy against RSV. However, there may be certain drawbacks on the in-silico method. Hence, these in-silico results may not correspond to the same promise in the in-vivo or in-vitro analysis. Despite all in-silico approach may speed up potential primary vaccine design against RSV along with many other diseases.

References

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., Donovan, C. O., Redaschi, N., & Yeh, L. L. (2004). The Universal Protein Resource (UniProt*). Swiss Institute of Bioinformatics, Centre Medical Universitaire.* https://doi.org/10.1093/nar/gkh131

Bachi, T. (1988). Direct Observation of the Budding and Fusion of an Enveloped Virus by Video Microscopy of Viable Cells Thomas. *Institute for Immunology and Virology, University of Zurich.* https://doi.org/10.1083/jcb.107.5.1689

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., … Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw1099

Bermingham, A., & Collins, P. L. (1999). The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proceedings of the National Academy of Sciences of the United States of America*, https://doi.org/10.1073/pnas.96.20.11259

Bhuiyan, M. A., Quayum, S. T., Ahammad, F., Alam, R., Samad, A., & Nain, Z. (2021). Discovery of potential immune epitopes and peptide vaccine design - a prophylactic strategy against Rift Valley fever virus [version 1; peer review: 2 approved with reservations]. *F1000Research*. https://doi.org/10.12688/F1000RESEARCH.24975.1

Borchers, A. T., Chang, C., Gershwin, M. E., & Gershwin, L. J. (2013). Respiratory syncytial virus - A comprehensive review. *Clinical Reviews in Allergy and Immunology*. https://doi.org/10.1007/s12016-013-8368-9

Buckner, C. A., Lafrenie, R. M., Dénommée, J. A., Caswell, J. M., Want, D. A., Gan, G. G., Leong, Y. C., Bee, P. C., Chin, E., Teh, A. K. H., Picco, S., Villegas, L., Tonelli, F., Merlo, M., Rigau, J., Diaz, D., Masuelli, M., Korrapati, S., Kurra, P., … Mathijssen, R. H. J. (2016) Liveness Detection in Biometrics. *InTech eBooks*. https://doi.org/10.5772/17205

Chen, X., & Jensen, P. E. (2008). The role of B lymphocytes as antigen-presenting cells. *Archivum Immunologiae et Therapiae Experimentalis*. https://doi.org/10.1007/s00005-008-0014-5

Collins, P. L., Fearns, R., & Graham, B. S. (2013). Respiratory syncytial virus: Virology, reverse genetics, and pathogenesis of disease. In *Current Topics in Microbiology and Immunology*

https://doi.org/10.1007/978-3-642-38919-1_1

Collins, P. L., & Graham, B. S. (2008). Viral and Host Factors in Human Respiratory Syncytial Virus Pathogenesis. *Journal of Virology*. https://doi.org/10.1128/jvi.01625-07

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*. https://doi.org/10.1155/2013/263952

Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*. https://doi.org/10.1186/1745-6150-8-30

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2 - A server for in silico prediction of allergens. *Journal of Molecular Modeling*. https://doi.org/10.1007/s00894-014-2278-5

Dowling, J. K., & Mansell, A. (2016). Toll-like receptors: The swiss army knife of immunity and vaccine development. *Clinical and Translational Immunology*. https://doi.org/10.1038/cti.2016.22

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

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkg563

Ghildyal, R., Ho, A., & Jans, D. A. (2006). Central role of the respiratory syncytial virus matrix protein in infection. *FEMS Microbiology Reviews*. https://doi.org/10.1111/j.1574-6976.2006.00025.x

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & Taylor, S. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition and Food Research*. https://doi.org/10.1002/mnfr.201500769

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0073957

Hossain, M. E., Rahman, M. Z., Islam, M. M., Hoque, A. F., Sumiya, M. K., Begum, M. N., Alam, M. M., Uddin, K. M. M., Hassan, M. Z., Rahman, M., Goswami, D. R., & Brooks, W. A. (2022). Pre COVID-19 molecular epidemiology of respiratory syncytial virus (RSV) among children in

