In Silico Construction and Evaluation of a Potential Multi-Epitope Subunit Vaccine Constructed Against Rabies Virus

By

Ridita
21146048

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
November 2025

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing my 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 main sources of help.

Student's Full Name & Signature:

_____

Ridita
21146048

Approval

The thesis titled "*In Silico* Construction and Evaluation of a Potential Multi-Epitope Subunit Vaccine Against Rabies Virus" submitted by Ridita (ID: 21146048) of Summer, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on November, 2025.

Supervised By:

Supervisor:

_____

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

Approved by:

Chairperson:

_____

Dr. Raushanara Akter
Professor and Acting Chairperson
School of Pharmacy
BRAC University

Dean:

_____

Professor A F M Yusuf Haider, PhD
Acting Dean, School of Pharmacy
BRAC University

Ethics Statement

The completion of this thesis was achieved by adhering to all the relevant ethical norms and standards. Neither animal nor human trials were conducted or involved in this project.

Abstract

Rabies is a lethal zoonotic disease which is mainly contracted through the bite of rabid animals, especially canines. In Asian countries including Bangladesh rabies is still an endemic listed by the World Health Organization which claims 59,000 lives annually. The current treatment mainly includes multiple dose post exposure prophylaxis vaccine or rabies immunoglobulin (RIG) transfusion with no available medications. Our research centers on computational approaches to design a multi-epitope vaccine in-silico as a preventive measure to combat Rabies. The vaccine construct included cytotoxic T lymphocytes, helper T lymphocytes and B cell lymphocytes epitopes which were screened for antigenicity, allergenicity and toxicity and universally acknowledged linkers were used to conjoin them. The final construct displayed a significant antigenicity increase and was further investigated for acceptable physico-chemical properties and structural validation using Ramachandran and Z-score analysis. Evaluation of our construct's interaction with immune receptors was done through molecular docking with TLR4. Finally, immune simulation was conducted computationally, which assured that our construct was efficacious enough to trigger cellular and humoral responses while offering long-term immunity against this virus.

Keywords: Rabies; Post-exposure prophylaxis; Pre-exposure prophylaxis; In Silico Vaccine; Molecular Docking; Immune Simulation

Dedication

I would like to thank my family for their constant support and reassurance, especially my siblings Rose, Rabbi and Rafi because they inspire me each day to keep going.

Acknowledgement

With this opportunity, I would like to genuinely thank my supervisor, Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, BRAC University, for his constant guidance, critical insights and unwavering support throughout my entire process of completing this thesis project. His guidance and advice have truly encouraged me to go on with this research, making it an immensely enlightening and fulfilling experience of my academic life.

Table of Contents

List of Tables

List of Figures

List of Acronyms

B Cell	B Lymphocyte
CTL	Cytotoxic T Lymphocyte ($CD8^+$ T cells)
FASTA	Fast Alignment
GRAVY	Grand Average of Hydropathicity
HTL	Helper T Lymphocyte ($CD4^+$ T cells)
IEDB	Immune Epitope Database
IFN-γ	Interferon Gamma
KDa	Kilo Dalton
MHC	Major Histocompatibility Complex
PDB	Protein Data Bank
PEP	Post-Exposure Prophylaxis
PrEP	Pre-Exposure Prophylaxis
RABV	Rabies Virus

Chapter 1

Introduction

Rabies virus is an extremely severe zoonotic pathogen which can quickly escalate to lethal encephalitis with a 100% fatality rate in symptomatic individuals who are not vaccinated (Alemayehu et al., 2024). According to the World Health Organization (WHO), Bangladesh is classified as one of the top-most countries (third) where rabies is still an endemic. An estimated 59,000 lives are claimed by rabies every year in more than 150 countries, with Asia and Africa having 95% of these cases where 40% of them are children under 15 years of age (Alemayehu et al., 2024). Even though, rabies is seriously deadly and almost always fatal once symptoms appear, it is still an alarmingly under-reported disease and falls in category for "Neglected Tropical Diseases" as stated by World Health Organization in 2022 (Alemayehu et al., 2024). 99% Transmission of rabies virus mainly occurs via saliva from bites or scratches of rabid dogs, among other animals making "Canine Rabies" the major cause behind global human rabies infection and death (Dürr et al., 2008). Bangladesh is immensely endangered, yet people neglect this disease and its symptoms due to lack of awareness, cost and time in efficiency of treatment, lack of treatment facility and multiple-dose vaccine regimen (Hampson et al., 2015). The Post Exposure Prophylaxis (PEP) regimen that are currently used as treatments against rabies are washing the infected wound, administration of polyclonal antibody and multiple doses of PEP rabies vaccine after being bitten. These inactivated vaccines do not guarantee lifelong protection and neutralizing antibody level usually starts to diminish after 1 - 5 years of application. No medications or antibiotics work to cure this disease as of yet (Odhar et al., 2023). After exposure through rabid animal saliva, RABV keeps replicating in muscles then moves towards the spinal cord and the brain through the neurons which are peripheral. After reaching the brain, it continues to spread in more tissues like the salivary gland. Consequently, inevitable death occurs after major neuronal dysfunction (Odhar et al., 2023). The symptoms

of this neuroinvasive and neurotropic virus includes agitation, dysphagia leading to hydrophobia, aerophobia, spasms of muscles, paralysis partially or completely and if not treated in time, ultimately death. Despite being this aggressive of a disease, proper treatment and vaccination is not timely or economically available in a developing country like Bangladesh (Dürr et al., 2008).

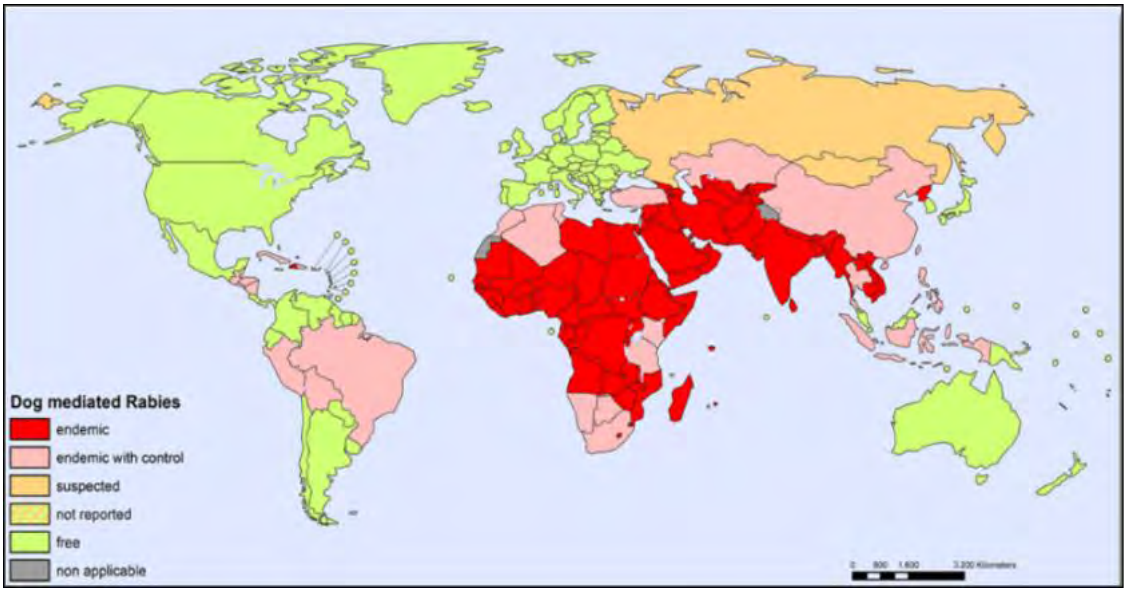


Figure 1: Rabies Endemic Regions (Occurrence of Rabies | Rabies - Bulletin - Europe, n.d.)

1.1 Genomic Structure of Rabies Virus

Rabies, the negative sense single stranded RNA virus belongs to the Rhabdoviridae family and is under the genus *Lyssavirus*, and order *Mononegavirales* (Walker et al., 2018). It portrays an enveloped bullet shaped virion and is about 75 nm diameters and 180 nm length-wise. The viral genome is about 12 kb long and includes five essential structural proteins namely- the nucleoprotein or N protein, glycoprotein (G), phosphoprotein (P), matrix protein (M) and lastly the large polymerase (L) encoding its genomic structure (Khalifa et al., 2021). All of these are functionally important and amongst these, the glycoprotein (G) aids in viral entry, while the nucleoprotein (N) has an indispensable role in viral replication and transcription by encapsidation of the viral RNA genome (Mehta et al., 2014). The matrix protein (M) aids in

budding as well as, virion assembly (Wang & Xing, 2024). Remaining proteins like the phosphoprotein (P) and large protein (L) operates as essential cofactors of RNA-dependent RNA polymerase to further aid the viral genome in their replication and transcription processes. The RABV nucleocapsid composed of single-stranded RNA are coated by nucleoprotein (N) is a helical structure (Wang & Xing, 2024). Inside the virus's bullet-shaped envelope its located and linked with other proteins like the phosphoprotein (P) and large polymerase (L) proteins. Fusion of these create the ribonucleoprotein (RNP) core. This particular interaction and co-ordination among the structural elements of RABV determines the virus life cycle, which includes host-cell attachment and entry, replication, assembly and budding which, consequently, results in severe neuropathology in the infected person (Tombari et al., 2025).

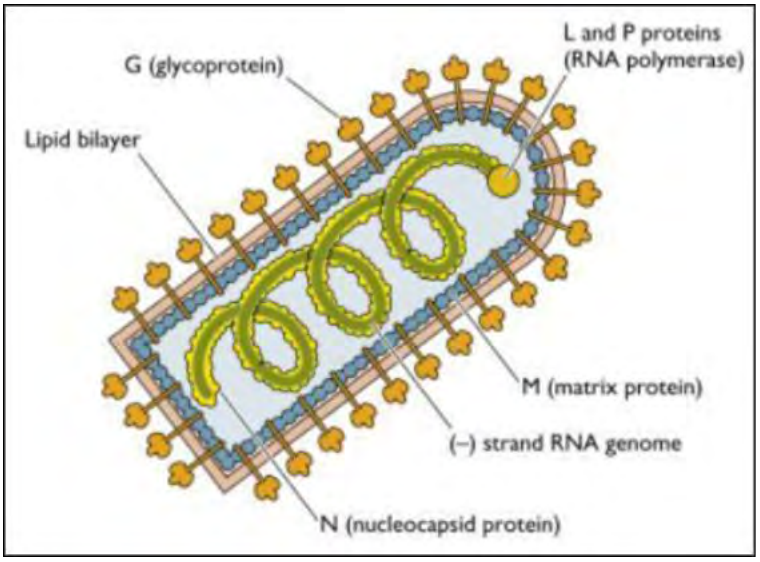


Figure 2: Genomic structure of Rabies Virus denoting all structural proteins (Ge et al., 2010)

1.2 Lifecycle of Rabies Virus

The viral lifecycle of RABV begins after it uses its G protein to bind to host neuron receptors like nicotinic acetylcholine receptor and endocytosis occurs, letting the viral entry into the cell. There the virus is engulfed and turned into an endosome. The fusion of viral envelopes with endosomal membrane occurs, releasing the viral RNA genome into host cytoplasm. The viral

RNA polymerase or L protein transcribes genomes into messenger RNAs which are then translated by host ribosomes into viral structural proteins like N, G, P, M and L proteins (Khalifa et al., 2021). The N protein's function is to encapsidate and protect the viral RNA and form nucleocapsid core which is quintessential for further viral replication. Full length genomes are synthesized, and they assemble towards the plasma membrane where virions are released from the host cell. These new and matured virions keep infecting the neighbouring neurons and moves towards CNS (Jadeja & Vanak, 2023).

