In Silico Epitope Prediction and Multi-Epitope Vaccine Design Targeting Human Elongation Factor 1-Alpha 1 (EF1A1)

By

Mottakian Ahamed Alif
22146020

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

School of Pharmacy

BRAC University

June 2026

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing 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/We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mottakian Ahamed Alif
22146020

Approval

The thesis titled "**In Silico Epitope Prediction and Multi-Epitope Vaccine Design Targeting Human Elongation Factor 1-Alpha 1 (EF1A1)"** submitted by Mottakian Ahamed Alif (22146020) of spring 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on June 2026.

Examining Committee:

Supervisor:

_____

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

Program Coordinator:

_____

Namara Mariam Chowdhury
Senior Lecturer and Program Coordinator
School of Pharmacy, BRAC University

Departmental Head
(Chairperson):

_____

Sharmind Neelotpol, PhD
Professor and Chairperson
School of Pharmacy, BRAC University

Ethics Statement

This thesis was conducted in compliance with all the relevant ethical standards and did not involve human or animal trials.

Abstract

Worldwide dissemination of SARS-CoV-2 have accentuated the necessity for effective vaccines. Here, a design for multiepitope vaccine (MEV) was applied. The epitope pools were computationally predicted, and many tests were done for CTL, HTL and B-cell epitopes. The chosen epitopes were connected to form a single construct using appropriate linkers in order to maintain structural configuration and maximise immunogenicity. The vaccine's physicochemical properties and stability were tested utilizing ProtParam, its 3D structure predicted through Phyre2 and validated via Ramachandran plots, ERRAT, QMEAN affirming a credible stereochemical conformation. The docking study of interaction with Toll-like receptor 8 (TLR8) is for binding affinity of the design. In silico simulations utilizing C-ImmSim indicated strong antibody responses (IgM, IgG1, IgG2), activated $CD4^+$ and $CD8^+$ T cells in balanced states with quality memory formation based on combined Th1/Th2-derived cytokine profiles. In summary, the in-silico analyses suggest that it is highly immunogenic, not structurally unstable. Now it should provide a logic basis for experimental validation and rapid vaccine development against current and emergent viral variants.

Keywords: SARS-CoV-2; In Silico Vaccine; Multi-epitope Protein; Molecular Docking.

Dedication

I want to dedicate my work to my parents, they supported me in all of the great events that happened during my life. I appreciate every one of you help and advice.

Acknowledgement

I would like to take this opportunity to thank my supervisor Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, BRAC University for his everlasting support patience and constructive criticism. He encouraged me to do additional research, which has turned into an educational opportunity.

I owe my sincere thanks and admiration to my parents who stand behind me and motivate me always. I would also like to thank my seniors and closest friends for all the help you have rendered.

Table of Contents

List of Figures

List of Tables

List of Acronyms

CTL Cytotoxic T Lymphocyte ($CD8^+$ T cell)

ERRAT Structure Verification Algorithm for Non-Bonded Atomic Interactions

HTL Helper T Lymphocyte ($CD4^+$ T cell)

QMEAN Qualitative Model Energy Analysis

TLR-8 Toll-Like Receptor 8

Chapter 1

Introduction

1.1 Viral Morphology and Pathophysiology

Vaccination is a successful intervention. By inducing an active acquired immunity, it generates protective immune responses against specific pathogens via vaccination. Active immunity can develop naturally after an infection or artificially after a vaccination (an introduction of a weakened, killed, or selected component of the disease-causing organism into the body that prompts immune recognition and memory [CDC 2024]). Vaccines can be either prophylactic to prevent infection pre-morbidity or therapeutic, aiming at enhancing vaccine-specific immune responses that function against an established disease condition. One option is, depending on the type of pathogen (bacterial or viral) and the immune response needed, there have been developments of different types of vaccines. COVID-19 is caused by SARS-CoV-2. However, it is transmitted by secretions and causes respiratory illness with varied severity, vaccine development targeting this virus continues as an active area of interest in biomedical research. Next generation vaccine design is increasingly reliant on the use of immunoinformatic, a computational approach. This strategy saves time as well as cost by in silico screening of potential vaccine candidates prior to experimental validation through suitable bioinformatics instruments (Khan et al., 2023).

SARS-CoV-2 comes from the Coronaviridae family, which is a positive single-stranded RNA virus and is an enveloped virus. SARS-CoV-2, according to the International Committee on Taxonomy of Viruses (ICTV) is grouped as well as annotated as belonging to the realm Riboviria; order Nidovirales; suborder Cornidovirineae; family Coronaviridae; subfamily Orthocoronavirinae; genus Betacoronavirus and subgenus Sarbecovirus. The virus: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Species Severe acute respiratory

syndrome-related coronavirus. Morphologically Coronaviruses have spherical, enveloped particles and a surface covered with a spike projection. The spike protein is what forms an appearance of the virus which is crown like when viewed through electron microscopy, and that is what provides coronavirus with its name. Virion characteristics: Members of Orthocoronavirinae are enveloped and generally contain large, positive-sense RNA genomes (22–36 kb), they are 80–160 nm in size. The SARS-CoV-2 genome has information to develop non-structural proteins that are important for viral replication, in addition to structural proteins such as: spike (S), membrane (M), envelope (E) and nucleocapsid proteins (N). The SARS-CoV-2 pathogenesis mostly starts when the spike viral protein connects itself to the angiotensin-converting enzyme 2 receptor of the target cells. Following receptor binding, host proteases (ex. TMPRSS2, cathepsins and furin-like proteases) cleave and trigger the spike protein for viral cell fusion or endocytosis aiding entry of virus into the cell. This entry into target human tissues is crucial in terms of defining the efficiency through which the virus can be internalized by human ACE2-expressing respiratory epithelial cells and other epithelia (Jackson et al., 2022; Li et al., 2024). Once inside a target cell, SARS-CoV-2 commandeers intracellular machinery to replicate its own RNA genome and make viral proteins. Some of the proteins go back into making new viral particles, which assemble and are released to infect neighboring cells. Clinical manifestations of infection can vary from uncomplicated respiratory symptoms to pneumonia, acute respiratory distress syndrome, multi-organ dysfunction and expiration. Severe Disease Is Associated with Aberrant Immune Response and Immunopathological Changes Abnormal immune response, high levels of inflammatory cytokines, loss of immune balance promote tissue damage leading to disease progression (Li et al., 2024). Coronaviruses are dispersed among warm-blooded animals. Alpha- and betacoronaviruses typically infect mammals (particularly bats), whereas the gamma- and deltacoronaviruses are predominantly found in birds (or/and other indeterminate origin [20]).

The broad host range along with mutation and recombination means that coronaviruses are major pathogens capable of causing interspecies transmission, leading to outbreaks in humans and animals. Thus, it is significant to characterize morphology, taxonomy (Woo et al., 2023) in order to facilitate vaccine design, development of antiviral drugs and supervision of future coronaviruses.

1.2 The Genome Organization and Proteomic Structure

SARS-CoV-2 has linear genome approximately 30 kb in length. Studies made in the last two years referred to SARS-CoV-2 genome as composed by ~29 viral proteins, including 16 nonstructural proteins, 4 principal structural and some further accessory proteins supporting for viral replication, host interaction and pathogenicity (Bai et al., 2022; Jhanwar et al., 2024). The genome contains UTRs (untranslated regions) on the 5′ and 3′ ends, a large replicase region, ORF1a/ORF1ab in the 5′ quarter that encodes polyproteins which are split into functional non-structural proteins. Its genome lacks hemagglutinin–esterase (HE) gene present in some betacoronaviruses (and typical of lineage A betacoronaviruses; see Jhanwar et al., 2024). ORF1a and ORF1ab translate into long polyproteins, respectively yielding a set of non-structural proteins or NSPS. These NSPS play a significant role in viral replication, transcription, immune evasion, and proteolytic processing of proteins. The principal among them is nsp3 having a papain-like protease domain, and nsp5 referred to as main protease or 3C-like protease. These virally-encoded proteases process the viral polyproteins to yield mature functional proteins and are thus potential antiviral drug development targets (Tamayo-Ordóñez et al., 2023; Yevsieieva et al., 2023). The nucleocapsid protein acts as packaging for viral RNA genome while spike, membrane and envelope proteins are with viral envelope. Spike protein, whose gene imparts binding to angiotensin-converting enzyme 2 receptor facilitating

viral access into target cells, is currently of particular mention. The spike protein includes two functional subunits with receptor-binding function in S1 and fusion function in S2 (Minigulov et al., 2024). The production of the membrane as well as envelope proteins are vital aspects of viral assembly, morphogenesis and budding. It is the structural protein most abundant and also functions as a central organizer during virion assembly through interactions with other viral components. Collectively, these structural proteins function to assemble infectious viral particles and maintain the viral life cycle in host cell (Zhang et al., 2022; Minigulov et al., 2024). Hence genome and proteomic organization of SARS-CoV-2 display a greatly coordinated system in which non-structural proteins predominantly control for replication and transcription in contrast to structural proteins that primarily preserve viral morphology, entry, assembly, and release. Such genomic and protein-based characteristics are important for drug, vaccine, and identifying development in resistance to SARS-CoV-2.

1.3 The overall process (Entry, Replication, and Pathogenesis) in the Host System

SARS-CoV-2, the causal virus of the COVID-19 disease, mainly enters into the host cells through interaction between the angiotensin-converting enzyme 2 receptor and the spike glycoprotein. When the spike protein is bound to ACE2, then host proteases (TMPRSS2, furin and endosomal cathepsins) cleave the spike protein which activates it for membrane fusion between viral and host-cell membranes. This entry can be via direct fusion at the cell surface or by the endosomal route (Jackson et al., 2022; Li et al., 2024). After the virus penetrates, its RNA genome gets released into the cytoplasm of the host cell. Both ORF1a and ORF1ab regions are encoded as large polyproteins, with the orf1ab in a single reading frame requiring programmed ribosomal frameshifting to produce the longer pp1ab polyprotein. The

polyproteins are cleaved into non-structural proteins that make up the replication-transcription complex, which is necessary for viral RNA synthesis, transcription and genome replication (V'kovski et al., 2021; Rehfeld et al., 2023; Nazir et al., 2024). After replication and translation of SARS-CoV-2, structural proteins are translocated into the ER and traverse through the ER-Golgi intermediate compartment. Viral assembly is the process by which newly synthesized genome that is encapsulated by nucleocapsid protein merges with spike, membrane and envelope proteins. These newly assembled virions are then encapsulated into vesicles that are secreted from the infected cell in a general secretory or lysosome-associated pathway. Studies have shown that SARS-CoV-2 forms around Golgi apparatus, ERGIC and vesicular transport (Scherer et al., 2022; Katiyar et al., 2024). Pathogenesis is not just about viral replication. Infected cells can have an effect on respiratory epithelial cells and initiate innate immune reactions. In mild infection, the viral clearance is supported by an adequate immune response but in severe COVID-19 patients, delayed interferon responses with excessive cytokine production leading to worsening epithelial and endothelial injury (cytokine storm), inflammation and coagulation abnormalities may underlie the pneumonia (Lamers & Haagmans, 2022; Li et al., 2024).