Bangladesh. *Heliyon*. https://doi.org/10.1016/j.heliyon.2022.e11043

Jason S. McLellan1, William C. Ray2, and M. E. P. (2014). *Structure and Function of RSV Surface Glycoproteins*. https://doi.org/10.1007/978-3-642-38919-1

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2016). Trabajo práctico Nº 13 . Varianzas en función de variable independiente categórica. *Nature Protocols*. https://doi.org/10.1038/nprot.2015-053

Keogan, M. T., Wallace, E. M., & O'Leary, P. (2021). Cytotoxic T Cells. *Concise Clinical Immunology for Healthcare Professionals*. https://doi.org/10.4324/9780203643341-19

Khalid, K., Irum, S., Ullah, S. R., & Andleeb, S. (2022). In-Silico Vaccine Design Based on a Novel Vaccine Candidate Against Infections Caused by Acinetobacter baumannii. *International Journal of Peptide Research and Therapeutics*. https://doi.org/10.1007/s10989-021-10316-7

Kumar, N., Admane, N., Kumari, A., Sood, D., Grover, S., Prajapati, V. K., Chandra, R., & Grover, A. (2021). Cytotoxic T-lymphocyte elicited vaccine against SARS-CoV-2 employing immunoinformatics framework. *Scientific Reports*. https://doi.org/10.1038/s41598-021-86986-6

Kumar, P., & Arya, A. (2018). Ramachandran plot - A simplified approach. *Pathfinder Research and Training Foundation, India, January*, 6. https://www.researchgate.net/publication/330158666_Ramachandran_plot-_A_simplified_approach

Kuo, L., Fearns, R., & Collins, P. L. (1997). Analysis of the gene start and gene end signals of human respiratory syncytial virus: quasi-templated initiation at position 1 of the encoded mRNA. *Journal of Virology*. https://doi.org/10.1128/jvi.71.7.4944-4953.1997

Kwakkenbos, M. J., Diehl, S. A., Yasuda, E., Bakker, A. Q., Van Geelen, C. M. M., Lukens, M. V., Van Bleek, G. M., Widjojoatmodjo, M. N., Bogers, W. M. J. M., Mei, H., Radbruch, A., Scheeren, F. A., Spits, H., & Beaumont, T. (2010). Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. *Nature Medicine*. https://doi.org/10.1038/nm.2071

Kwofie, S. K., Dankwa, B., Odame, E. A., Agamah, F. E., Doe, L. P. A., Teye, J., Agyapong, O., Miller, W. A., Mosi, L., & Wilson, M. D. (2018). In silico screening of isocitrate lyase for novel anti-buruli ulcer natural products originating from Africa. *Molecules*. https://doi.org/10.3390/molecules23071550

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*.

https://doi.org/10.1186/1471-2105-8-424

Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A. C., Neveu, V., & Wishart, D. S. (2009). T3DB: A comprehensively annotated database of common toxins and their targets. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp934

Lin, H. H., Ray, S., Tongchusak, S., Reinherz, E. L., & Brusic, V. (2008). Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunology*. https://doi.org/10.1186/1471-2172-9-8

Magdeldin, S., Yoshida, Y., Li, H., Maeda, Y., Yokoyama, M., Enany, S., Zhang, Y., Xu, B., Fujinaka, H., Yaoita, E., Sasaki, S., & Yamamoto, T. (2012). Murine colon proteome and characterization of the protein pathways. *BioData Mining*. https://doi.org/10.1186/1756-0381-5-11

Manavalan, B., Govindaraj, R. G., Shin, T. H., Kim, M. O., & Lee, G. (2018). iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2018.01695

Martinelli, D. D. (2022). In silico vaccine design: A tutorial in immunoinformatics. *Healthcare Analytics*. https://doi.org/10.1016/j.health.2022.100044

Mathew, S., Taleb, S., Eid, A. H., Althani, A. A., & Yassine, H. M. (2021). *In silico virtual screening of lead compounds for major antigenic sites in respiratory syncytial virus fusion protein*. https://doi.org/10.1007/s42247-021-00213-6