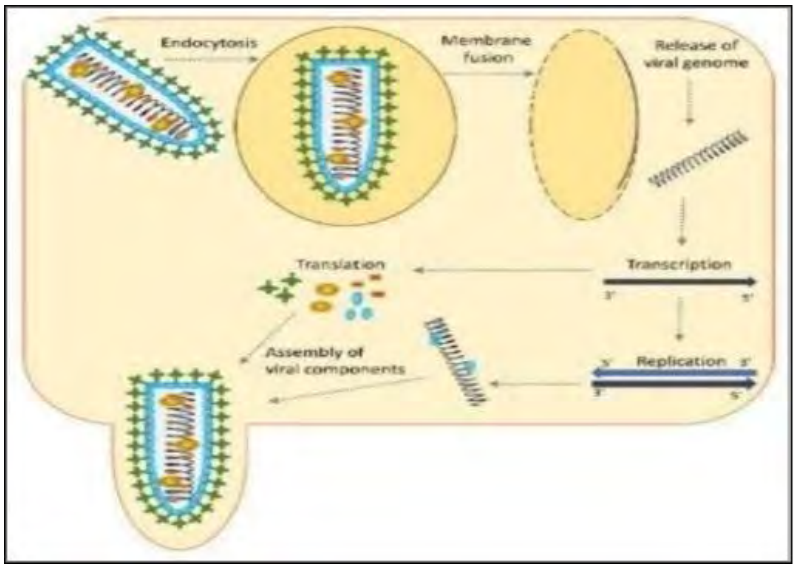


Figure 3: Viral Lifecycle of Rabies(Jadeja & Vanak, 2023)

1.3 Pathogenesis of Rabies Virus

Rabies virus predominantly targets the peripheral and central nervous system and mainly enters through the nicotinic acetylcholine receptors and continues to spread along the axons towards the brain. As a result, it can cause fatal encephalitis in the host if it is left untreated. The way the disease develops in the host is after being bitten or scratched deep by the rabid animal, mostly dogs. The saliva containing the virus comes in contact with the host's broken skin or mucous membranes right after the bite occurs, and that is how the viral entry takes place

(Bastos et al., 2023). The RABV, at first, starts to replicate locally at the site of muscle cells where the host was bitten. The viral G protein binds to the nicotinic acetylcholine receptor and neural cell adhesion molecules located at the neuromuscular junction. This leads to further attachment and uptake right into the motor neurons. The N protein encapidates the rabies viral genome and protects it from cellular enzymes (Kiflu, 2024). It also acts as an immune system antagonist as it blocks any antiviral responses which promotes further replication. Inside the neurons, this virus gets enclosed within a vesicle and then goes through retrograde axonal transport towards the central nervous system via the peripheral nervous system. When it ultimately reaches to spinal cord and brain, viral replication takes place which leads to inflammation and dysfunction of neurons and causes severe symptoms like hydrophobia (fear of water), aggression and agitation, paralysis and can eventually become fatal. After that, the virus keeps distributing further away from the CNS to extranural tissues like the salivary gland which could allow virus filled saliva aiding in further transmission of the disease (Bastos et al., 2023).

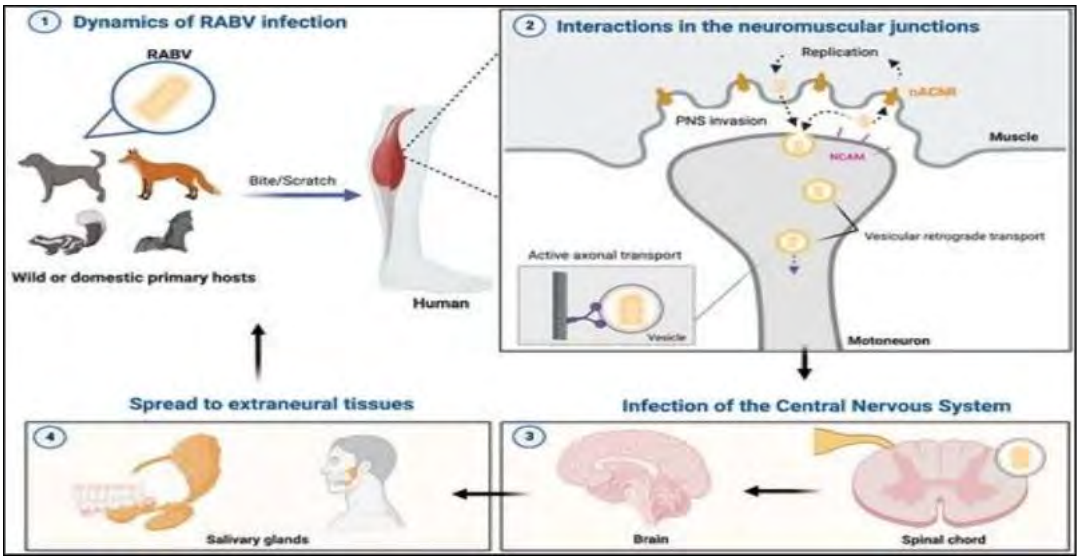


Figure 4: Rabies Virus Pathogenesis how the infection spreads from rabid animal to host body through neuromuscular junction leading to central nervous system and spreading to more extraneural tissues(Bastos et al., 2023).

Chapter 2

Methodology

The steps that were serially employed during the construction of our in-silico multi-epitope vaccine are schematically shown, as follows:

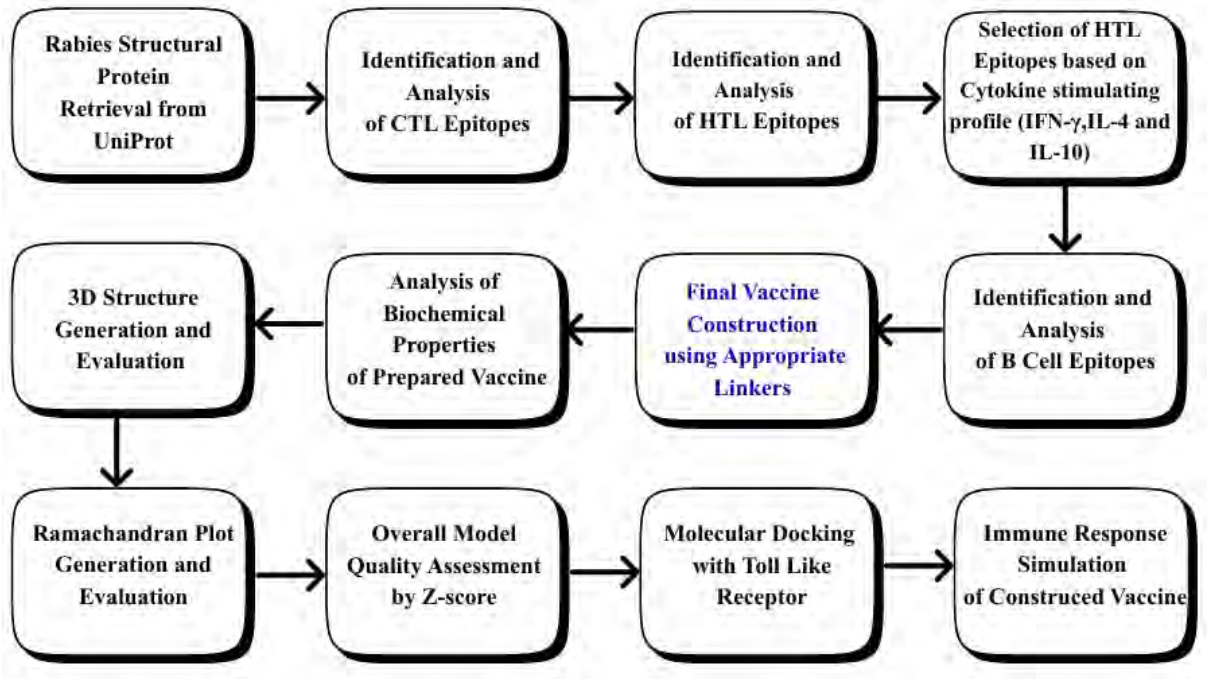


Figure 5: Approaches entailing the in-silico vaccine construction method using various computational tools.

2.1 Suitable Protein Sequence Selection

For the purpose of choosing the protein, we used the Uniprot protein database which is an extensive database that provides well attributed and high quality sequences of protein. Uniprot currently displays 106 Swiss-Prot reviewed proteins when searched for rabies *(Uniprot*, n.d.*)*. The rabies protein that we chose was retrieved in FASTA (canonical) format. We took the sequence which had 450 amino acids. After selection we submitted our protein in Vaxijen 2.0 tool to check whether it was antigenic enough or not. Since we are working with rabies, we

selected “virus” as our target organism and the threshold was set to 0.5. This threshold was set considering vaxijen’s ability to become more stringent while detecting antigenicity so it does not give false positives and give us only those proteins that should be studied in detail while working with in-silico vaccine construction process (Doytchinova & Flower, 2007).

2.2 Identification of Cytotoxic T Lymphocyte (CTL) Epitopes

CTLs or cytotoxic T lymphocytes which are also called $CD8^+$ T cells detect antigens that are displayed on the MHC class I molecules. Upon detection, they impair host cells that are infected by destroying them, hence triggering immune activation. For identification of CTL epitopes, NetCTL 1.2 tool was initially used. This tool comprises of three important factors which are TAP transport efficiency, efficiency of proteasomal C-terminal cleavage and binding affinity of MHC-I class (Stranzl et al., 2010). Then CTL epitopes were found after providing our protein sequence in the format of FASTA where epitope identification threshold was set to 0.75 as stated by and combined score was selected which ascertained only the most relevant and usable epitopes (Larsen et al., 2007).

2.3 Analysis of CTL Epitopes

NetMHCpan-4.1 tool was equipped to additionally validate the predicted epitopes. Over a comprehensive range of HLA class I molecules, this tool gives us vigorous pan-allele based predictions. It evaluates peptides of many ranges, but we focused on 9-mer ones. For the purpose of getting Strong Binders a threshold of 0.5 was kept, 2 was kept in the event of weak binders and Binding Affinity (BA) predictions were specifically included (Reynisson et al., 2020). This led to an output where peptides were arranged and sorted based on their scores of binding affinity. Epitopes which met the set threshold conditions and were denoted as “SB” or strong binders were further selected for evaluation. The CTL epitopes that were chosen required to be antigenic in nature, which was determined by vaxijen 2.0 tool again having 0.5

as the threshold. To ensure that these were neither allergenic nor toxic, we used the computational tools named AllerTOP v2.1 and ToxinPred, respectively (Gupta et al., 2013).

2.4 Identification of Helper T Lymphocyte (HTL) Epitopes

Helper T lymphocyte or HTL epitopes primarily function by activating and coordinating other important immune cells like CTLs and B cells through cytokine release. It is also known as $CD4^+$ T cell and can mediate humoral response as it aids to activate B cell's antibody production. Epitope prediction was done as we utilized the tool named NetMHCIIpan-4.0 and 15-mer peptides were taken. We took the first 20 alleles against our protein sequence and then took the rest of the 14 alleles, covering all the allele based predictions (Reynisson et al., 2020). The threshold of strong binders was kept set at 1%, for weak binders it was 5% and binding affinity predictions were considered. Peptide scanning was performed through running those against HLA-DR/DP/DQ alleles. Then SB designated peptides were chosen and further ran for antigenicity, allergenicity and toxicity analysis using the same tools as used for CTL (Reynisson et al., 2020).

2.5 Analysis of HTL Epitopes

Once we confirm that our selected epitopes are antigen, non-allergen and non-toxin. We need to assess those further for cytokine response induction which are highly essential for immune response generation and amplification. IFN-γ, IL-4 and IL-10 are the cytokine releases that were considered for our project which are known as interferon gamma, interleukin 4 and interleukin 10 respectively (Capsoni et al., 1995). IFN-γ was detected with the server IFN epitope and only the positives were taken since interferon is necessary for macrophage activation and Th1-type immune response. Further evaluation of IL-4 and IL-10 was performed with IL4Pred and IL10Pred and we took the epitopes which induce both of these. Interleukin

4 stimulate antibody production and Th2-type response and interleukin 10 is vital for being the anti inflammatory cytokine (Capsoni et al., 1995).

2.6 Identification of B cell Epitopes and Analysis

B cell makes antibodies and forms memory B cells which are required for long term humoral immunity and that is exactly why we need it for our constructed vaccine. For achieving b cell epitopes we use IEDB Analysis Resource where we specifically choose predicitive tools under b cell aiming for linear epitopes from our original sequence. Advanced tool “Bepipred Linear Prediction 2.0” was chosen (Jespersen et al., 2017). The result we got showed us an antigenicity based graph and peptide length based data table where 7 to 30 amino acids length of peptides was taken into consideration. The selected B cell epitopes were further detected for antigenicity, allergenicity and toxicity using the same computational tools which are Vaxijen 2.0, AllerTOP v2.1 and ToxinPred respectively.

2.7 Final Vaccine Construction with Appropriate Linkers

Our multi epitope vaccine was finally constructed in-silico after following a very distinct construction pattern where our adjuvant which was our primary sequence was paired with EAAAK linker to add CTL for rigidity, GPGPG linker was chosen after this to join our HTL for flexibility and enhanced processing and KK linker was used to connect B cell epitopes which increased solubility (Ayyagari et al., 2022). Our followed process is schematically represented below.

Figure 6: Application of Linkers with Epitopes in Final Vaccine Construction

2.8 Constructed Vaccine's Evaluation for its Antigenicity, Allergenicity and Toxicity

After the vaccine has been constructed with linkers we must again check for its antigenicity and notice if the antigenicity has increased significantly from the previous protein sequence's antigenicity. This was again done by the tool Vaxijen 2.0 where the threshold was set to 0.5 and target organism to "virus". Furthermore, we evaluated it's allergenicity using "AllergenOnline" server and non toxicity was ensured by T3DB server. This whole process ensures us that our prepared vaccine is immunogenic.