Chapter 2

Methodology

2.1 Suitable protein selection

Sequences of viral proteins are usually obtained in the FASTA format and computationally screened for immunoinformatic based vaccine design from reliable biological databases such as NCBI Protein database and Uniprot. Recently, related SARS-CoV-2 vaccine-design studies have also employed protein sequences obtained from public databases to generate appropriate antigenic regions and predict B-cell and T-cell epitopes suitable for use (Sarvmeili et al., 2024; Deepthi et al., 2025). The antigenicity was checked by VaxiJen (v2) after selection of sequences. Using the "virus" model, VaxiJen, are an alignment-independent tool for predicting protective antigenicity. In the most recent SARS-CoV-2 multi-epitope vaccine studies, the Vaxine thresholds of 0.5 under the viral model for proteins or vaccine constructs have been treated as antigens and were retained for downstream immunoinformatic evaluation (Sarvmeili et al., 2024). The target protein prepared in the present study gave a predicted score as an overall protective antigen prediction of 0.5179, which is capable of producing the possible antigen; therefore, it is appropriate for epitope prediction (Sarvmeili et al., 2024; Deepthi et al., 2025).

2.2 CTL epitope identification

CTLs are the main effectors of antiviral immunity, recognizing short peptide fragments displayed on endogenous major histocompatibility complex class I (MHC-I) molecules. These peptide-MHC complexes are recognized by $CD8^+$ T cells and induce apoptosis of infected host

cells, limiting viral replication (Grifoni et al., 2020; Mateus et al., 2020). Prediction of CTL epitopes using the NetCTL 1.2 server: This is by far the most commonly utilised tool for immunoinformatic-assisted design (Larsen et al., 2007). The SARS-CoV-2 spike protein sequence from the NCBI Protein database in FASTA file format was submitted to NetCTL 1.2. For epitope prediction, the sensitivity (TP/ (TP + FN)) and specificity (TN/ (TN + FP)) threshold were set at 0.75. Because of their critical role in regulating virus spread and inducing protective immunity (Grifoni et al., 2020), CTL epitopes were chosen based on predicted $CD8^+$ T-cell responsiveness. The in silico predictions are helpful for faster selection of putative epitopes to further validate them in laboratory experiments and facilitate multi-epitope vaccines against SARS-CoV-2 (Mateus et al., 2020; Peeples, 2021). Next, the CTL epitopes that overlapped were filtered on the basis of antigenicity, toxicity and allergenicity; resulting in determining epitope sequences capable of being utilized for immunological tests in sequential vaccine design. This provides a rational guided in silico immunization strategy to the highest probability of inducing safe, effective cellular immunity.

2.3 Analyzing the Prediction of CTL epitope

CTLs are a major component of antiviral immunity; they specifically recognize short peptide fragments that are presented on MHC class I molecules (Grifoni et al., 2020; Mateus et al., 2020). For SARS-CoV-2 based Vaccine designing approaches in immunoinformatics, NetCTL 1.2 server is often launched for prediction and identification of CTLs epitopes as one more common method. This tool is a combination of input to predict MHC class I binding affinity (Larsen et al., 2007), leading to predictions peptide sequences. The sequence of the SARS-CoV-2 spike protein was extracted in FASTA format from NCBI Protein database and used to submit it to NetCTL 1.2 for this study. We determined the optimal thresholds for epitope

prediction in a range of 0.47–0.48 that gave good sensitivity and specificity values. The CTL epitopes predicted were then selected for their efficiency in inducing $CD8^+$ T-cell responses, which are important for controlling viral proliferation and improving protective immunity (Grifoni et al., 2020). These in silico predictions can funnel candidate epitopes before lab confirmation (Mateus et al., 2020; Peeples, 2021). The resulting CTL epitopes were assessed for antigenicity, immunogenicity, toxicity, and allergenicity to obtain the best potential candidates representative of sequences that prompted optimal immune response (immunophenotype). Our framework offers a rational method to guide the design of vaccines that elicit effective T cell responses but limit potential safety issues.

2.4 HTL Epitopes Evaluating

HTLs are important for immune responses through the activation of both CTLs and B cells. HTLs augment both cellular and humoral immunity with eosinophils and polyclonal B-cell production of neutralizing antibodies, which are critical for targeting viral pathogens (Grifoni et al., 2020; Mateus et al., 2020).

In this study NetMHCIIpan 4.0 server was used. This tool will predict peptide binding affinity to MHC Class II molecules and integrates information from multiple HLA-DR, HLA-DP and HLA-DQ alleles that allows a more complete assessment of potential helper T-cell epitopes (Reynisson et al., 2020). Fifteen-mer peptides from the SARS-CoV-2 spike protein sequence in but FASTA format were generated and binding affinities against 34 HLA alleles were predicted, including the 20 most common alleles and another 14 to ensure broad population coverage. From those results, high (top 1%) and low performing (top 5%) binding predictions were filtered in order to shortlist the candidate HTL epitopes. Some of the selected best binders

were subsequently used for bioinformatics, antigenicity, allergenicity and toxicity evaluation in the same computational workflow as CTL epitopes. This step helps to incorporate only immunologically relevant and safe epitopes into the final multi-epitope vaccine design (Deepthi et al., 2025, Sarvmeili et al., 2024). In silico prediction of HTL epitopes holds the key to designing effective vaccines that successfully induce strong $CD4^+$ T-cell responses, which are essential in establishing memory leading to long-term immunity and production.

2.5 Analysis of HTL Epitopes

In this study, the cytokines of interest were the pivotal members of Th1- and Th2-type immune responses: interferon-gamma (IFN-γ), interleukin-4 (IL-4), and interleukin-10 (IL-10) (Sallusto et al., 2004; Cox et al., 2021). IFN-γ is chiefly linked to the Th1-type of response and upregulates macrophage activity, therefore enhancing cell-mediated mechanisms required for clearing viruses. IFNepitope server was used for the prediction of IFN-γ-inducing epitopes, and only positive peptides were chosen for further analysis. As another example, IL-4 and IL-10 are also critical for Th2-type immune responses and antibody production. IL-4 is a potent B-cell activator and class-switching immunoglobulin mechanism, whereas IL-10 serves as an anti-inflammatory cytokine that regulates immune responses to avoid overt inflammation (Cox et al., 2021; Li et al., 2022).

2.6 Selection of B cell Epitopes and Analysis

Long-term humoral immunity is mediated by B cells, which are capable of producing antibodies and generating memory B cells (Sette & Crotty, 2021). 1, Immunoinformatics-based vaccine design: The IEDB Analysis Resource was used. The Bepipred Linear Epitope Prediction 2.0 was used, which ranks amino acid residues and produces an antigenicity profile

along the protein using machine-learning methods (Jespersen et al., 2017). Lastly, only peptides of 7 to 30 amino acids were selected in order to reduce the possibilities in downstream analysis and also provide biologically relevant epitope lengths. Antigenicity, allergenicity and toxicity of selected B-cell epitopes were assessed. This workflow guarantees the B-cell epitopes incorporated in a vaccine are able to elicit strong humoral immunity while minimizing side effects (Deepthi et al., 2025; Sarvmeili et al., 2024).

2.7 Final Vaccine Construction

The desired CTL, HTL and B-cell epitopes were connected and with specific linkers for high immunogenicity. An adjuvant (a built-in molecular component computationally fused to the vaccine's antigen sequence) was used at the N-terminal to help the immune system. EAAAK linker, Rigid and proper folding added between the adjuvant and CTL epitope. GPGPG linker, Linker used between the HTL epitopes to enhance processing and flexibility. KK linker, solubilized and made antibody accessible by linking B-cell epitopes.

The modular structure guarantees the correct spacing of the epitopes, their structural stability and the induction of both humoral and cellular immunity (Ayyagari et al., 2022; Sarvmeili et al., 2024).

So, the process is given below:

Figure 1: The process of vaccine construction

2.8 Evaluation of Constructed Vaccine for Antigenicity, Allergenicity, and Toxicity

The multi-epitope vaccine was then tested for immunogenic properties after assembly with appropriate linkers. Quantification of antimicrobial peptide–vaccines: The antigenicity was re-assessed using VaxiJen 2.0 server with the cutoff, 0.5 in viral model and our vaccine maintained or further improved its antigenicity. Antigenicity was re-evaluated through the viral model, where a cut-off of 0.5 was presented, showing retained or enhanced antigenicity of the vaccine. Allergenicity was validated in Allergen Online server and toxicity in T3DB. Non-allergenic and non-toxic epitope selection is critical for the safety and immunogenicity (Deepthi et al., 2025; Ayyagari et al., 2022). This step in the in-silico evaluation is important to test whether the designed multi-epitope vaccine can induce immune response and diminish possible adverse effect.

2.9 Constructed Vaccine (Biochemical Properties)

The biochemical and the physicochemical characteristics were performed by ProtParam tool. Some of the important parameters are the total number of amino acids, molecular weight, theoretical isoelectric point (pI), atomic composition, and amino acid frequencies (Wilkins et al., 1999). Half-life of the construct was also determined to obtain prediction of its stability in vivo. The instability index is a prediction of whether the vaccine is likely to remain stable in vitro or not: the lower it is, the more stable the vaccine is likely to be. The GRAVY (Grand Average of Hydropathicity) score determines hydrophilicity or hydrophobicity; negative values are hydrophilic and soluble proteins, and positive values are hydrophobic (Kyte & Doolittle, 1982). Also, aliphatic index, which is based on the volume of aliphatic side chains,

gives an estimation of thermostability for stable constructs, with higher values (>60) preferred. (Deepthi et al., 2025; Sarvmeili et al., 2024).

2.10 Constructed Vaccine (3D Structure Prediction)

Predicting the 3D structure of the designed multi-epitope vaccine has been carried out by using the Phyre2 server based on homology modelling. This approach uses a known database of structures to find structurally similar proteins, which is used to generate a reliable model (Kelley et al., 2015). This tool gives two important parameters; confidence and coverage which show the reliability of the predicted model. A score of 100% and a level of coverage greater than 80% can be taken as an indicator of a very reliable structure. The vaccine sequence was submitted to Phyre2 to obtain 3D structure, which was downloaded in PDB format. The structure was then visualized and analyzed with the help of the software PyMOL which helped to assess the overall folding and structural integrity of the vaccine construct. Therefore, an important step in 3D modelling is to ensure the correct display and accessibility of the epitopes (Deepthi et al., 2025; Sarvmeili et al., 2024).