Munoz, F. M., & Englund, J. A. (2019). Respiratory syncytial virus. *Maternal Immunization, January*. https://doi.org/10.1016/B978-0-12-814582-1.00011-5

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*. https://doi.org/10.1038/srep42851

Pandaya, M. C., Callahan, S. M., Savchenko, K. G., & Stobart, C. C. (2019). A Contemporary View of Respiratory Syncytial Virus. *Pathogens*. https://doi.org/10.3390/pathogens8020067

Parvizpour, S., Pourseif, M. M., Razmara, J., Rafi, M. A., & Omidi, Y. (2020). Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discovery Today*. https://doi.org/10.1016/j.drudis.2020.03.006

Peebles, R. S., & Graham, B. S. (2005). Pathogenesis of respiratory syncytial virus infection in the murine model. *Proceedings of the American Thoracic Society*. https://doi.org/10.1513/pats.200501-002AW

Proença-Módena, J. L., Acrani, G. O., Snider, C. B., & Arruda, E. (2011). Respiratory Viral Infections. In *Tropical Infectious Diseases: Principles, Pathogens and Practice* (Thrid Edit). Elsevier Inc. https://doi.org/10.1016/B978-0-7020-3935-5.00058-6

Rahman, N., Ali, F., Basharat, Z., Shehroz, M., Khan, M. K., Jeandet, P., Nepovimova, E., Kuca, K., & Khan, H. (2020). Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: An immunoinformatics approach. *Vaccines*. https://doi.org/10.3390/vaccines8030423

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0009862

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

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., Peters, B., Nielsen, M., & Nielsen, M. (2020). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *Journal of Proteome Research*. https://doi.org/10.1021/acs.jproteome.9b00874

Sameer, A. S., & Nissar, S. (2021). Toll-Like Receptors (TLRs): Structure, Functions, Signaling, and Role of Their Polymorphisms in Colorectal Cancer Susceptibility. *BioMed Research International*, *2021*. https://doi.org/10.1155/2021/1157023

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gki481

Shafat, Z., Faizan, I., Tazeen, A., Farooqui, A., Deeba, F., Aftab, S., Hamza, A., Parveen, S., Islam, A., Broor, S., Ahmed, & Parveen, S. (2017). In-silico analysis of ectodomain G protein of Respiratory Syncytial Virus . *Indian Journal of Health Sciences and Care*. https://doi.org/10.5958/2394-2800.2017.00023.2

Shang, Z., Tan, S., & Ma, D. (2021). Respiratory syncytial virus: From pathogenesis to potential therapeutic strategies. *International Journal of Biological Sciences*. https://doi.org/10.7150/ijbs.64762

Sheik, S. S., Sundararajan, P., Hussain, A. S. Z., & Sekar, K. (2002). Ramachandran plot on the web. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/18.11.1548

Spann, K. M., Tran, K.-C., Chi, B., Rabin, R. L., & Collins, P. L. (2004). Suppression of the Induction of Alpha, Beta, and Gamma Interferons by the NS1 and NS2 Proteins of Human Respiratory Syncytial Virus in Human Epithelial Cells and Macrophages. *Journal of Virology*. https://doi.org/10.1128/jvi.78.8.4363-4369.2004

Srinivasakumar, N., Ogra, P. L., & Flanagan, T. D. (1991). Characteristics of fusion of respiratory syncytial virus with HEp-2 cells as measured by R18 fluorescence dequenching assay. *Journal of Virology*. https://doi.org/10.1128/jvi.65.8.4063-4069.1991

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2017.00292

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm290

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*. https://doi.org/10.1093/nar/gku1004

Yeo, K. T., Yung, C. F., Khoo, P. C., Saffari, S. E., Sng, J. S. P., How, M. S., & Quek, B. H. (2021). Effectiveness of Palivizumab against Respiratory Syncytial Virus Hospitalization among Preterm Infants in a Setting with Year-Round Circulation. *Journal of Infectious Diseases*. https://doi.org/10.1093/infdis/jiaa749