2.9 Constructed Vaccine's Evaluation for Biochemical Properties

ProtParam tool is utilized for computing the biochemical or physicochemical parameters of our constructed vaccine. It calculates significant aspects of our construction like the number of its amino acids, theoretical pI or the iso electric point, molecular weight in kDa (kilodalton) unit and extensive amino acid composition in detail (Wilkins et al., 1999). It also tells us atomic composition, formula and total atom numbers. The half-life is also computationally estimated. The most important information this tool provides us is it's estimated instability index and GRAVY value which stands for grand average of hydropathicity. The instability index predicts whether our will be stable in-vitro or not and indicates a protein as stable when it is computed to be any value less than 40. GRAVY is the number which dictates whether our construct is hydrophilic or hydrophobic where a positive gravy indicates hydrophobicity and negative GRAVY means our construct is hydrophilic and soluble (Kyte & Doolittle, 1982). In addition to that our construct's thermostability by regarding the volume of aliphatic side chain is indicated by the aliphatic index where more than 60 value is preferred.

2.10 Constructed Vaccine's 3D Structure Generation

For predicting our vaccine construct's 3D image model, we implemented the server known as Phyre 2.2. This generates 3D structure of a provided amino acid sequence by applying a method called homology model where it finds an already known structure from its comprehensive database that looks similar and generates a model using that as a template (Kelley et al., 2015). Along with structure creation it gives us two significant parameters which are known as the confidence percentage and coverage percentage. Structural reliability is highly ensured when the confidence is 100% and coverage is 80% or more. The 3D structure is generated in a PDB format. We input our vaccine sequence and got a PDB formatted structure for it that was envisioned through PyMOL software (Kelley et al., 2015).

2.11 Constructed Vaccine's Ramachandran Plot Generation and Validation

For quality evaluation for our predicted model by Phyre 2.2, we further decide to implement SwissModel Expasy where we upload the PDB structure of Phyre which gives us Ramachandran plots against phi (φ) and psi (ψ) which are the allowed angles for the generated model as stated by Keifer et al. (2008). Good steoreochemistry and conformation is implied when most of the model's residues fall in the regions which are ramachandran favoured and further evaluates less unstable bond angles. 90-95% or more ramachandran favoured score is the goal where outliers (areas requiring refinement) are supposed to be less than 2%.

2.12 Constructed Vaccine's Z-score Assessment

Ascertainment of our vaccine design was done again by generating a Z-score analysis with the use of the tool ProSa Web. Our PDB structure is compared against experimentally validated structures of protein. It also checks the energy distribution and deviation of model provided and certifies final quality. A graph is generated against residue number and knowledge based position of sequence versus energy which is crucial to assess model's overall quality. It also

tells us if our structure falls in the X-ray or the NMR region (Wiederstein & Sippl, 2007). A z-score of - 5 to -10 is desired.

2.13 Analysis of Constructed Vaccine's Molecular Docking

To determine the extent of our created vaccine construct and immunological receptors (toll like receptors) binding affinity the server called ClusPro was chosen to generate the molecular docking process. For the ligand molecule section of the Cluspro we put in the PDB structure we got from Phyre 2.2 and for the receptor molecule section, Toll like receptor (TLR4) was sourced from the comprehensive protein databank titled RCSB (Reynisson et al., 2020). Output result showed us figures of docked complex models and their scores could also be viewed. The model which we acquired had the highest binding energy scoring and was taken for further study (Comeau et al., 2004).

2.14 Simulation of Immune Response of the Constructed Vaccine

C-ImmSim 10.1 web server was implemented as the concluding step, which utilizes its algorithms to simulate how our constructed vaccine would act in a functional immune system. It also gives us an idea about the vaccine's immunogenicity as it simulates the humoral and cellular reactions of the vaccine we designed (Gupta et al., 2013). Our construct was input in FASTA format thrice and each time steps were modified with the first injection's time step being 1, second injection's time step being 84 and third injection's time step being 168 where each step upholds eight hours. The result generated by C-ImmSim showed us numerous graphs regarding antigen and immune response, concentration of antibody, B cell populations, $CD4^+$ and $CD8^+$ cell, cytokine and interleukin response, and last but not the least gives a summary graph combining antibody titers and memory cell formation and response (Rapin et al., 2010). It also generates separate graphs highlighting the cytotoxic T cell population, helper T cell population and B lymphocyte population which tells us how strongly the cellular and humoral

responses are being upheld and if our vaccine immunity is longer lasting against the disease or not. The concentration of cytokine and interleukin graph entails the induction of interferon gamma along with all the necessary interleukin induction like interleukin- 4, 10,2 ad 12. When we applied our vaccine sequences thrice in FASTA format C-ImmSim predicted how our created vaccine of three doses would behave in a biological system and it passed every parameter within good and acceptable results.

2.15 Considerations and Reflection on Methodology

Our rabies vaccine design is prepared through computationally predicting and filtering various aspects and validated successfully through in silico method ensuring reliability. Only properly reviewed and annotated protein sequence was considered to avoid errors. We used an amalgamation of various widely used immunoinformatic tools and servers which are validated in numerous literatures which ascertains that these aid to identify potential vaccine candidates through efficient means. Even though these are well acclaimed because of their capability to model and evaluate molecular interactions related to antigen processing and presentation using computational techniques, it may be insufficient due to immune response complexities in various host dynamics. On that account, our study exclusively concentrates on proposing a pre-exposure prophylaxis vaccine designed against the rabies virus which should be investigated and validated further. Hence, this study only forefronts a prospective and potential vaccine proposal against this deadly rabies and further and proper in-vitro and in-vivo tests are encouraged to be conducted to ensure its proper efficacy, safety and stability.

Chapter 3

Results

3.1 Protein Sequence Selection based on Antigenicity Check

By accessing the broad protein sequence resource, Uniprot, we chose our Rabies protein (Uniprot ID: P16285) sequence in FASTA format. It was picked because of demonstrating antigenicity in vaxijen 2.0 amid our initial screening. The protein sequence is also displayed below.

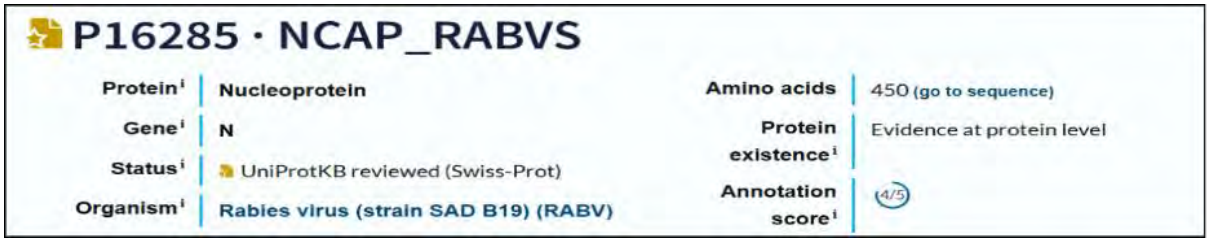


Figure 7: Rabies protein retrieval from UniProt (UniProt, n.d.)

Retrieved sequence:

MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPCITLGKAPDLNKAYKSVL
SGMSAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIARKGDKITPGSLVEIKRT
DVEGNWALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKTNIADRIEQ
IFETAPFVKIVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGTV
VTAYEDCSGLVSFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAVPHSYFIH
FRSLGLSGKSPYSSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSVLGGY
LGEEFFGKGTFERRFFRDEKELQEYEAAELTKTDVALADDGTVNSDDEDYFSGETRS
PEAVYTRIMMNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYSSDS

Vaxijen 2.0 is an essential web based tool which anticipates whether protein sequences are antigenic or not based on the threshold and organism that has been set. For our initial

antigenicity screening our chosen protein passed as it passed as antigenic when the 0.5 threshold was applied and the target organism we selected for this is “virus” as we are working with RABV. The antigenicity our chosen sequence displayed was 0.5010.

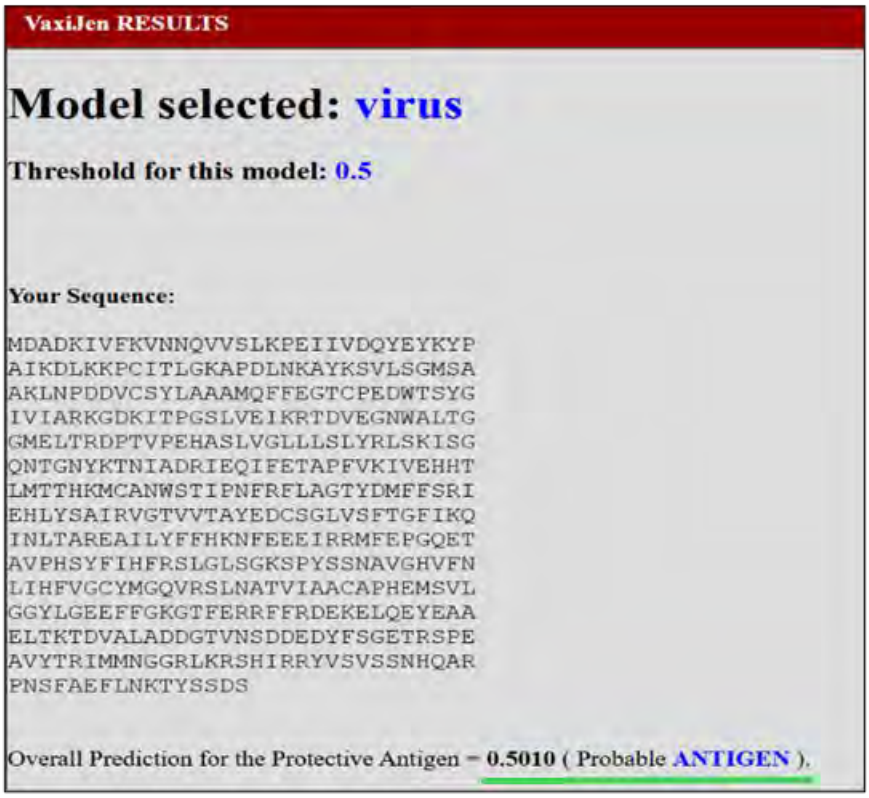
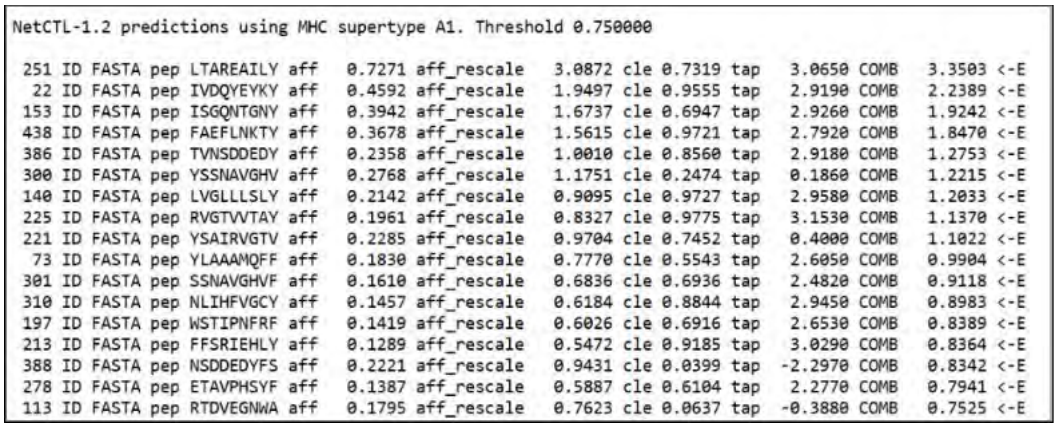


Figure 8: Initial Antigenicity of protein with 0.5 threshold application in Vaxijen 2.0 (Doytchinova & Flower, 2007).

3.2 Identifying and Evaluating CTL Epitopes

At first we implemented NetCTL 1.2 tool to select CTL epitopes as it helps to precisely predict these sequences considering different criterias. For MHC class I selection A1 supertype was selected along with C Terminal cleavage weight being 0.15, TAP transport efficiency weight being 0.05 and lastly epitope identification threshold being 0.75 (Larsen et al., 2007). We

further manually selected “combined score” on sort by score option which generated us with results that have seventeen epitopes with “<-E” assigned to them. This “<-E” represents best predicted epitopes from our provided sequence along with their combined scores. These were accounted for further evaluation.