2.11 Ramachandran Plot Generation and Structural Validation

The 3D structure model of the vaccine was predicted by using Phyre2 and then validated by using SwissModel. It calculates the backbone dihedral angles (phi(φ) and psi(ψ)) of each residue from the protein structure (Kiefer et al. 2008). 90–95% residues in favoured regions and <2% in outlier or disallowed regions represent a high quality protein model. Ramachandran plot show both visually, and numerically at what the structure is relatively reliable, while also showing areas that can be improved. Validating this is important because for a vaccine to work, the Multi-epitope Vaccine Construct must fold correctly, be thermally stable and present epitopes to the host. This is important to validate that it should be a stereo chemically accurate

confirmation and it folds into a stable source of epitopes in order to effectively stimulate an immune response, for this reason it is required.

2.12 3D Model Quality Assessment by ERRAT and QMEAN

Further assessment of the quality of the determined 3D structure pertaining to the vaccine construct was performed through ERRAT and QMEAN analysis methods. ERRAT assesses non-bonded atomic interactions that may suggest structural discrepancy and which are then compared with those observed experimentally with a score (Benkert et al., 2011; Colovos & Yeates, 1993) using the model at both global and local levels, QMEAN. Usually, an ERRAT score > 80 % indicates a good protein model and the QMEANDisCo global score is about 0–1. (Deepthi et al. 2025, Sarvmeili et al., 2024).

2.13 Molecular Docking (Toll-Like Receptors)

The binding affinity of the designed epitopes with the immune receptors (Lymphocyte receptors) of host by molecular docking was studied through ClusPro server. Phyre2 was employed to retrieve the vaccine construct in PDB format (Comeau et al., 2004; Reynisson et al., 2020) and Toll-like receptor 8 (TLR8) was retrieved from RCSB Protein Data Bank. ClusPro runs several docked complexes and provides a ranking of the best to worst based on the binding energies. The model with the highest binding affinity was selected for further analysis. In the prediction of such interaction with innate immune receptor, downstream capacity of signal transducer is prerequisite to induce strong immune reaction (Deepthi et al., 2025; Sarvmeili et al., 2024). The orientation and access of epitopes for immune recognition can be confirmed via visual inspection.

2.14 Simulation of the Vaccine (Immune Response)

We evaluated the immunogenic properties in a platform called C-ImmSim server. This tool is capable of prediction of humoral immunologic responses (activation of B-cells, antibody formation), CD4+ (HTL) activity and CD8+ (CTL) (Rapin et al., 2010). The simulation helps to assess how well the vaccine can create long-term immunological memory, and determine its potential interactions with molecules that trigger adaptive immune responses, such as Toll-like receptors (TLRs) that are crucial for adaptive immune system activation (Deepthi et al., 2025; Sarvmeili et al., 2024). C-ImmSim has the ability to model these responses computationally, thus giving critical insight on the likely efficacy and safety of the vaccine prior to experimental testing and saving time and resources in guiding vaccine design strategies.

2.15 Key Aspects, Materials, and Methods

The design of the multi-epitope vaccine was done in-silico, integrating epitope prediction, vaccine construction and structural validation. Immunoinformatics tools were used to find CTL, HTL, and B-cell epitopes. The cytokine-inducing potential of the selected epitopes was also checked to ensure that they could induce both Th1 and Th2 type immune responses (Deepthi et al., 2025; Rapin et al., 2010). The vaccine construct was designed with proper linkers (EAAAK, GPGPG, KK) for structural stability and exposing the epitopes. The biochemical properties were studied using ProtParam, and the 3D structure was generated by Phyre2 and checked with the Ramachandran plot, Z score (ProSA) and ERRAT/QMEAN scoring. The innate immune activation prediction was carried out by molecular docking by ClusPro. Lastly, predictions of the formation of immunological memory were performed using

C-ImmSim simulations, creating a comprehensive pipeline for the evaluation of vaccine immunogenicity prior to experimental testing (Sarvmeili et al., 2024; Deepthi et al., 2025).

Chapter 3

Results

3.1 Antigenicity Check of protein sequence

From the UniProt website, which is known as a large protein resource, we got our protein sequence (UniProt ID: P68104) in FASTA format.

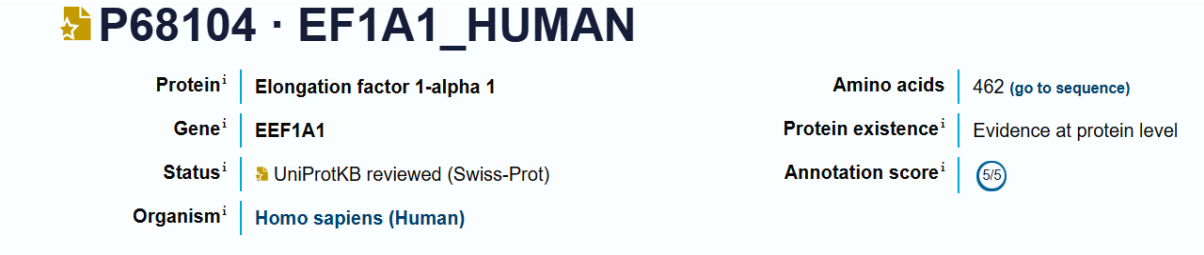


Figure 2: Protein retrieval from UniProt (UniProt, n.d.)

Now, by the help of the Vaxijen 2.0 tool, we got the antigenicity score, which is 0.5179, where our targeted organism was a virus.

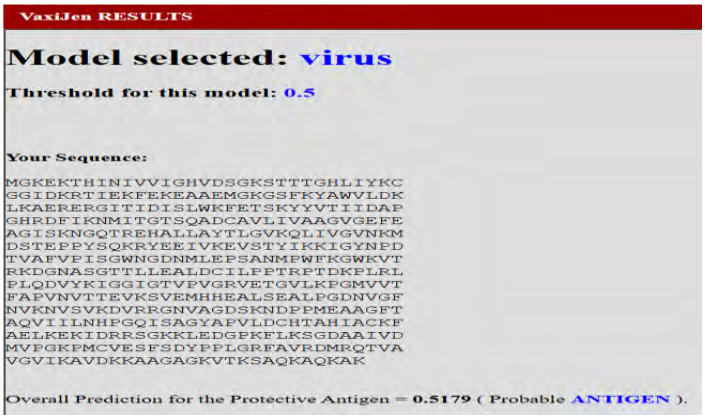


Figure 3: Protein's initial antigenicity using Vaxijen 2.0's 0.5 threshold application (Doytchinova & Flower, 2007).

3.2 Identifying and Evaluating CTL Epitopes

In this regard, CTL epitopes were predicted by the NetCTL 1.2 server as previously reported: prediction of MHC class I binding affinity, proteasomal cleavage (±600 Kcal/mole) and TAP (transporter associated with antigen processing) transport efficiency (Larsen et al., 2007). In this study, we chose the A1 supertype, where the C-terminal cleavage weight is 0.15, TAP transport weight = 0.05 and the epitope identification threshold is 0.75. We used the combined score sorting option to pick the epitopes with the highest predicted scores. In these analyses, seventeen human CTL epitopes labelled with "<-E" from the SARS-CoV-2 spike protein emerged as the top-scoring candidates. The remaining epitopes were evaluated further for antigenicity, allergenicity, toxicity and cytokine-inducing capabilities to only include the immunologically relevant sequences in a final multi-epitope vaccine construct (Deepthi et al., 2025; Sarvmeili et al., 2024).