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

 251 ID FASTA pep LTAREAILY aff   0.7271 aff_rescale   3.0872 cle 0.7319 tap   3.0650 COMB   3.3503 <-E
  22 ID FASTA pep IVDQYEYKY aff   0.4592 aff_rescale   1.9497 cle 0.9555 tap   2.9190 COMB   2.2389 <-E
 153 ID FASTA pep ISGQNTGNY aff   0.3942 aff_rescale   1.6737 cle 0.6947 tap   2.9260 COMB   1.9242 <-E
 438 ID FASTA pep FAEFLNKTY aff   0.3678 aff_rescale   1.5615 cle 0.9721 tap   2.7920 COMB   1.8470 <-E
 386 ID FASTA pep TVNSDDEDY aff   0.2358 aff_rescale   1.0010 cle 0.8560 tap   2.9180 COMB   1.2753 <-E
 300 ID FASTA pep YSSNAVGHV aff   0.2768 aff_rescale   1.1751 cle 0.2474 tap   0.1860 COMB   1.2215 <-E
 140 ID FASTA pep LVGLLLSLY aff   0.2142 aff_rescale   0.9095 cle 0.9727 tap   2.9580 COMB   1.2033 <-E
 225 ID FASTA pep RVGTVVTAY aff   0.1961 aff_rescale   0.8327 cle 0.9775 tap   3.1530 COMB   1.1370 <-E
 221 ID FASTA pep YSAIRVGTV aff   0.2285 aff_rescale   0.9704 cle 0.7452 tap   0.4000 COMB   1.1022 <-E
  73 ID FASTA pep YLAAAMQFF aff   0.1830 aff_rescale   0.7770 cle 0.5543 tap   2.6050 COMB   0.9904 <-E
 301 ID FASTA pep SSNAVGHVF aff   0.1610 aff_rescale   0.6836 cle 0.6936 tap   2.4820 COMB   0.9118 <-E
 310 ID FASTA pep NLIHFVGCY aff   0.1457 aff_rescale   0.6184 cle 0.8844 tap   2.9450 COMB   0.8983 <-E
 197 ID FASTA pep WSTIPNFRF aff   0.1419 aff_rescale   0.6026 cle 0.6916 tap   2.6530 COMB   0.8389 <-E
 213 ID FASTA pep FFSRIEHLY aff   0.1289 aff_rescale   0.5472 cle 0.9185 tap   3.0290 COMB   0.8364 <-E
 388 ID FASTA pep NSDDEDYFS aff   0.2221 aff_rescale   0.9431 cle 0.0399 tap  -2.2970 COMB   0.8342 <-E
 278 ID FASTA pep ETAVPHSYF aff   0.1387 aff_rescale   0.5887 cle 0.6104 tap   2.2770 COMB   0.7941 <-E
 113 ID FASTA pep RTDVEGNWA aff   0.1795 aff_rescale   0.7623 cle 0.0637 tap  -0.3880 COMB   0.7525 <-E
```

Figure 9: CTL epitopes generated by NetCTL1.2 (Larsen et al., 2007)

These were then arranged sequentially and placed in NetMHCpan 4.1 immunoinformatic tool to find out compatible binding alleles and gave us 9-mer peptide sequences across every HLA alleles of MHC I class. The threshold for strong binders and weak binders was kept to 0.5 and 2 respectively and binding affinity prediction was selected. The output gave us eleven CTL epitopes after removing redundancy of recurring epitopes.

CTL Epitopes	Antigenicity	Allergenicity	Toxicity
LTAREAILY	Non-Antigen	Non-Allergen	Non-Toxin
IVDQYEYKY	Antigen	Allergen	Non-Toxin
ISGQNTGNY	Antigen	Non-Allergen	Non-Toxin
FAEFLNKTY	Non-Antigen	Non-Allergen	Non-Toxin
TVNSDDEDY	Antigen	Allergen	Non-Toxin
RVGTVVTAY	Antigen	Allergen	Non-Toxin
YLAAAMQFF	Non-Antigen	Non-Allergen	Non-Toxin
NLIHFVGCY	Non-Antigen	Allergen	Non-Toxin
ETAVPHSYF	Non-Antigen	Non-Allergen	Non-Toxin
SSNAVGHVF	Non-Antigen	Non-Allergen	Non-Toxin
WSTIPNFRF	Antigen	Non-Allergen	Non-Toxin

Table 1: CTL Epitope Selection based on Antigenicity, Allergenicity and Toxicity filtering

From the table 1, we can see the assessment of our 9-mer epitopes based on antigenicity using Vaxijen 2.0, allergenicity using AllerTOP v2.1, and toxicity with ToxinPred, all of which are essential bioinformatics tools. We only selected the ones which met the criteria of being antigenic, non-allergenic, and nontoxic. The selected epitopes also displayed a high combined score where ISGQNTGNY had 2.9190 and WSTIPNFRF had 2.6530 respectively.

3.3 Identifying and Evaluating HTL Epitopes

Helper T Lymphocytes not only recognizes antigen displayed on MHC CLASS II molecules but also play an indispensable role in immune response activation by cytokine release like interferon gamma (macrophage activation and antigen presentation), interleukin-4 (induce B cell response promotion) and interleukin-10 (inflammation regulation). NetMHC-IIpan 4.0 tool was employed for achieving HTL epitopes across all alleles where threshold for strong binder was kept 1 (% Rank) and weak binders had 5 (% Rank) and we also incorporated “BA” or binding affinity predictions. The generated results showed us a vast list of 15-mer peptides out of which we chose the “<=SB” designated ones after omitting redundancy. After taking the epitopes, we further screened them for antigenicity, allergenicity and toxicity along with ensuring whether they were interferon gamma positive and IL-4 and IL-10 inducer or not.

Epitopes	Antigenicity	Allergenicity	Toxicity	IFN-γ	IL-4	IL-10
VNNQVVSLKPEIIVD	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-inducer
NNQVVSLKPEIIVDQ	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
NQVVSLKPEIIVDQY	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
YKSVLSGMSAAKLNP	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KSVLSGMSAAKLNPD	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KVNNQVVSLKPEIIV	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer

AYKSVLSGMSAAKLN	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
SVLSGMSAAKLNPDD	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
CYMGQVRSLNATVIA	Antigen	Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
MGQVRSLNATVIAAC	Antigen	Non-Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
TSYGIVIARKGDKIT	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
PHSYFIHFRSLGLSG	Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
HSYFIHFRSLGLSGK	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Inducer
SYFIHFRSLGLSGKS	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
RRYVSVSSNHQARPN	Antigen	Non-Allergen	Non-toxin	Positive	Non-Inducer	Inducer
KPEIIVDQYEYKYPA	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Non-Inducer
GGMELTRDPTVPEHA	Non-Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
TFERRFFRDEKELQE	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
FERRFFRDEKELQEY	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
ERRFFRDEKELQEYE	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
RRFFRDEKELQEYEA	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
NKAYKSVLSGMSAAK	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
VSFTGFIKQINLTAR	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
TGFIKQINLTAREAI	Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
DVALADDGTVNSDDE	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
FVKIVEHHTLMTTHK	Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Non-Inducer
LKRSHIRRYVSVSSN	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
KRSHIRRYVSVSSNH	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Non-Inducer
DVEGNWALTGGMELT	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Inducer

VEGNWALTGGMELTR	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
EGNWALTGGMELTRD	Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
GNWALTGGMELTRDP	Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
EDYFSGETRSPEAVY	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
DKIVFKVNNQVVSLK	Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer
KIVFKVNNQVVSLKP	Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Non-Inducer
REAILYFFHKNFEEE	Non-Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Inducer
EAILYFFHKNFEEEI	Non-Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-inducer
NTGNYKTNIADRIEQ	Non-Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Non-Inducer
TGNYKTNIADRIEQI	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
SYGIVIARKGDKITP	Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
TPGSLVEIKRTDVEG	Antigen	Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
PGSLVEIKRTDVEGN	Antigen	Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
LKPEIIVDQYEYKYP	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
DYFSGETRSPEAVYT	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Inducer
KITPGSLVEIKRTDV	Antigen	Allergen	Non-toxin	Positive	Inducer	Non-Inducer
ITPGSLVEIKRTDVE	Antigen	Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
GSLVEIKRTDVEGNW	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
TAPFVKIVEHHTLMT	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
APFVKIVEHHTLMTT	Non-Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
GFIKQINLTAREAIL	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
FIKQINLTAREAILY	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
IKQINLTAREAILYF	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer

KQINLTAREAILYFF	Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
DEKELQEYEAAELTK	Antigen	Allergen	Non-toxin	Positive	Non-Inducer	Non-Inducer
RDEKELQEYEAAELT	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
VPHSYFIHFRSLGLS	Antigen	Non-Allergen	Non-toxin	Positive	Inducer	Non-Inducer
GKSPYSSNAVGHVFN	Non-Antigen	Non-Allergen	Non-toxin	Negative	Non-Inducer	Inducer
KSPYSSNAVGHVFNL	Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Inducer
SPYSSNAVGHVFNLI	Non-Antigen	Allergen	Non-toxin	Negative	Non-Inducer	Inducer
VDQYEYKYPAIKDLK	Antigen	Allergen	Non-toxin	Negative	Inducer	Non-Inducer
FSRIEHLYSAIRVGT	Antigen	Non-Allergen	Non-toxin	Negative	Inducer	Inducer

Table 2: HTL Epitope selection based on Interferon gamma, Interleukin-4 and 10 inductions along with Antigenicity, Allergenicity and Toxicity test.

These HTL epitopes were filtered for antigenicity using Vaxijen 2.0 with threshold 0.5, allergenicity was checked by AllerTOP v2.1 and toxicity by ToxinPred. Additionally, these were screened for its ability to be IFN, IL-4 and IL-10 inducers through the tools IFN epitope server, IL4Pred and IL10Pred respectively. The epitopes which perfectly met these six parameters were chosen for further evaluation.

3.4 B Cell Epitope Prediction and Evaluation

B cells or B lymphocytes play the primary role in the generation of humoral immune response as they are involved in antibody production against pathogens. The epitopes of B cells from our sequence were achieved using IEDB Analysis Resource where we submitted our sequence and selected “Bepipred Linear Prediction 2.0” approach (Fleri et al., 2017). The result that we got included a graphical representation of a score versus position graph, where the yellow spikes indicated potential epitopes and green downward spikes indicated non-antigenic

epitopes. It also displayed a data box of predicted peptides from which we chose seven to thirty amino acid sequences as this is the optimal range for antibody binding and antigenic stability.

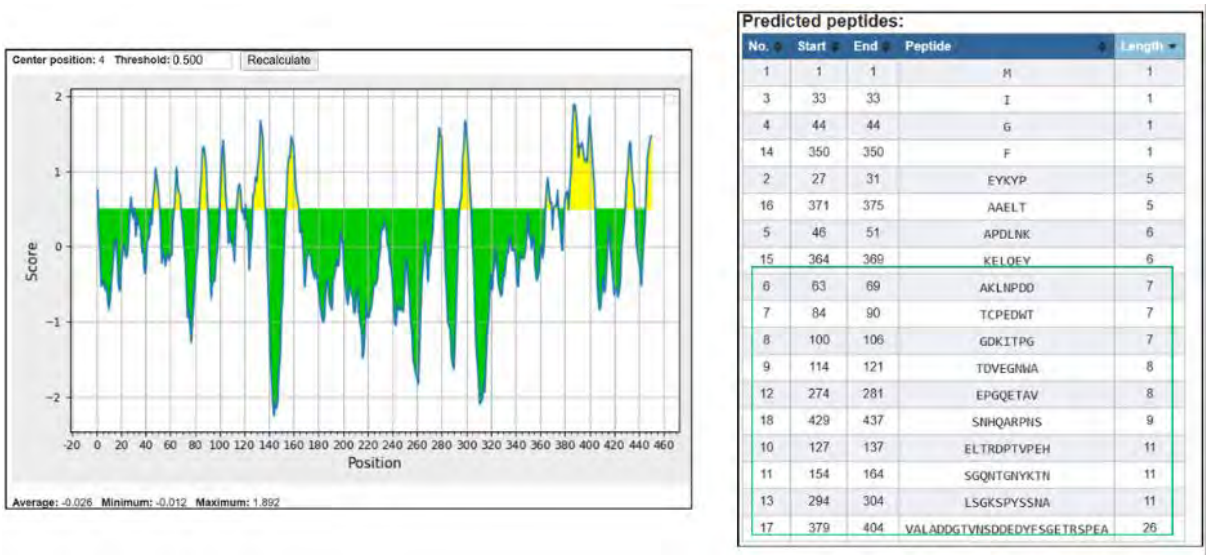


Figure 10: Score vs Position graph and achieved epitopes from IEDB Analysis Resource (Fleri et al., 2017).

Additional assessment was done for assuring antigenicity through Vaxijen 2.0, non-allergenicity through AllerTOP v2.1 and non-toxicity through ToxinPred and only those peptides were taken as B cell epitopes which met all three of these criteria which is shown in the tabular figure below.

EPITOPES	LENGTH	ANTIGENICITY	ALLERGENICITY	TOXICITY
AKLNPDD	7	Antigen	Non-allergen	Non-toxin
TCPEDWT	7	Non-Antigen	Allergen	Non-toxin
GDKITPG	7	Non-Antigen	Allergen	Non-toxin
TDVEGNWA	8	Antigen	Non-allergen	Non-toxin
EPGQETAV	8	Non-Antigen	Non-Allergen	Non-toxin
SNHQARPNS	9	Antigen	Non-allergen	Non-toxin
ELTRDPTVPEH	11	Non-Antigen	Non-Allergen	Non-toxin
SGQNTGNYKTN	11	Non-Antigen	Non-Allergen	Non-toxin
LSGKSPYSSNA	11	Antigen	Non-allergen	Non-toxin
VALADDGTVNSDDEDYFSGETRSPEA	26	Non-Antigen	Allergen	Non-toxin

Table 3: B cell epitope selection based on Length, Antigenicity, Allergenicity and Toxicity screening.

3.5 Final Vaccine Construction Using Appropriate Linker

After getting all the desired CTL, HTL and B cell epitopes which meet all the required criteria, we combined them with our primary sequence with appropriate linkers. Linkers are crucial when we are working with multi epitope vaccine construction as they aid the presentation of each epitope distinctly and retain their structure restoring immunogenicity. This ensures proper folding; hence, structural integrity can be preserved preventing steric hindrance among epitopes. We combined the CTL epitopes with "EAAAK" linker, which is known for its rigidity, HTL epitopes with the flexible linker "GPGPG" which efficiently promoted HTL presentation and lastly "KK" polar linker was used for connecting B cell epitopes which ensured solubility. All of these elements play a key role in making our vaccine construct more stable and effective immunologically (Ayyagari et al., 2022). We took the epitopes with the highest scoring antigenicity to boost our construct's antigenicity which led us to use one cytotoxic T lymphocyte epitope, one helper T lymphocyte epitope and two B lymphocyte epitopes. The final vaccine design is displayed below.

FINAL VACCINE CONSTRUCTION:

MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPCITLGKAPDLNKAYKSVL
SGMSAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIARKGDKITPGSLVEIKRT
DVEGNWALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKTNIADRIEQ
IFETAPFVKIVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGTV
VTAYEDCSGLVSFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAVPHSYFIH
FRSLGLSGKSPYSSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSVLGGY
LGEEFFGKGTFERRFFRDEKELQEYEAAELTKTDVALADDGTVNSDDEDYFSGETRS
PEAVYTRIMMNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYSSDSEAAAKWST
IPNFRFGPGPGHSYFIHFRSLGLSGKKKAKLNPDDKKTDVEGNWA

3.6 Assessment of Antigenicity, Allergenicity and Toxicity

After we completed designing our vaccine sequence, we submitted it to Vaxijen 2.0 again with the threshold set to 0.5 where the aim was to check whether our newly constructed multi-epitope rabies vaccine was more antigenic than the protein we took initially. The generated result displayed a significant increase in our antigenicity compared to initial protein implying it is recognizable by immune system and obtained protective response generation.

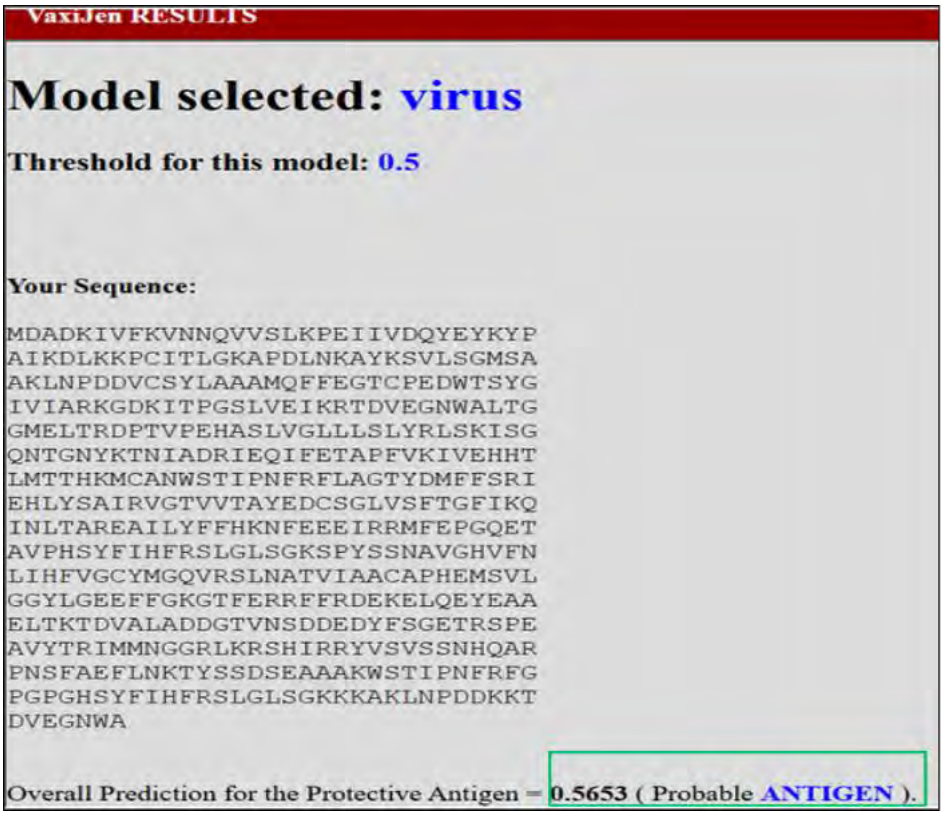


Figure 11: Increased Antigenicity of Vaccine Construct based on 0.5 threshold in Vaxijen 2.0 server (Doytchinova & Flower, 2007).

For checking the allergenicity of our newly constructed sequence we used AllergenOnline server which compared our FASTA input among its broad database of allergenic hits (Goodman et al., 2016). The output result displayed “Number of sequences with hits: 0” which validated that our vaccine design was infact non-allergenic.

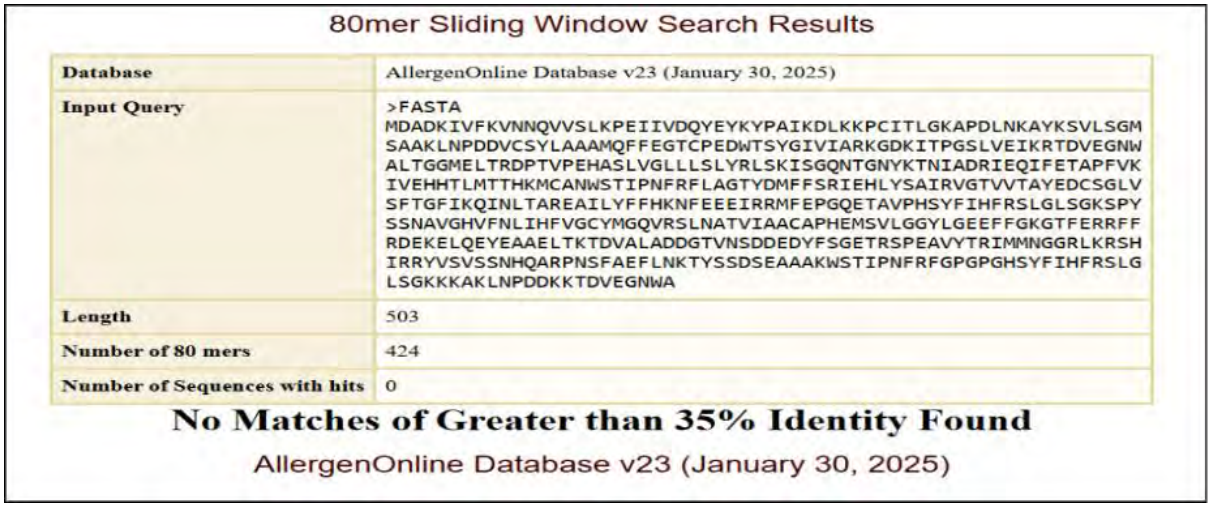


Figure 12: Allergenicity Check of Constructed Vaccine displaying 0 hits (Goodman et al., 2016)

Furthermore, for toxicity check of our vaccine construct we used T3DB server where we paste our sequence in FASTA format (Wishart et al., 2015). The result we got ensured our vaccine was non-toxic because this server generated "your search returned no results".

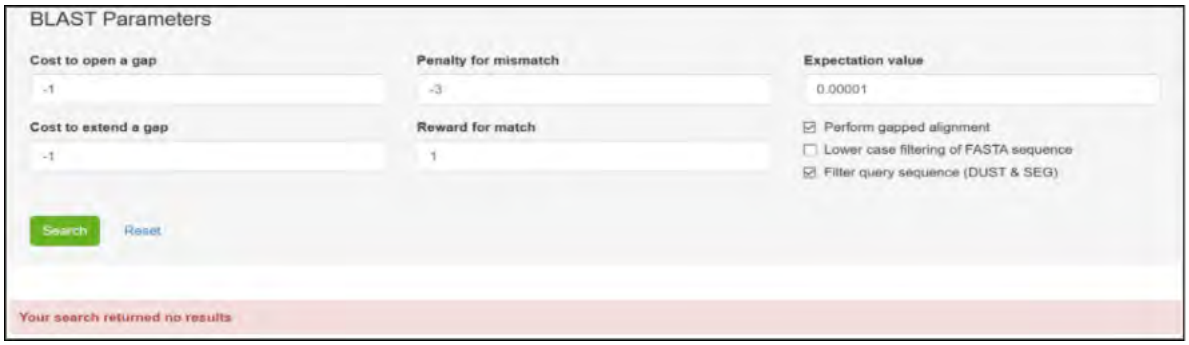
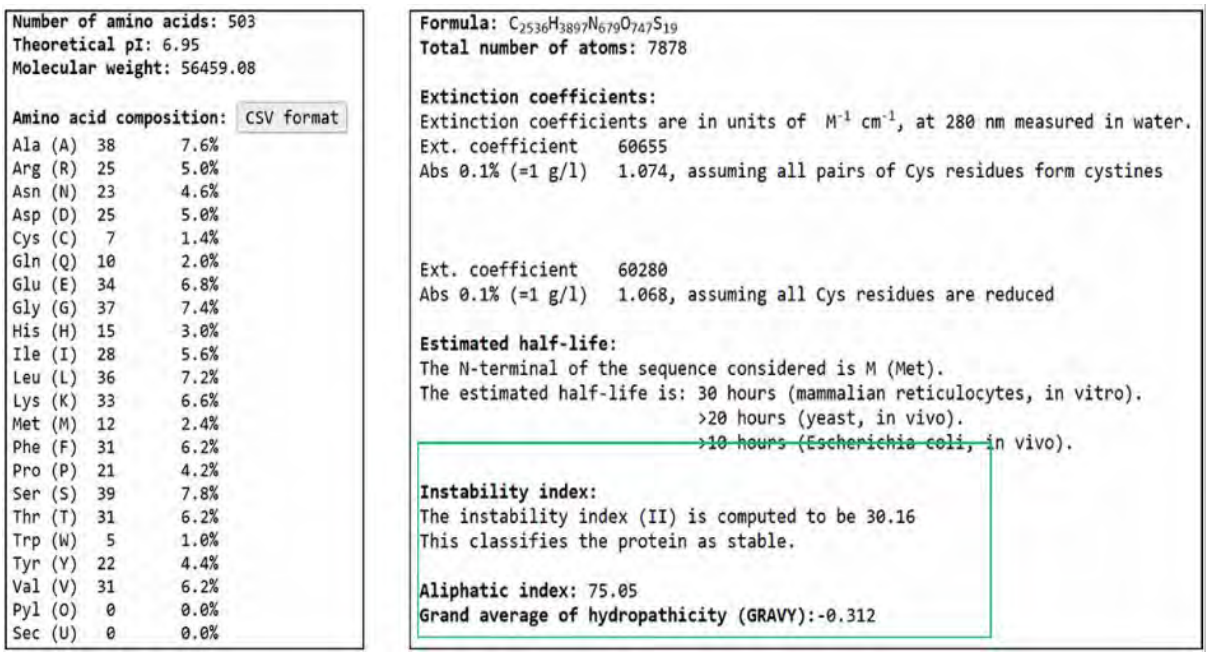


Figure 13: T3DB server displaying no results solidifying the vaccine construct is non-toxic (Wishart et al., 2015)

3.7 Assessment of Constructed Vaccine's Biochemical Properties

By biochemical properties of our constructed vaccine sequence, we mean the overall physical and chemical attributes of it which estimates how it is going to behave in actual biological systems. We determine these properties by ProtParam tool where we submit our sequence and it tells us several physico-chemical attributes of it including- number of amino acids present and its particular composition, isoelectric point which dictates whether our construct is acidic,