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

  21 ID FASTA pep STTTGHLIY aff   0.7541 aff_rescale   3.2017 cle 0.9772 tap   3.0210 COMB   3.4993 <-E
 175 ID FASTA pep STYIKKIGY aff   0.2592 aff_rescale   1.1005 cle 0.9341 tap   3.1900 COMB   1.4001 <-E
  72 ID FASTA pep TIDISLWKF aff   0.2567 aff_rescale   1.0900 cle 0.8930 tap   2.4750 COMB   1.3477 <-E
 364 ID FASTA pep HTAHIACKF aff   0.2601 aff_rescale   1.1043 cle 0.6520 tap   2.4320 COMB   1.3237 <-E
 154 ID FASTA pep KMDSTEPPY aff   0.2362 aff_rescale   1.0028 cle 0.9440 tap   3.1180 COMB   1.3003 <-E
 108 ID FASTA pep QADCAVLIV aff   0.2792 aff_rescale   1.1854 cle 0.6070 tap   0.2580 COMB   1.2894 <-E
  48 ID FASTA pep EMGKGSFKY aff   0.1459 aff_rescale   0.6193 cle 0.9640 tap   2.8330 COMB   0.9055 <-E
 133 ID FASTA pep TREHALLAY aff   0.1457 aff_rescale   0.6187 cle 0.8810 tap   2.9170 COMB   0.8967 <-E
 286 ID FASTA pep TTEVKSVEM aff   0.1715 aff_rescale   0.7283 cle 0.9426 tap   0.1650 COMB   0.8779 <-E
 169 ID FASTA pep EIVKEVSTY aff   0.1126 aff_rescale   0.4781 cle 0.9178 tap   2.7600 COMB   0.7538 <-E
```

Figure 4: CTL epitopes from NetCTL1.2 (Larsen et al., 2007)

The NetMHCpan 4.1 immunoinformatic tool was used to identify suitable binding alleles, resulting in 9-mer peptide sequences for all HLA alleles in the MHC I class.

CTL	Antigenicity	Allergenicity	Toxicity
STTTGHLIY	Antigen	Allergen	Non-Toxin
STYIKKIGY	Antigen	Non-Allergen	Non-Toxin
KMDSTEPPY	Antigen	Non-Allergen	Non-Toxin
HTAHIACKF	Antigen	Non-Allergen	Non-Toxin
EMGKGSFKY	Antigen	Allergen	Non-Toxin
EIVKEVSTY	Non-Antigen	Allergen	Non-Toxin
HTAHIACKF	Antigen	Non-Allergen	Non-Toxin

Table 1: CTL Epitope Selection Based on Antigenicity, Allergy, and Toxicity Filtering

3.3 Identifying and Evaluating HTL Epitopes

HTLs, $CD4^+$ T cells recognise antigens presented on MHC class II molecules and activates the immune system via cytokine release, including IFN-γ, IL-4, and IL-10. It mediate macrophage activation, B-cell stimulation, and inflammation regulation, respectively (Grifoni et al., 2020; Mateus et al., 2020). HTL epitopes were predicted using the NetMHCIIpan 4.0 server across multiple HLA-DR, HLA-DP, and HLA-DQ alleles. Strong binders were selected at the top 1% threshold, and weak binders at 5%, incorporating binding affinity (BA) predictions. Fifteen-mer peptides were generated, and top-scoring epitopes labelled "<=SB" were selected after removing redundancy.

The selected HTL epitopes were further screened for antigenicity, allergenicity andtoxicity and cytokine-inducing potential (IFNepitope for IFN-γ, IL4Pred for IL-4, IL10Pred for IL-10). Epitopes, meeting all six conditions were included, ensuring strong cellular and humoral immune activation with balanced cytokine responses (Deepthi et al., 2025; Sarvmeili et al., 2024).

Peptide	Antigenicity	Allergenicity	Toxicity	IFN-γ	IL-4	IL-10
RTIEKFEKEAAEMGK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
TIEKFEKEAAEMGKG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
IEKFEKEAAEMGKGS	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
EKFEKEAAEMGKGSF	Antigen	Allergen	Non-toxin	positive	Inducer	Inducer
RYEEIVKEVSTYIKK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
YEEIVKEVSTYIKKI	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
PDTVAFVPISGWNGD	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
DTVAFVPISGWNGDN	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
TVAFVPISGWNGDNM	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
EEIVKEVSTYIKKIG	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Inducer
KEKIDRRSGKKLEDG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
VRDMRQTVAVGVIKA	Antigen	Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer
RERGITIDISLWKFE	Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer
ERGITIDISLWKFET	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
IKKIGYNPDTVAFVP	Non-Antigen	NAllergen	Non-toxin	positive	Non-Inducer	Non-Inducer
KKIGYNPDTVAFVPI	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
KIGYNPDTVAFVPIS	Antigen	Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer

ETSKYYVTIIDAPGH	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
TSKYYVTIIDAPGHR	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer

SKYYVTIIDAPGHRD	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KYYVTIIDAPGHRDF	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
HRDFIKNMITGTSQA	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
RDFIKNMITGTSQAD	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
AQVIILNHPGQISAG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
QVIILNHPGQISAGY	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
FAELKEKIDRRSGKK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VAVGVIKAVDKKAAG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
AVGVIKAVDKKAAGA	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VGVIKAVDKKAAGAG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
GHRDFIKNMITGTSQ	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
DFIKNMITGTSQADC	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
FETSKYYVTIIDAPG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
MVVTFAPVNVTTEVK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VVTFAPVNVTTEVKS	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
YYVTIIDAPGHRDFI	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
DGPKFLKSGDAAIVD	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
GPKFLKSGDAAIVDM	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
RFAVRDMRQTVAVGV	Non-Antigen	Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer
FAVRDMRQTVAVGVI	Non-Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer

AVRDMRQTVAVGVIK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
RDMRQTVAVGVIKAV	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
YSQKRYEEIVKEVST	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
SQKRYEEIVKEVSTY	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
QKRYEEIVKEVSTYI	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
ACKFAELKEKIDRRS	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
EAAGFTAQVIILNHP	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
AAGFTAQVIILNHPG	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
AGFTAQVIILNHPGQ	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VKQLIVGVNKMDSTE	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
KQLIVGVNKMDSTEP	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KRYEEIVKEVSTYIK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
GVIKAVDKKAAGAGK	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
THINIVVIGHVDSGK	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
HINIVVIGHVDSGKS	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Inducer
IDKRTIEKFEKEAAE	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
DKRTIEKFEKEAAEM	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KRTIEKFEKEAAEMG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
GVKQLIVGVNKMDST	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
FTAQVIILNHPGQIS	Non-Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer

TAQVIILNHPGQISA	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
TVAVGVIKAVDKKAA	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VIKAVDKKAAGAGKV	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