basic or neutral designated by theoritical pI and molecular weight. It also estimates atomic formula, extinction coefficient and half-life. Most importantly it tells us about the instability index of our construct where a score below 40 means our protein is stable in-vitro. It also gives us an idea about the thermostability denoted in Aliphatic index which should be more than 60 in value and grand average of hydropathicity or GRAVY score. GRAVY score indicates whether our sequence is hydrophilic (GRAVY score negative) or hydrophobic (GRAVY score positive) where we always aim for a hydrophilic construct. Below is our vaccine construct's ProtParam results:



Number of amino acids: 503
Theoretical pI: 6.95
Molecular weight: 56459.08

Amino acid composition: CSV format

Ala (A)	38	7.6%
Arg (R)	25	5.0%
Asn (N)	23	4.6%
Asp (D)	25	5.0%
Cys (C)	7	1.4%
Gln (Q)	10	2.0%
Glu (E)	34	6.8%
Gly (G)	37	7.4%
His (H)	15	3.0%
Ile (I)	28	5.6%
Leu (L)	36	7.2%
Lys (K)	33	6.6%
Met (M)	12	2.4%
Phe (F)	31	6.2%
Pro (P)	21	4.2%
Ser (S)	39	7.8%
Thr (T)	31	6.2%
Trp (W)	5	1.0%
Tyr (Y)	22	4.4%
Val (V)	31	6.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Formula: $C_{2536}H_{3897}N_{679}O_{747}S_{19}$
Total number of atoms: 7878

Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 60655
Abs 0.1% (=1 g/l) 1.074, assuming all pairs of Cys residues form cystines

Ext. coefficient 60280
Abs 0.1% (=1 g/l) 1.068, 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 30.16
This classifies the protein as stable.

Aliphatic index: 75.05
Grand average of hydropathicity (GRAVY):-0.312

Figure 14: Biochemical Property Analysis values by ProtParam tool (Wilkins et al., 1999).

Our sequence is within the required optimal range of 401 to 600 amino acids being 503 amino acids which is long enough to attain diverse epitopes and compact enough to remain structurally efficient and stable. Our isoelectric point value is 6.95 which is close enough to be near neutral as it is preferred for solubility and stability in a physiological system. Half-life estimated in mammalians is shown to be 30 hours which ensures in vivo stability. Moreover, the molecular weight of our construct is 56.46 kDa approximately. In addition to that, our

computed score in instability index is 30.16 indicating our construct is significantly stable. It is also thermostable (75.05) as the aliphatic score is more than 60 and our GRAVY score is –0.312 which indicates hydrophilicity.

3.8 3D Model Generation of Constructed Vaccine

For homology modelling of our sequence and translating it into a 3D structure we used the immunoinformatic tool Phyre 2.2 where we pasted our sequence and it generated us a 3D model in PDB format. For assuring reliability, it also has two important parameters which are confidence percentage and coverage percentage. Structural visualization was done by the software PyMOL. The model it generated based on our input had a coverage score of 83% (80% or more is desired) where 420 residues of our sequence had been modelled with 100% confidence further validating our constructed multi epitope vaccine.

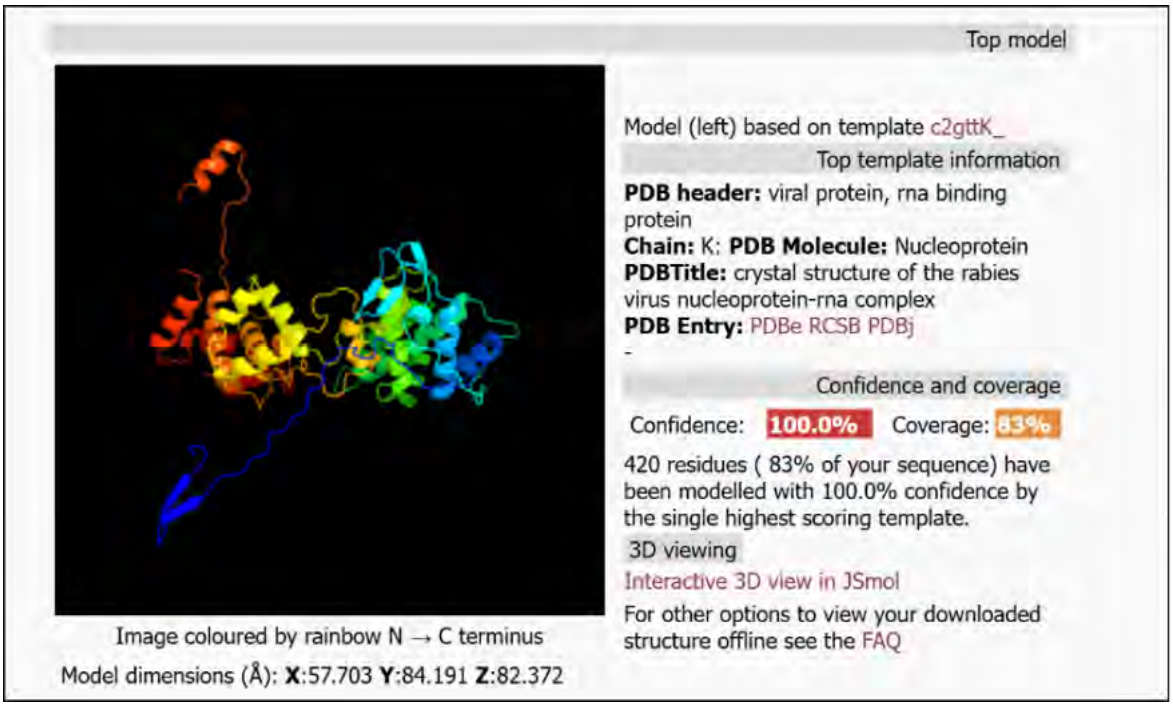


Figure 15: Phyre 2.2 generated 3D model with 83% residue coverage at 100% confidence (Kelley et al., 2015).

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

Employment of Swiss Model Expasy server was done in order to estimate the stereochemistry of our generated PDB model obtained from Phyre 2.2. This is required to ensure the protein's correct folding and overall structural reliability. The data achieved from Swiss Model Expasy is given below:

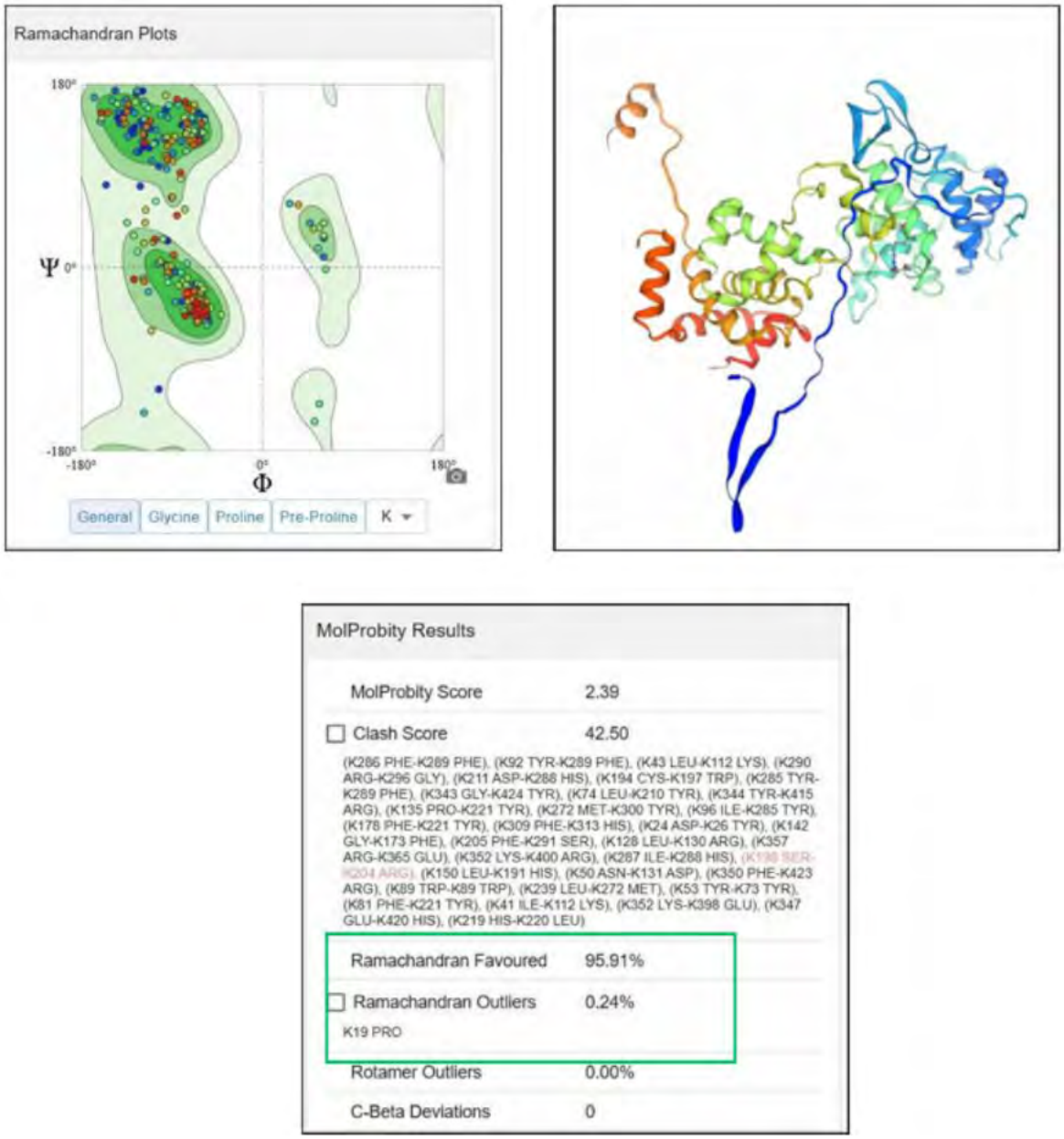


Figure 16: Ramachandran plot generation along with scores and model by Swiss Model Expasy (Kiefer et al., 2009)

We can see that our Ramachandran favored score is 95.91% which is exceptionally high as its required to be above 90% and Ramachandran outliers is 0.24% which is required to be below 2%. What these basically mean is that more than 95% of our sequences residues falls in the energetically favored regions which are the green regions in the graph indicating good folding. Analyzation of this was done by phi (φ) and psi (ψ) dihedral angles distribution. The outliers meaning disallowed portions are only 0.24% indicating less possibility of any structural error or strain. C-Beta deviations are also in acceptable range (< 0.25 Å).

3.10 Assessment of Z-score of Vaccine's 3D model

Determination of Z-score was done by ProSa Web which stands for protein structure analysis, compares our PDB formatted model with real and experimentally validated structures. It has two distinctly colored regions for X-ray crystallography and NMR spectroscopy which tells us which approach our structure falls. The graph is plotted against residue numbers versus Z-score (Wiederstein & Sippl, 2007). After inputting our Phyre's PDB structure in ProSa web, the results generated was – 8.09 Z-score which falls within the acceptable range of -5 to -10. Our model also falls in the light blue x ray region which aids to validate high resolution structure. Local model quality graph (knowledge-based energy versus sequence position) is also generated where most of the residues take place in the negative region further reiterating our Z-score validation.

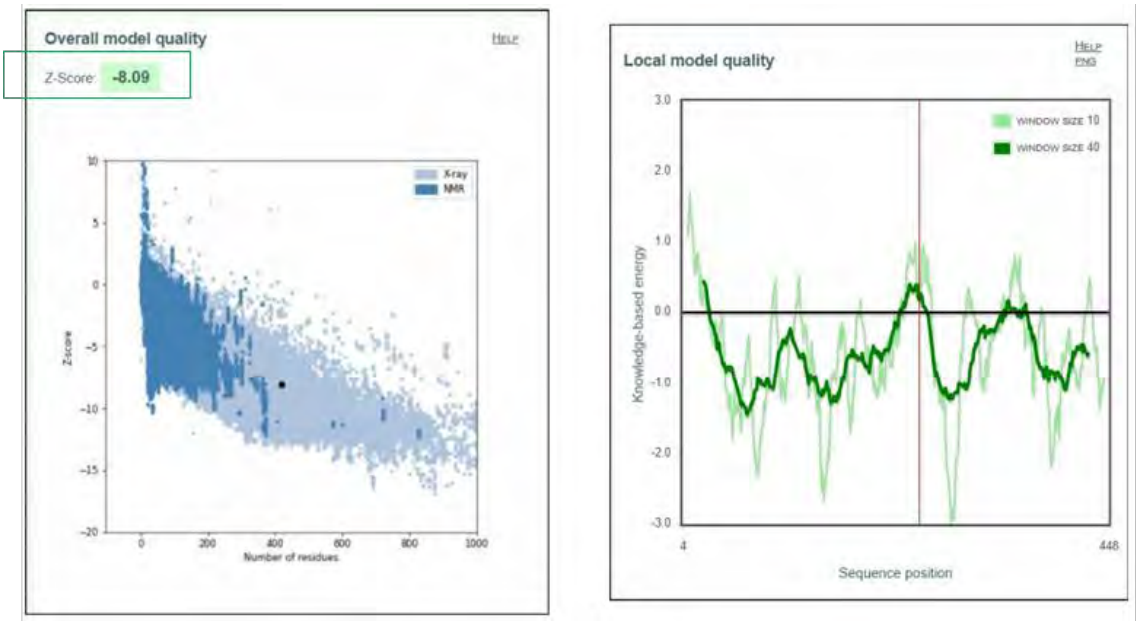
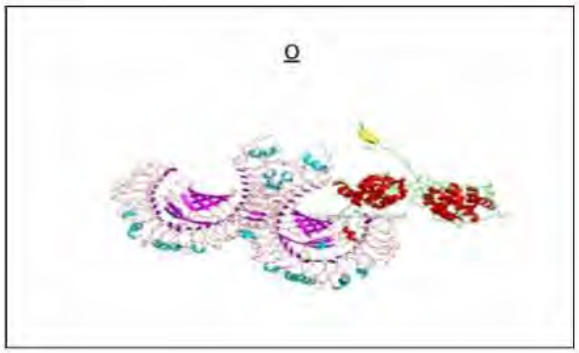


Figure 17: Z-score graph and local model quality graph by ProSa Web (Wiederstein & Sippl, 2007).

3.11 Molecular Docking with Toll like Receptor

To computationally figure out protein-protein docking between our vaccine construct and compatible tool like receptor, we use the server ClusPro in order to figure out whether our vaccine could actually bind with receptors and initiate immune response or not (Kiflu, 2024). There were two molecular sections in ClusPro for receptor molecule and ligand molecule. In the receptor molecule we put Toll like receptor 4 (TLR 4) which was the best fit for our rabies vaccine construct, and we obtained it from the protein databank called RCSB in PDB format. For the ligand section we again put in the Phyre 2.2 PDB structure, then we proceeded to dock (Tombari et al., 2025). The result generated several structures and model scores of those structures could be viewed. We chose the “0” number model which upholds the best visualization of the central (most typical) pose of our 0 cluster which basically entails that “0” has the prime structural representation of our cluster. The cluster we choose is also entitled “0” where all docking poses are clustered based on how identical they are. Cluster 0 is by default the most stable with the greatest number of members (number of cluster’s binding orientations

which are highly similar) in it. Our 0 cluster had 62 member in it suggesting reliability. The representative center (average central conformation's energy) and lowest energy (most stable conformation's energy) both showed weighted score –998.1 indicating a very well compacted cluster with strong binding affinity which has steady reproducible docked conformations. Despite cluster 3 having slightly higher score in lowest energy, it was not considered as it had a significantly lower number of member which statistically lowered reliability. So, we selected 0 cluster due to its balance between high binding energy and more cluster members, making it reliable for further evaluation.



Cluster	Members	Representative	Weighted Score
0	62	Center	-998.1
		Lowest Energy	-998.1
1	53	Center	-989.1
		Lowest Energy	-1003.6
2	31	Center	-817.2
		Lowest Energy	-971.2
3	31	Center	-933.7
		Lowest Energy	-1077.5
4	30	Center	-834.3
		Lowest Energy	-1025.9
5	28	Center	-844.5
		Lowest Energy	-967.8

Figure 18: Docked complex and model energy scores achieved by ClusPro (Comeau et al., 2004)

3.12 C-ImmSim generated Immune Simulation