IVKEVSTYIKKIGYN	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer

VKEVSTYIKKIGYNP	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KEVSTYIKKIGYNPD	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
GDNVGFNVKNVSVKD	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KYAWVLDKLKAERER	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
APVNVTTEVKSVEMH	Non-Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Inducer
PVNVTTEVKSVEMHH	Non-Antigen	Allergen	Non-toxin	positive	Non-Inducer	Inducer
KNVSVKDVRRGNVAG	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
NVSVKDVRRGNVAGD	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KPGMVVTFAPVNVTT	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
PGMVVTFAPVNVTTE	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
VAFVPISGWNGDNML	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
YIKKIGYNPDTVAFV	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
IGYNPDTVAFVPISG	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
INIVVIGHVDSGKST	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
NIVVIGHVDSGKSTT	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
KGSFKYAWVLDKLKA	Non-Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer

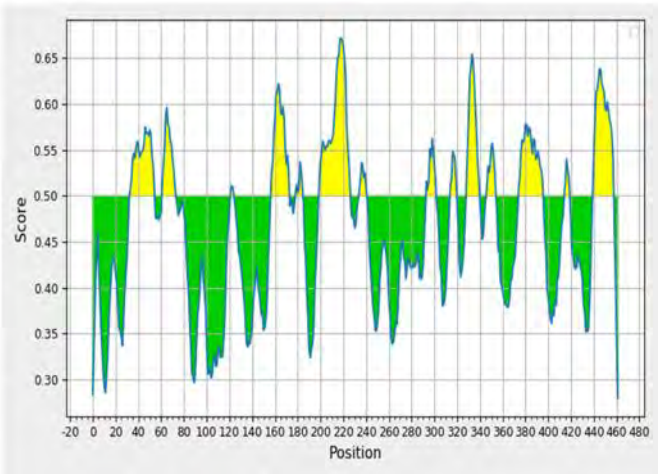
QTVAVGVIKAVDKKA	Antigen	Non-Allergen	Non-toxin	positive	Non-Inducer	Non-Inducer
VESFSDYPPLGRFAV	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer

Table 2: HTL epitope selection

3.4 Epitope of B cell Evaluation

The function of B lymphocytes, which are responsible for producing antibodies for disease fighting against pathogens and memory B cells. For SARS-CoV-2, IEDB Analysis Resource was used (Fleri et al., 2017). This creates a score vs residue position plot, where peaks correspond to regions of the protein that are likely binding sites (antigenic epitopes) for antibodies and valleys correspond to non-antigenic regions.

Based on the results of these searches, we selected peptides containing 7–30 amino acids as optimal for antibody recognition and antigenic stability. Antigenicity of these candidate epitopes were evaluated. Also, allergenicity and toxicity were tested at levels that guarantee just non-risky, immunogenic arrangements were remembered for the multi-epitope immunization development. By this pipeline, the vaccine can powerfully elicit humoral immune responses and cellular immunity.



Predicted peptides:

No.	Start	End	Peptide	Length
1	34	55	IDKRTIEKFEKEAAEMGKGSFK	22
2	62	73	KLKAERERGITI	12
3	122	124	EFE	3
4	158	173	TEPPYSQKRYEEIVKE	16
5	179	185	KKIGYNP	7
6	200	227	NMLEPSANMPWFKGWKVTRKDGNASGTT	28
7	235	241	ILPPTRP	7
8	294	302	MHHEALSEA	9
9	315	320	VSVKDV	6
10	329	340	SKNDPPMEAAGF	12
11	347	355	LNHPGQISA	9
12	375	396	LKEKIDRRSGKKLEDGPKFLKS	22
13	415	419	FSDYP	5
14	440	458	AVDKKAAGAGKVTKSAQKA	19

Figure 5: Graph of B cell from protein

Additional analyses were performed for further validation so that fulfilled all three of these criteria were called B cell epitopes Editable table, with those determined to be better for process and helpful being included.

Epitope	Length	Antigenicity	Allergenicity	Toxicity
IDKRTIEKFEKEAAEMGKGSFK	22	Non-Antigen	Allergen	Non-Toxic
KLKAERERGITI	12	Non-Antigen	Allergen	Non-Toxic
TEPPYSQKRYEEIVKE	16	Non-Antigen	Allergen	Non-Toxic
KKIGYNP	7	Antigen	Non-Allergen	Non-Toxic
NMLEPSANMPWFKGWKVTRKDGNASGTT	28	Non-Antigen	Allergen	Non-Toxic
ILPPTRP	7	Non-Antigen	Allergen	Non-Toxic
MHHEALSEA	9	Non-Antigen	Allergen	Non-Toxic
SKNDPPMEAAGF	12	Antigen	Non-Allergen	Non-Toxic
LNHPGQISA	9	Non-Antigen	Allergen	Non-Toxic
LKEKIDRRSGKKLEDGPKFLKS	22	Non-Antigen	Allergen	Non-Toxic
AVDKKAAGAGKVTKSAQKA	19	Antigen	Non-Allergen	Non-Toxic

Table 3: Selection of epitopes of B cell

3.5 The Final Construction of vaccine Using Appropriate Linker

Once all the antigenicity, non-allergenicity and non-toxicity of CTL, HTL and B-cell epitopes were evaluated and chosen optimally. In multi-epitope designs, linkers are key components because they preserve the native conformational status of each epitope and proper folding while avoiding steric hindrance to maintain immunogenicity (Ayyagari et al., 2022). For the final construct, the EAAAK linker was to be interlinking the adjuvant with CTL epitope that facilitated structural separation whilst providing rigidity. GPGPG linker was used between

HTL epitopes to improve flexibility and promote efficient presentation. When combining B-cell epitopes, a KK linker was employed to increase solubility and guarantee ability for antibody recognition. For maximal antigenicity, only the top-ranked epitopes were selected, yielding a construct with 3 CTL epitopes, 1 HTL epitope and 2 B-cell epitopes. Such a strategy allows the design of stable, immunogenic and structurally sound vaccine candidates that would elicit strong cellular (T cell) or humoral (B cell) immune responses (Deepthi et al., 2025; Sarvmeili et al., 2024).

So, the final vaccine design is given below:

MGKEKTHINIVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAEMGKGSFKYAWVL
DKLKAERERGITIDISLWKFETSKYYVTIIDAPGHRDFIKNMITGTSQADCAVLIVAAGV
GEFEAGISKNGQTREHALLAYTLGVKQLIVGVNKMDSTEPPYSQKRYEEIVKEVSTYIKK
IGYNPDTVAFVPISGWNGDNMLEPSANMPWFKGWKVTRKDGNASGTTLLEALDCILPPTR
PTDKPLRLPLQDVYKIGGIGTVPVGRVETGVLKPGMVVTFAPVNVTTEVKSVEMHHEALS
EALPGDNVGFNVKNVSVKDVRRGNVAGDSKNDPPMEAAGFTAQVIILNHPGQISAGYAPVL
DCHTAHIACKFAELKEKIDRRSGKKLEDGPKFLKSGDAAIVDMVPGKPMCVESFSDYPPLRF
AVRDMRQTVAVGVIKAVDKKAAGAGKVTKSAQKAQKAKEAAAKKMDSTEPPYEAAAKHTA
HIACKFEAAAKHTAHIACKFGPGPGEKFEKEAAEMGKGSFKKKKIGYNPKKAVDKKAAGAG
KVTKSAQKA

3.6 The process of checking (Antigenicity, Allergenicity and Toxicity)

Once a final sequence was constructed, it was evaluated again for antigenicity value using VaxiJen 2.0 with a threshold of 0.5. The analysis revealed a considerable augmentation of its antigenicity as compared to the original protein; this indicates that this construct can be potentially recognized by the immune system and may induce protection (Deepthi et al., 2025; Sarvmeili et al., 2024).

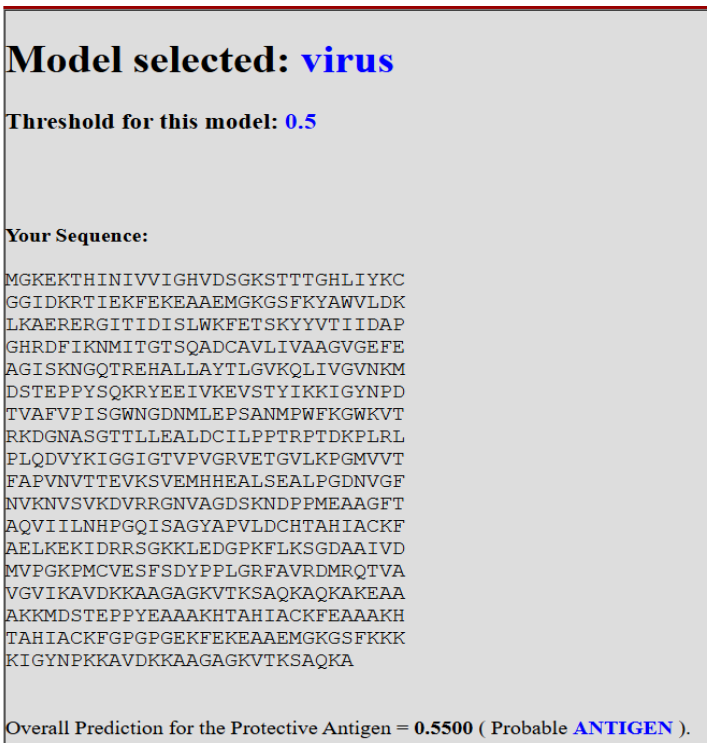


Figure 6: Increased Antigenicity of constructed Vaccine

Allergenicity was predicted by the AllergenOnline server via comparing the input FASTA sequence with a database of allergenic proteins. The output stated "Number of sequences with hits: 0", thus proving the vaccine construct to be non-allergic.

Database	AllergenOnline Database v24 (January 26, 2026)
Input Query	>FASTA MGKEKTHINIVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAEMGKGSFKYAWVL DKLKAERERGITIDISLWKFETSKYYVTIIDAPGHRDFIKNMITGTSQADCAVLIVAAGV GEFEAGISKNGQTREHALLAYTLGVKQLIVGVNKMDSTEPPYSQKRYEEIVKEVSTYIKK IGYNPDTVAFVPISGWNGDNMLEPSANMPWFKGWKVTRKDGNASGTTLLEALDCILPPTR PTDKPLRLPLQDVYKIGGIGTVPVGRVETGVLKPGMVVTFAPVNVTTEVKSVEMHHEALS EALPGDNVGFNVKNVSVKDVRRGNVAGDSKNDPPMEAAGFTAQVIILNHPGQISAGYAPV LDCHTAHIACKFAELKEKIDRRSGKKLEDGPKFLKSGDAAIVDMVPGKPMCVESFSDYPP LGRFAVRDMRQTVAVGVIKAVDKKAAGAGKVTKSAQKAQKAKEAAAKKMDSTEPPYEAAA KHTAHIACKFEAAAKHTAHIACKFGPGPGEKFEKEAAEMGKGSFKKKKIGYNPKKAVDKK AAGAGKVTKSAQKA
Length	554
Number of 80 mers	475
Number of Sequences with hits	0

No Matches of Greater than 35% Identity Found

Figure 7: Allergenicity Check of Constructed Vaccine