We submitted our constructed vaccine sequence into a server titles C-ImmSim 10.1 where we alter the simulation steps to 300 and set time-steps for three doses. The generated results predicted graphical outputs of how our multi-epitope vaccine could cause immune response. The first graph shows us 3 black spikes indicating antigens were introduced as doses. In the graphical figure we can see days are plotted in the X axis and antigen and antibody titers are

along Y axis where the black spikes represents antigen introduction via doses and antibody titers in right Y axis show us the production and response of immunoglobulins (Rapin et al., 2010). At day 0 when the antigen was first introduced, it reached a high peak of antigen concentration while simultaneously being primed for further recurrence. IgM (green line) was the antibody first created and total antibody response dictated by the yellow line sees a slight rise indicating our vaccine triggered primary immune response successfully. After the second dose at day 28, the IgM and IgG (total antibody response indicated by the yellow line) significantly rise which states that memory response was formed. Prevalence of IgG assures preserved humoral responses solidifying lasting immunity. We also see rise in IgM (green line) and IgG1+IgG2 (blue line) which validates most long term antibody response. Consequently, at day 58-60 when the third dose was administered, we can see we can see an even better IgM+IgG response and antigen clearance occurs way faster. The other immunoglobulins are also seen spiking significantly. Remarkable and quicker drop is seen in antigen levels after each dose. We can say our rabies vaccine construct initiated a strong immunogenic response and was able to form memory cells for further protection against re-exposure predicting it as efficacious. The figure is given below:

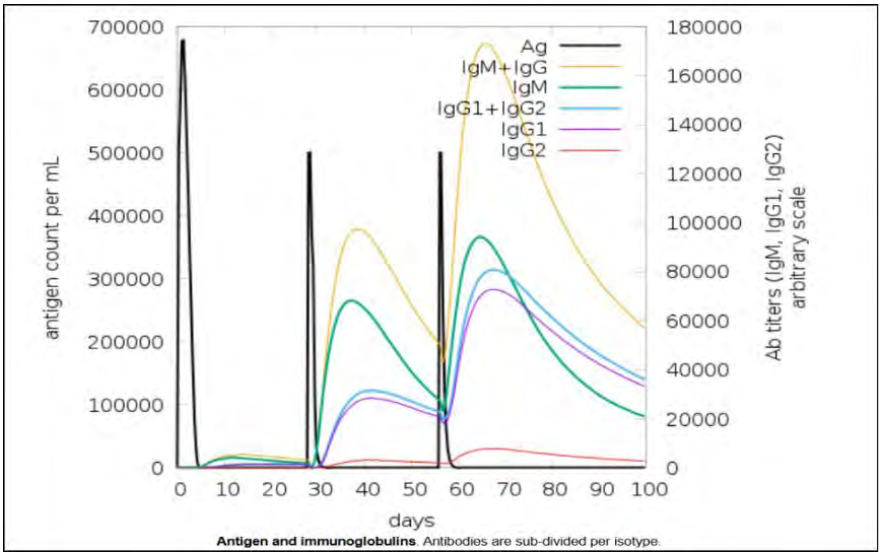


Figure 19: C-ImmSim generated Immune Simulation (Rapin et al., 2010)

The cytokine and interleukin concentration graph shows us remarkably raised levels of interferon gamma and interleukins- 4,10 and 12 which indicates strong cell mediated Th1-type response, Th2-type humoral response, inflammation regulation and promoting Th1 differentiation aiding better production of interferon gamma respectively. An inset graph box is futher observed where spikes in interleukin 2 indicates active T cell proliferation.

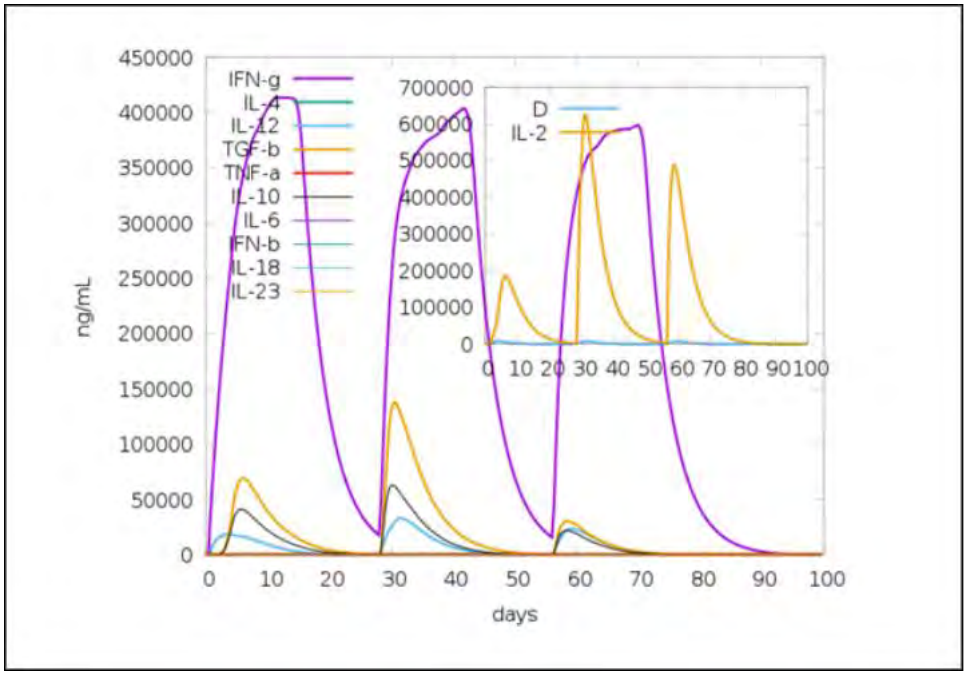


Figure 20: Cytokine and Interleukin Concentration graph containing IFN gamma, Interleukin 4 and 10 (Rapin et al., 2010)

We also analyzed the B lymphocyte population graph where we could see significant rises in total B cell which assures us that our vaccine is effective in producing humoral immune response in host and aiding in prolonged immunity by production of memory cells. Antibody maturation is indicated by isotype switching where IgM spikes first telling us primary immunological response has been initiated along with shifting to constant increasement of the IgG which points maturation of B cell. Again, memory cells indicated in green line spikes which ensure protection in case of re-exposure.

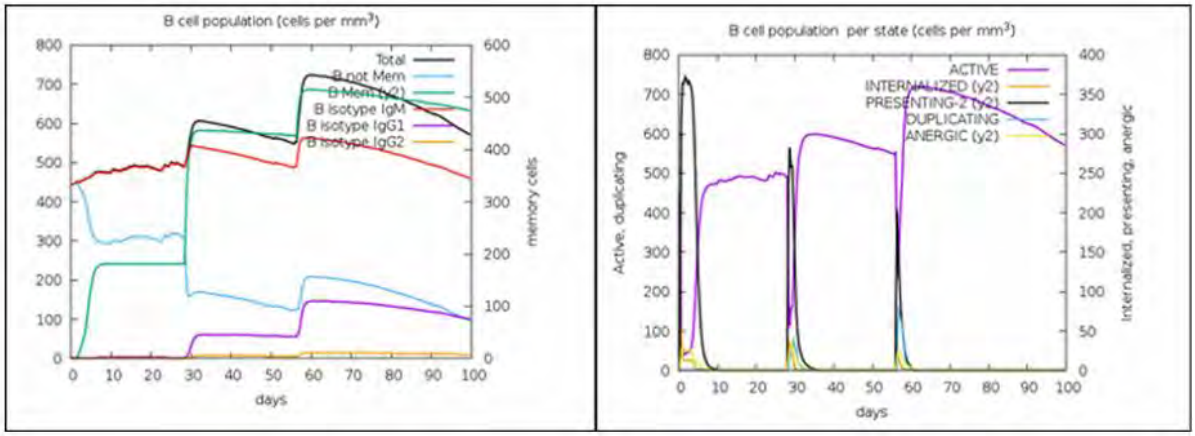


Figure 21: B cell Lymphocyte population graph (Rapin et al., 2010)

Further set of graphs explore the $CD4^+$ T helper (Th) cells where it is observed that significant spikes in Th memory, active and duplicating cell population after the doses had been administered which aids highly in T cell activation. All of this combined, validates that our vaccine could generate protective response when the host has been re-exposed to the pathogen.

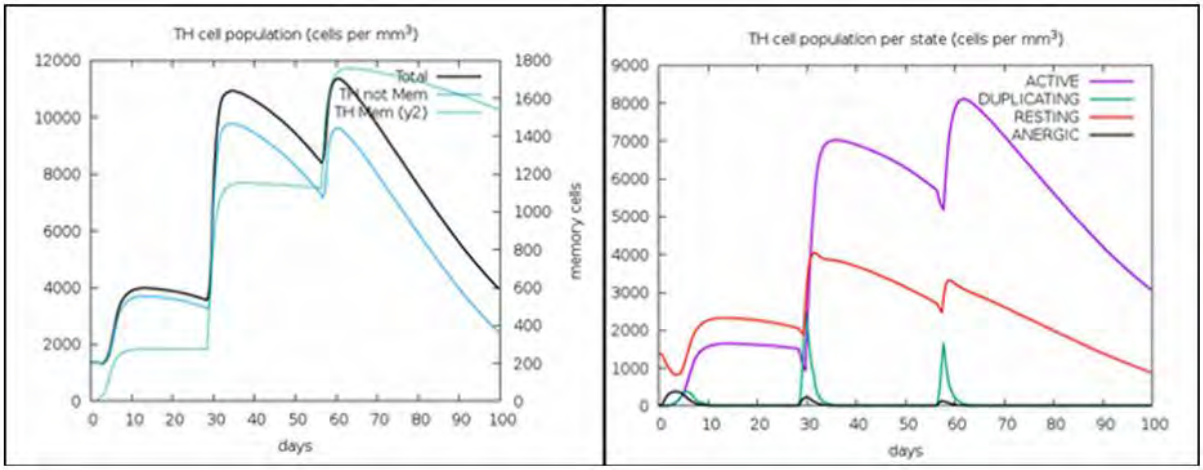


Figure 22: CD4 T helper Lymphocyte population graph (Rapin et al., 2010)

The last set of graph entails the $CD8^+$ cytotoxic lymphocytes and their population. Dominant Increase of active population and gradual decline of resting population annotes long-term and

sustained immunological response further. Steady anergic response which confirms regulated and steady baseline for immune response additionally reiterating that our vaccine is potent and prevent overactivation.

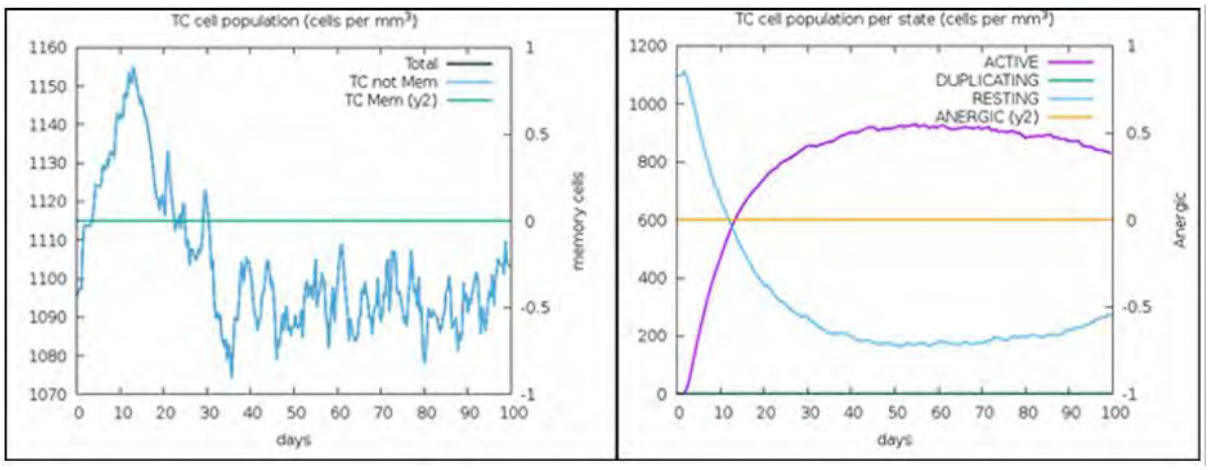


Figure 23: CD8 cytotoxic T Lymphocyte population graph (Rapin et al., 2010)

Further evaluation of all the graphs generated by C-ImmSim according to our constructed vaccine visibly portrays the generation of both cellular and humoral response within our number of doses where the first dose was applied to prime our immune system against Rabies virus. Along with that, long term immunity is also assured as memory cell population was available. In the combined graph, we can see that the concentration of antigen was raised thrice and each time the clearance peak got steeper ensuring that due to our vaccine introduction efficiently cleared antigen and faster than before. All of this graph computationally ensures the efficacy of our proposed constructed vaccine.

Chapter 4

Discussion

Rabies is a life-threatening virus and infection from it leads to acute infection of the brain and spinal cord which later culminates to Rabies Encephalitis. It is an endemic in Bangladesh and yet people are not aware about its devastation (Ghosh et al., 2024). The usual treatment after being infected with rabies includes multiple dose of PEP regimen also known as post exposure prophylaxis and RIG or rabies immunoglobulin administration. There are available pre-exposure prophylaxis (PrEP) vaccines, but they are not accessible in general setting. Bangladeshi patients do not even comply to the PEP regimen as well where only 78% of the bitten victims completed the full PEP treatment and even among these only 20% received vaccine on the day of exposure and other 66% delayed it between 1-3 days (Tamanna et al., 2023). So, it is essential for us to design a pre-exposure prophylaxis vaccine which would be an accessible, cost-effective, reproducible and would have less number of doses (Ghosh et al., 2024). Our study aims to propose an in silico multi-epitope based subunit vaccine design against rabies virus. We selected rabies nucleoprotein which is a structural protein and a good vaccine target (along with glycoprotein, phosphoprotein and matrix protein) as it binds viral RNA (Tombari et al., 2025). Moreover, (Tombari et al., 2025), describes N protein as a good choice for being highly conserved among rabies strains and good T cell induction which compliments antibody mediated immunity. Cytotoxic T lymphocytes, Helper T lymphocytes and B lymphocytes epitopes were selected and consequently ran through antigenicity, allergenicity and toxicity screening. Helper T cell epitopes were further validated based on whether they could induce cytokine and interleukin (Interferon gamma, IL-4 and IL-10) stimulation. All four selected epitopes (one from CTL, one from HTL and two from B cell) were only chosen when they passed all these screenings and had exceptionally high antigenic scores. These were conjoined by well validated linkers. The constructed sequence having 503