The T3DB database was accessed to subject the sequence given in FASTA format for any potential hits, which returned no results and consequently deemed the construct as non-toxic (Wishart et al., 2015). An integrated assessment enables assurance that the vaccine designed is immunogenic, safe and appropriate for subsequent experimental demonstration.

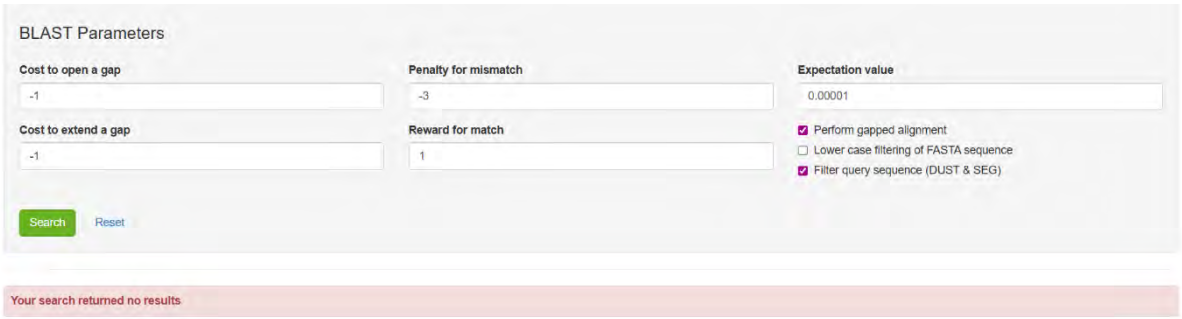


Figure 8: T3DB server displaying no results solidifying the vaccine construct is non-toxic

3.7 Evaluation of Constructed Vaccine's Biochemical Properties

The biochemical nature of the multi-epitope vaccine provides an explanation for its functioning in biological systems. These properties were calculated using ProtParam, which calculates various physicochemical parameters from the amino acid sequence (Wilkins et al., 1999). Important parameters include total amino acids, amino acid composition, theoretical pI (which tells you if the protein is acidic/basic/neutral), and molecular weight. ProtParam also predicts atomic composition, the extinction coefficient and half-life of the protein. The instability index evaluates in vitro protein stability with proteins below 40 (stable). The aliphatic index (thermostability) shows with values greater than 60 indicating higher stability. The GRAVY score (Grand Average of Hydropathicity) reflects hydrophilicity or hydrophobicity; negative values indicate hydrophilic and soluble proteins, which are favourable. for vaccine constructions. These properties assure designed stability, solubility, and the ability to perform additional immunological tests (Deepthi et al., 2025; Sarvmeili et al., 2024).

PortParam results are given below:

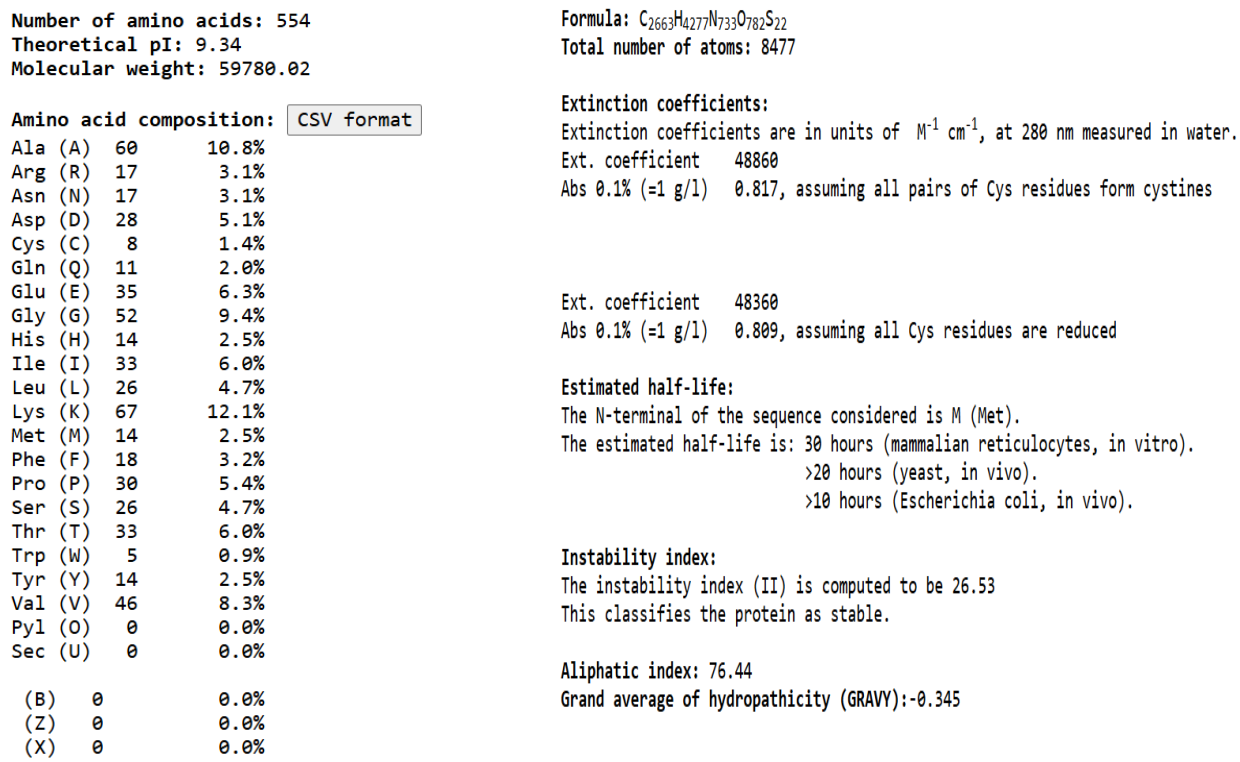


Figure 9: Biochemical Property Analysis values by the ProtParam tool (Wilkins et al., 1999).

The length of our sequence is 554 amino acids long, also fitting into the optimal length range of between 400-600 amino acids, allowing for potential diversity among epitopes while remaining compact enough to maintain structural functionality and stability. The isoelectric point value of 9.34 is also sufficiently close to neutrality, as it is favourable for solubility and stability in a physiological system, which allows us to clot blood. Half-life estimated in mammals is shown to be 30 hours, which guarantees an in vivo stability. Additionally, the molecular weight of our construct is approximately 59.78 kDa. Even more so, our computed score on the instability index is 26.53, meaning, in comparison, the construct, such as our score, is very significantly stable with the assumed level. Furthermore, according to the aliphatic

index score (> 60), it is not only thermostable (76.44) but also has a hydrophilic property since our GRAVY score is -0.345.

3.8 Constructed Vaccine's Model Generation in 3D

Multi-epitope vaccine tertiary structure prediction was performed using Phyre2, an immunoinformatics tool that uses homology modelling to generate the structure for proteins from amino acid sequences (Kelley et al., 2015). Phyre2 was used to find structurally similar templates from its database and generate a PDB-format model for the vaccine sequence.

Structural reliability was assessed, involving two parameters, coverage and confidence. The predicted model was 82% covered, meaning 82% of the sequence was modelled with a high confidence of 100% for the included 456 residues in the structure. The 3D model was visualised and analysed by PyMOL, as it folds correctly and the epitope is accessible to immunogenicity. This validation confirms that it reaches a stable three-dimensional structure compatible with downstream analyses, like docking and immune simulations (Deepthi et al. 2025; Sarvmeili et al. 2024).

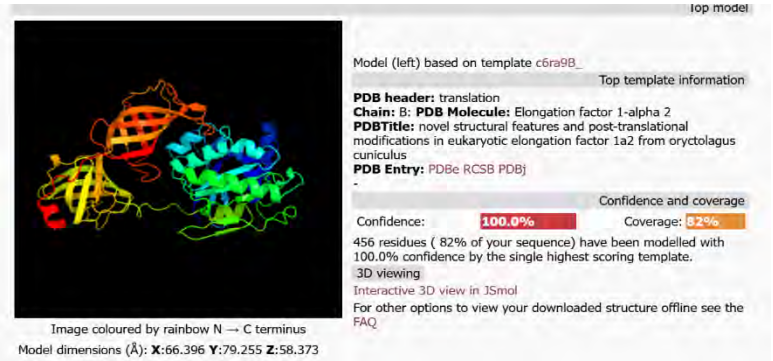


Figure 10: Phyre2.2 generated 3D model with 82% residue coverage at 100% confidence

3.9 Ramachandran Plot generation for 3D Model's Quality Assessment

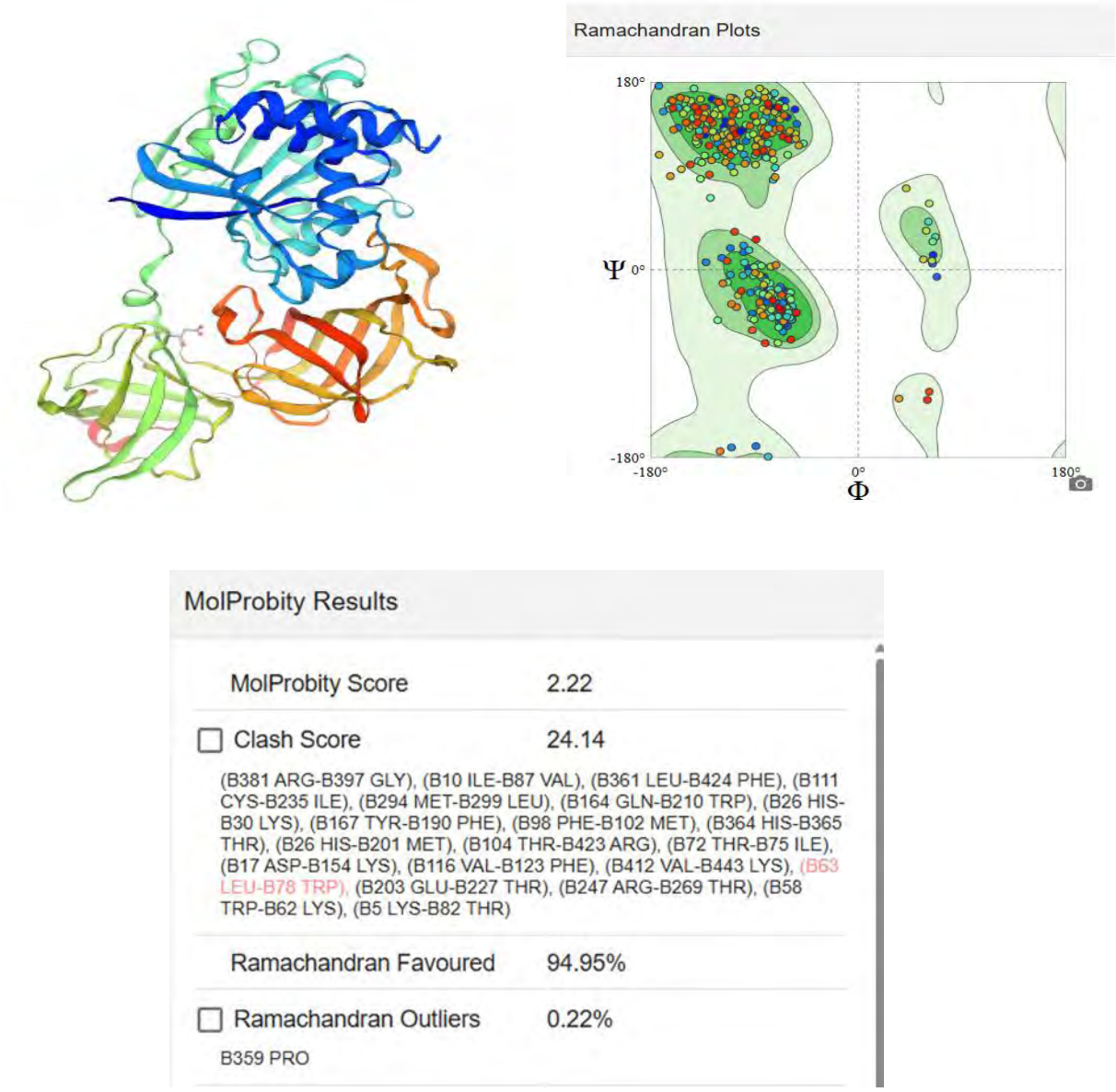


Figure 11: Ramachandran plot generation by Swiss-Model Expasy

Phyre2-derived models were assessed using the SwissModel Expasy server. The Ramachandran plot guarantees both correct protein folding and general structural reliability through assessment (Kiefer et al., 2008).

It indicated that 94.95% of residues fell in energetically favoured regions (greater than 90%) while only 0.22% of residues were in disallowed regions (less than 2%). These findings suggest correct folding with little structural stress. Furthermore, C-beta deviations were less than 0.25 Å, which also confirms the structural reliability. This validation confirms that it reaches a consistent stereochemical conformation appropriate for further modelling studies like molecular docking and immune simulations (Deepthi et al., 2025; Sarvmeili et al., 2024).

3.10 Analysis of ERRAT and QMEAN

ERRAT analysis showed the stereochemical quality. ERRAT is used for assessing non-bonded atomic interactions between different atom types. The construct compared favourably to previously published high-resolution models, with an average quality factor of 88.0631% (at the 80% r.m.s.d. threshold for contact point score) being above the threshold considered indicative of a high-quality model that is comparable to crystallographic structures. This corroborated that non-bonded interactions were mainly in tolerable limits and thus lent credibility to the predicted structural model (Colovos & Yeates, 1993; Mahapatra et al., 2023).

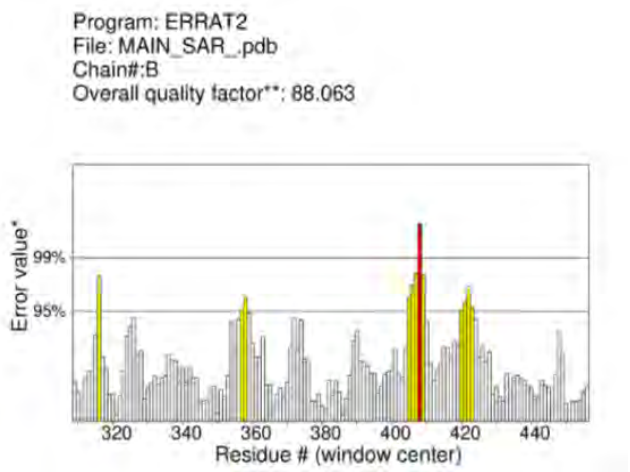


Figure 12: Overall quality factor analysis by ERRAT

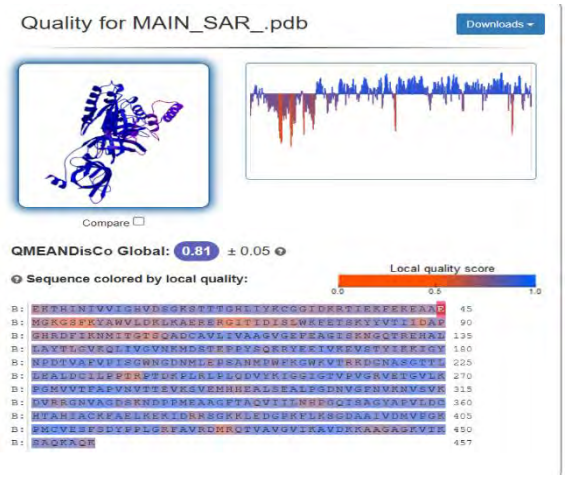


Figure 13: QMEANDisCo Global score

QMEANDisCo's global quality assessment further investigation into qualities returned a global score of 0.81 (±0.09), which is generally within the acceptable response range from 0 to 1 (Dijkstra et al., 2023). Scores approaching 1 mean similar alignment with the high-resolution experimental structure, while negative scores indicate a dubious degree of quality. The score we got is closer to 1, which indicates it is a high-quality model and has a high-resolution experimental structure (Benkert et al., 2011).

3.11 Molecular Docking

To assess the binding ability of the multi-epitope vaccine to immune receptors, we further performed molecular docking with ClusPro to identify stable protein-protein interactions (Kiflu, 2024; Tombari et al., 2025). PDB Extracts: Receptor representation using Toll-like receptor 8 (TLR8) and the modelled vaccine construct using Phyre2. A large number of docking poses based on structural similarity were generated using ClusPro. Out of all identified clusters, cluster 0 was selected as a representative model due to the largest number of members

(77), indicating a high degree of reliability, demonstrating that it also had a stable structure and strong binding affinity for this ligand. Through cluster selection, we can select cluster 0, which offers a balance of high binding energy as well as reproducibility, thereby validating the full vaccine construct for the ability to interact with TLR8 and inducing innate immune activation leading to enhanced immunogenicity.

Cluster	Members	Representative	Weighted Score
0	77	Center	-1034.0
		Lowest Energy	-1039.7
1	72	Center	-780.3
		Lowest Energy	-955.6
2	49	Center	-766.2
		Lowest Energy	-886.4
3	38	Center	-922.2
		Lowest Energy	-926.2
4	32	Center	-795.6
		Lowest Energy	-927.1
5	19	Center	-816.4
		Lowest Energy	-836.4
6	19	Center	-766.3
		Lowest Energy	-883.3
7	18	Center	-750.6
		Lowest Energy	-875.9
8	18	Center	-802.1
		Lowest Energy	-859.2
9	17	Center	-777.4
		Lowest Energy	-793.5
10	17	Center	-876.6
		Lowest Energy	-876.6

Figure 14: Docked complex and model energy scores using ClusPro

3.12 Simulation of Immune

The C-ImmSim server was used to evaluate the immune profile of a designed multi-epitope vaccine based on epitope sequences and different parameters of the immune system (Rapin et al., 2010). Antigen reduction (black peaks) through clearance at the two doses is clearly depicted by both simulation and experiment. IgM antibodies predominated during the primary response, but a dramatic increase in IgG1 and IgG2 subtypes was observed due to successful

memory B cell activation. Boosting exacerbated the relative IgG + IgM response, confirming long-lived immunological memory was established. The presence of IgG reflects long-lasting humoral immunity, an important factor for persistent protection. Such findings indicate that the immunogenicity of the vaccine is robust enough to elicit primary and memory immune responses, leading to optimal recognition (and removal) of antigen following a re-challenge via data supporting protective efficacy of vaccine formulation(s) (Deepthi et al., 2025; Sarvmeili et al., 2024; Rapin et al., 2010).

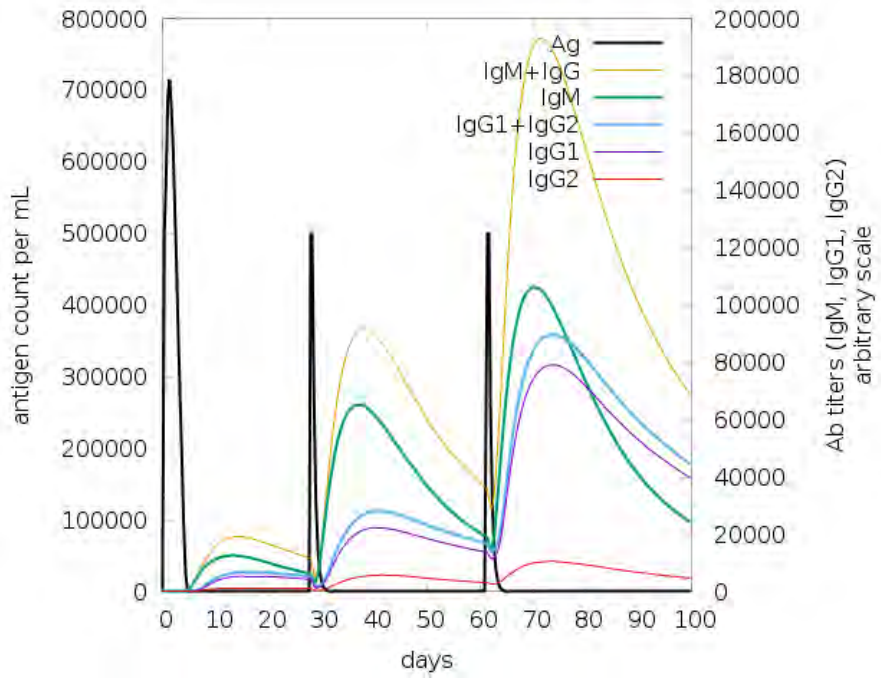


Figure 15: Immune simulation (C-IMMSIM)

The immunostimulation of IMDH-B701 multi-epitope vaccine suggested a remarkably high level of IFN-γ and IL-2, indicating robust Th1-type responses and activation of cytotoxic T lymphocytes. The abundance of these cytokines in the different epithelial compartments also reflects a balanced activation of helper T cells and immune regulation with moderate amounts of IL-10 and IL-12 levels, preventing overt inflammation through uncontrolled cellular responses. Because the vaccine also supports humoral (antibody-mediated) immune responses,

these data suggest that the vaccine effectively activates both the development of strong cellular immunity as well as the production of antibodies, indicating its potential to induce comprehensive and balanced protective immunity. (Deepthi et al., 2025; Sarvmeili et al., 2024; Rapin et al., 2010).

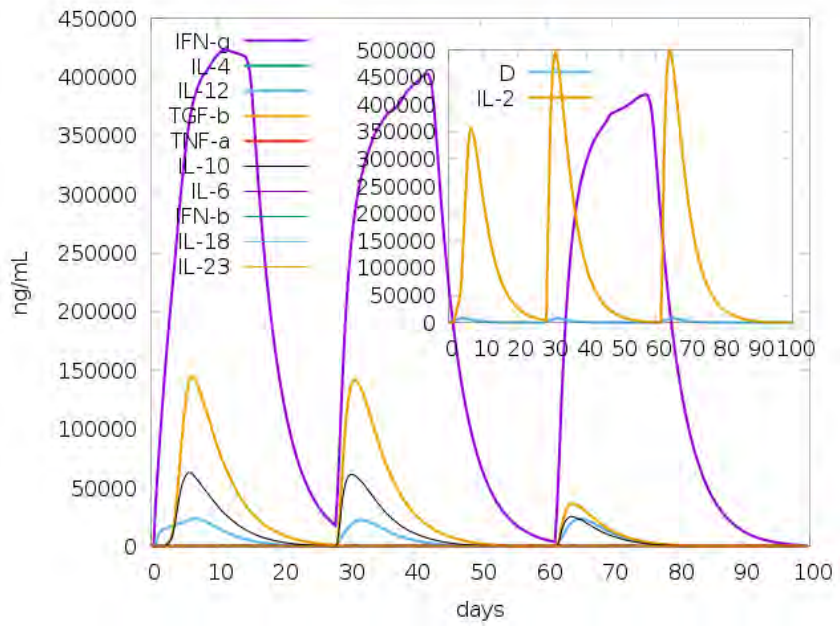


Figure 16: Levels of cytokines by C-IMMSIM

So, this immune simulation shows that it can produce memory B cells, which in turn ensures a sustained humoral immunity. This was supported by further successful IgM to IgG isotype switching and functional antibody activity measured after administering the vaccine. Production of IgM, as well as subtypes IgG1 and IgG2, was ongoing, with an increase in memory B cells. That will likely ensure a long-lasting immune response through continued antigen processing. (Deepthi et al., 2025; Sarvmeili et al., 2024; Rapin et al., 2010).

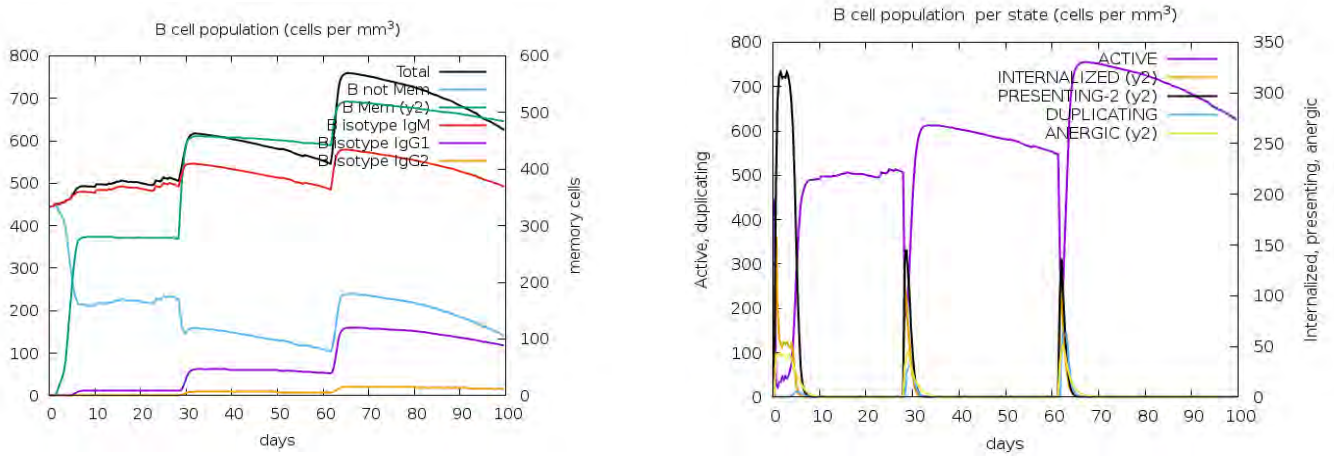


Figure 17: Total count of B lymphocytes

The multi-epitope vaccine's simulation indicated strong activation and memory formation of $CD4^+$ T-helper cells. There was a statistically significant increase in total $CD4^+$ T-helper cells and memory $CD4^+$ T-helper cell populations following each vaccination dose. Such a pattern reflects enduring cellular immunity, which is essential in sustaining immune protection through effective coordination of humoral and cytotoxic responses (Deepthi et al., 2025; Sarvmeili et al., 2024; Rapin et al., 2010).

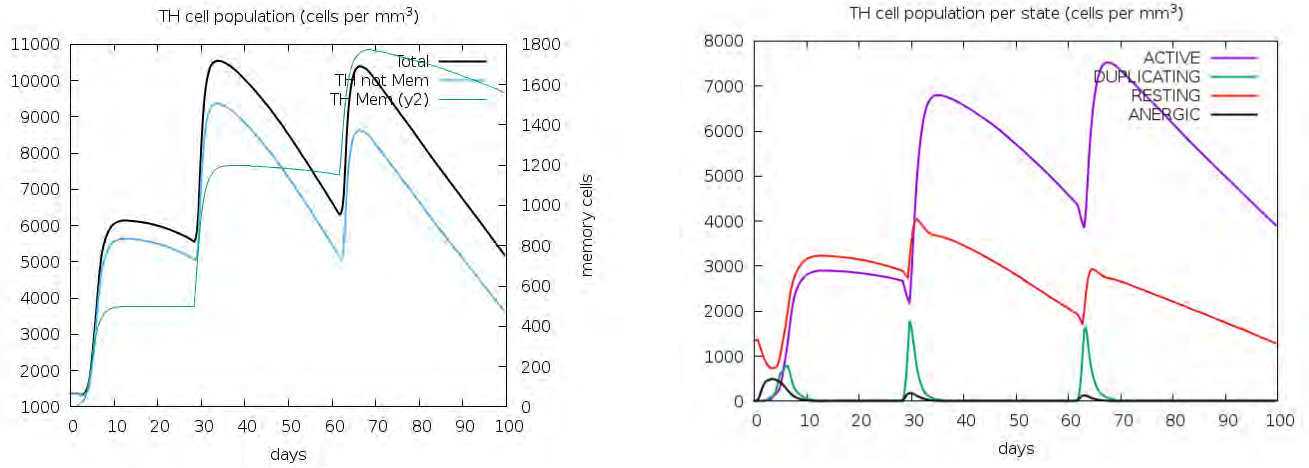


Figure 18: Number of CD4 T-helper lymphocytes

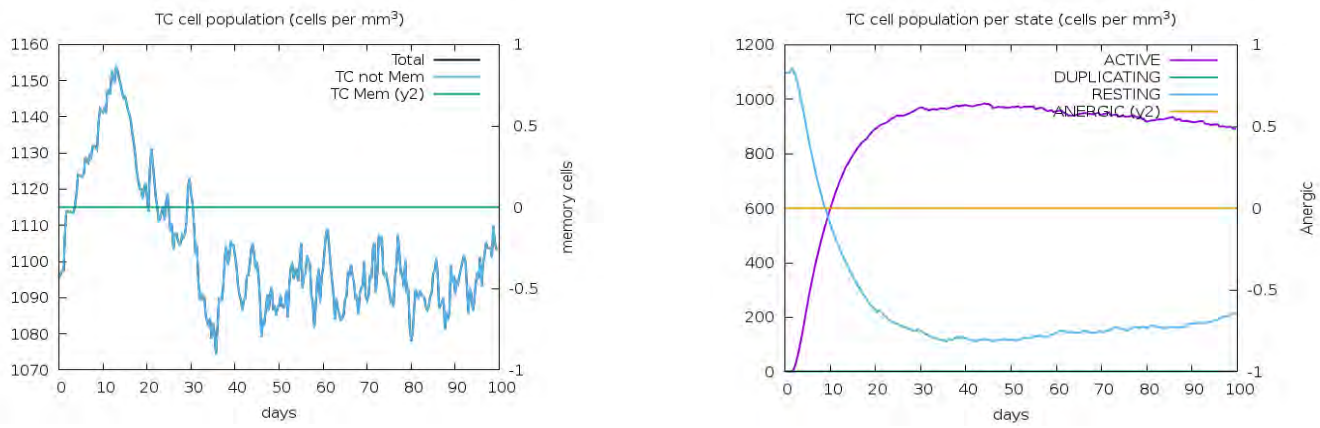


Figure 19: Number of CD8 T-cytotoxic lymphocytes

After vaccination, CTLs undergo proliferation, and many of the resultant cells activate early and eliminate abnormal or pathogen-infected cells. Part of these cells become memory CTLs that remain in the organism for a lifetime, offering long-term protection. The immune response is strong but restrained in the increase of anergic CTLs that contain excessive cytolytic activity and therefore allow potent yet controlled cellular immunity (Deepthi et al., 2025; Sarvmeili et al., 2024; Rapin et al., 2010).

Chapter 4

Discussion

SARS-CoV-2 has risen into its current global pandemic status. The antigenicity score of the final construct (0.55) indicates that this will most likely elicit an immune response, which is consistent with previously described studies emphasizing that activating appropriate epitopes is crucial for vaccine effectiveness (Mukherjee et al., 2022; Zhang et al., 2023). With CTL epitopes binding to MHC class-I and HTL epitopes eliciting MHC class II pathways, robust cellular immune activation was ensured by predictions for both. Both Th1- and Th2-mediated immunity are stimulated, an essential step needed for protection from SARS-CoV-2 through a balanced immune response (Li et al., 2021). The computer-based C-ImmSim simulations targeted the B-cell epitopes selective for induction of neutralizing antibodies (achievable by T-cell responses, with simultaneous immunity and immediate memory). The physicochemical analysis showed an instability index of 26.53, a thermostable (aliphatic index = 76.44) and a hydrophilic nature (GRAVY = –0.345), indicating a good behaviour in the biological systems (Wilkins et al., 1999). The Phyre2-constructed models and 3D structure prediction were combined with PyMOL visualization, which indicates that the protein exists in its most native state, with a coverage of 83%, and predicts it folds with 100% confidence. The modelled vaccine structure was found to have good stereochemical quality as assessed by structural validation using Ramachandran plots, ERRAT, and QMEAN (Kumar et al., 2023). The weighted score for molecular docking with TLR8 showed high binding affinity as well as the ability to cluster reliably; that will initiate downstream adaptive responses. Immune simulations suggested strong humoral and cell-mediated immune responses, rapid IgM production, class-switched IgG1 and IgG2, and memory B-cell generation.

Overall, these computational analyses proposed the designed multi-epitope vaccine as an immunogenic construct that can induce effective, potent and balanced immune responses, including long-lived memory cells, to clear antigens rapidly with potential cross-protection. So, multi-epitope vaccines confer broader protection and a lower risk of escape from immune detection than single-antigen approaches (Farid et al., 2023; Ahmed et al., 2022). The integrated immunoinformatic pipeline presented here provides a rational approach for accelerated vaccine design against emerging pathogens.

Chapter 5

Conclusion

Herein, the current study illustrates a rational design of a system which indicates an optimistic multiepitope vaccine to combat SARS-CoV-2. The constructed vaccine is predicted so that it can give high cellular immune action. The stereochemical stability was tested by structural validation like homology modelling, Ramachandran plot analysis, ERRAT, QMEAN and Z-score calculation. In our research, docking of molecular studies showed high binding to TLR8, suggesting efficient engagement of innate immune receptors and adequate antibody production along with long-lived memory B and T cell development. The importance of rapid antigen clearance, positive cytokine profiles and immunological memory to confer protection and its emerging versions was highlighted with the vaccine construct. Taken together, these data support the process of this system with potential applicability in experimental platforms and lay a foundation for rapid acceleration of SARS-CoV-2 vaccine development. However, future studies comprising in vitro and in vivo assessments are imperative to prove their usefulness and safeness as well as immunogenicity against biological systems. The computer pipeline described here illustrates the potential of immunoinformatic to help inform rational vaccines for emerging infectious diseases.

References

Centers for Disease Control and Prevention. (2024). *Explaining how vaccines work.*

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines

Khan, M. S., et al. (2023). Immunoinformatics design of B- and T-cell epitope-based vaccine against SARS-CoV-2.

World Health Organization. (n.d.). *Coronavirus disease (COVID-19).*

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species *severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2.

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells.

Li, X., Mi, Z., Liu, Z., & Rong, P. (2024). SARS-CoV-2: Pathogenesis, therapeutics, variants, and vaccines.

Woo, P. C. Y., de Groot, R. J., Haagmans, B., Lau, S. K. P., Neuman, B. W., Perlman, S., Sola, I., van der Hoek, L., Wong, A. C. P., & Yeh, S.-H. (2023). ICTV virus taxonomy profile: Coronaviridae 2023.

Bai, C., Zhong, Q., & Gao, G. F. (2022). Overview of SARS-CoV-2 genome-encoded proteins.

Jhanwar, A., Sharma, D., & Das, U. (2024). Unravelling the structural and functional dimensions of SARS-CoV2 proteins in the context of COVID-19 pathogenesis and therapeutics.

Minigulov, N., Boranbayev, K., Bekbossynova, A., Gadilgereyeva, B., & Filchakova, O. (2024).

Tamayo-Ordóñez, M. C., Rosas-García, N. M., Ayil-Gutiérrez, B. A., Bello-López, J. M., Tamayo-Ordóñez, F. A., Anguebes-Franseschi, F., Damas-Damas, S., & Tamayo-Ordóñez, Y. d. J. (2023). Non-structural proteins: A marker for detection of human coronavirus families.

Yevsieieva, L. V., Kulesh, D. A., & Konevega, A. L. (2023). Main and papain-like proteases as prospective targets for antiviral therapy of SARS-CoV-2 infection. *Biomedicines, 11*(11), 3002.

Zhang, Z., Nomura, N., Muramoto, Y., Ekimoto, T., Uemura, T., Liu, K., Yui, M., Kono, N., Aoki, J., Ikeguchi, M., Noda, T., & Iwata, S. (2022). Structure of SARS-CoV-2 membrane protein essential for virus assembly.

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells.

Katiyar, H., Arduini, A., Li, Y., & Liang, C. (2024). SARS-CoV-2 assembly: Gaining infectivity and beyond.

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis.

Nazir, F., Kombe Kombe, A. J. K., Khalid, Z., Bibi, S., Zhang, H., Wu, S., & Jin, T. (2024). SARS-CoV-2 replication and drug discovery.

Rehfeld, F., Maticzka, D., Grosser, R., Knauff, P., Eravci, M., Vida, I., Backofen, R., Mendell, J. T., & Caliskan, N. (2023). CRISPR screening reveals a dependency on ribosome recycling for efficient SARS-CoV-2 programmed ribosomal frameshifting and viral replication.

Scherer, K. M., Mascheroni, L., Carnell, G. W., Wunderlich, L. C. S., Makarchuk, S., Brockhoff, M., Mela, I., Fernandez-Villegas, A., Barysevich, M., Stewart, H., Chlanda, P., & others. (2022).

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2.

Deepthi, V., Sasikumar, A., Mohanakumar, K. P., & Rajamma, U. (2025). Computationally designed multi-epitope vaccine constructs targeting the SARS-CoV-2 spike protein elicit robust immune responses in silico.

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

Peeples, L. (2021). News Feature: Avoiding pitfalls in COVID-19 vaccine design.

Crotty, S., Weiskopf, D., Sette, A. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCIIpan-4.0: Improved pan-specific HLA class II antigen presentation prediction.

Cox, R. J., Brokstad, K. A., & Ogra, P. (2021). Influenza virus: Immunity and vaccination strategies.

Li, C., Shi, Y., & Yang, J. (2022). Computational analysis of cytokine-inducing epitopes for viral vaccines.

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design.

Sallusto, F., Geginat, J., & Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: Function, generation, and maintenance.

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes.

Sette, A., & Crotty, S. (2021). Adaptive immunity to SARS-CoV-2 and COVID-19.

Ayyagari, V., Gupta, S., & Mohan, S. (2022). Design of multi-epitope vaccines using immunoinformatics: Linking epitopes for enhanced immunogenicity.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction, and analysis.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2008). The SWISS-MODEL repository and associated resources.

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models.

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions.

Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: A fully automated algorithm for protein–protein docking.

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.

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

Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020).

Sarvmeili, J., Baghban Kohnehrouz, B., Gholizadeh, A., Shanehbandi, D., & Ofoghi, H. (2024). Immunoinformatics design of a structural protein-driven multi-epitope candidate vaccine against different SARS-CoV-2 variants based on fynomer.

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design.

Ayyagari, V., Gupta, S., & Mohan, S. (2022). Design of multi-epitope vaccines using immunoinformatics: Linking epitopes for enhanced immunogenicity.

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., & Baumert, J. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity.

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., & Wilson, M. (2015). T3DB: The toxic exposome database.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server.

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models.

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions.

Kiflu, T. A. (2024). Computational approaches for vaccine design and molecular docking.

Tombari, R., Sethi, A., & Kumar, V. (2025). Molecular docking and structural analysis of multi-epitope vaccine candidates against viral pathogens

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.

Ahmed, T., Al-Khater, K., & Elhassan, G. (2022). Immunoinformatics-based design of multi-epitope vaccines against SARS-CoV-2.

Farid, M., Hossain, M., & Rahman, S. (2023). Multi-epitope vaccine design targeting SARS-CoV-2 spike and nucleocapsid proteins using computational approaches.

Mukherjee, S., Sinha, S., & Ray, P. (2022). Immunoinformatics-guided design of a multi-epitope subunit vaccine against SARS-CoV-2.