amino acids was again tested by Vaxijen 2.0 ensuring increased antigenicity. After checking its allergenicity and toxicity further, we validated its molecular weight, stability, solubility (negative GRAVY score) and thermostability by ProtParam which indicated acceptable physico chemical content of our prepared vaccine sequence. 3D model of our construct was generated with Phyre 2.2 and tested parameters in confidence and coverage percentage. Our model yielded 83% coverage in homology modelling at 100% confidence ensuring good structural reliability. 95.91% Ramachandran favored value along with only 0.24% outliers indicated our model had accurate protein folding and good stereochemistry further strengthening structural validation. Our model was further compared with experimentally validated models solved by X ray crystallography and NMR spectroscopy using ProSa Web, where our model fell in the x ray region which solved high resolution structures. The -8.09 Z score indicated a good quality and reliable model with proper energy distribution. Afterwards, Molecular docking was done through ClusPro, where we chose our phyre structure as ligand molecule and TLR4 as receptor (Tombari et al., 2025). We chose model 0 and cluster 0 which displayed high binding affinity score being –998.1. Moreover, for host system evaluation of constructed vaccine we computationally simulated its immune responses utilizing C-ImmSim 10.1 server (Bernaschi & Castiglione, 2001). It generated well acceptable results where elevated cytotoxic and helper T cells were displayed along with memory B cell population and incredibly pronounced and steady IgM+IgG response which culminates to our construct being an efficacious vaccine against rabies virus. The cytokine and interleukin concentrations were also well elevated, which is desired. After concluding all these evaluation steps, we can acclaim that our vaccine design is potent and can generate desired immunogenic responses along with providing long term protection all well validated by multiple in silico computational approaches.

Chapter 5

Conclusion

Our research was set to propose an *in silico* based multi epitope subunit vaccine design against rabies virus since rabies is an endemic in Bangladesh and it is immensely necessary to be protected against it prior infection as it is a 100% fatal once the patient becomes symptomatic. Our study included numerous computational ways to identify highly antigenic cytotoxic T cell, helper T cell and B cell epitopes which are validated as non-allergenic and non-toxic. The linkers used (EAAAK, GPGPG and KK) are universally optimal linkers which aid to connect the vaccine construction for proper epitope presentation while protecting structural integrity. Once the vaccine was constructed, we analyzed its biochemical properties using ProtParam which yielded desirable results in every parameter. Further methodology was applied to generate our vaccine's model by Phyre 2.2 which was displayed high confidence and coverage percentage. The structure validity was additionally checked by finding out its residue position and stereochemical balance by analyzing Ramachandran scores. ProSa Web was used to identify that our structure had an acceptable Z score and fell within X ray region. Application of ClusPro was introduced for molecular docking where TLR4 was docked with our vaccine model and the results showed favorable structure and high binding scores. Lastly, C-ImmSim was applied to estimate how our vaccine would act in a host system which ensured strong humoral and cellular immunity along with long-lasting protection.

Even though our research has a strong computational foundation we need to consider its biggest limitation that this study is based on all computational algorithm. For determining our vaccine's actual capacity it needs to be involved in *in-vitro* and *in-vivo* experimentations. If its efficacy and safety could be validated, it could emerge as an accessible and reproducible pre exposure prophylaxis measure against this menacing virus which would need fewer doses than PEP.

References

Alemayehu, T., Oguttu, B., Rupprecht, C. E., & Niyas, V. K. M. (2024). Rabies vaccinations save lives but where are the vaccines? Global vaccine inequity and escalating rabies-related mortality in low- and middle-income countries.

Ayyagari, V. S., Venkateswarulu, T. C., Abraham Peele, K., & Srirama, K. (2022). Design of a multi- epitope based vaccine targeting M protein of SARS CoV2: an immunoinform atics approach.

Bastos, V., Pacheco, V., Rodrigues, É. D. L., Moraes, C. N. S., Nóbile, A. L., Fonseca, D. L. M., Souza, K. B. S., do Vale, F. Y. N., Filgueiras, I. S., Schimke, L. F., Giil, L. M., Moll, G., Cabral-Miranda, G., Ochs, H. D., Vasconcelos, P. F. da C., de Melo, G. D., Bourhy, H., Casseb, L. M. N., & Cabral-Marques, O. (2023). Neuroimmunology of rabies: New insights into an ancient disease.

Bernaschi, M., & Castiglione, F. (2001). Design and implementation of an immune system simulator.

Capsoni, F., Minonzio, F., Ongari, A. M., Carbonelli, V., Galli, A., & Zanussi, C. (1995). IL-10 up-regulates human monocyte phagocytosis in the presence of IL-4 and IFN-gamma.

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

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

Dürr, S., Naïssengar, S., Mindekem, R., Diguimbye, C., Niezgoda, M., Kuzmin, I., Rupprecht, C. E., & Zinsstag, J. (2008). Rabies diagnosis for developing countries.

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.

Ge, P., Tsao, J., Schein, S., Green, T. J., Luo, M., & Zhou, Z. H. (2010). Cryo-EM Model of the Bullet-Shaped Rabies Virus.

Ghosh, S., Hasan, M. N., Nath, N. D., Haider, N., Jones, D. H., Islam, M. K., Rahaman, M. M., Mursalin, H. S., Mahmud, N., Kamruzzaman, M., Rabby, M. F., Kar, S., Ullah, S. M., Ali Shah, M. R., Jahan, A. A., Rana, M. S., Chowdhury, S., Uddin, M. J., Sunil, T. S., … Islam, M. N. (2024). Rabies control in Bangladesh and prediction of human rabies cases by 2030: a One Health approach.

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.

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.

Hampson, K., Coudeville, L., Lembo, T., Sambo, M., Kieffer, A., Attlan, M., Barrat, J., Blanton, J. D., Briggs, D. J., Cleaveland, S., Costa, P., Freuling, C. M., Hiby, E., Knopf, L., Leanes, F., Meslin, F. X., Metlin, A., Miranda, M. E., Müller, T., … Dushoff, J. (2015). Estimating the Global Burden of Endemic Canine Rabies.

Jadeja, N. B., & Vanak, A. T. (2023). High-throughput techniques to understand evolution and transmission trends of rabies virus in Asian countries.

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

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.

Khalifa, M. E., Unterholzner, L., & Munir, M. (2021). Structural and Evolutionary Insights Into the Binding of Host Receptors by the Rabies Virus Glycoprotein.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources.

Kiflu, A. B. (2024). The Immune Escape Strategy of Rabies Virus and Its Pathogenicity Mechanisms.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein.

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.

Mehta, S., Samant, L., & Chowdhary, A. (2014). Computational modeling and in silico characterization of rabies virus glycoprotein and nucleoprotein.

Occurrence of rabies | Rabies – Bulletin – Europe. Retrieved October 10, 2025.

Odhar, H. A., Hashim, A. F., Humadi, S. S., & Ahjel, S. W. (2023). Design and construction of multi epitope- peptide vaccine candidate for rabies virus.

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.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.

Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. (2010). NetCTLpan: pan-specific MHC class I pathway epitope predictions.

Tamanna, S., Yasmin, D., Ghosh, S., Rahaman, M. M., Dey, A. K., Das, T. K., & Chowdhury, S. (2023). Evaluating adherence to government recommendations for post-exposure rabies vaccine among animal-bite victims: A hospital-based study in Bangladesh.

Tombari, W., Khamessi, O., Othman, H., Kallala, O., Mahjoub, R., Ghedira, K., & Trabelsi, A. (2025). A novel mRNA-based multi-epitope vaccine for rabies virus computationally designed via reverse vaccinology and immunoinformatics.

UniProt. Retrieved October 10, 2025.

Walker, P. J., Blasdell, K. R., Calisher, C. H., Dietzgen, R. G., Kondo, H., Kurath, G., Longdon, B., Stone, D. M., Tesh, R. B., Tordo, N., Vasilakis, N., & Whitfield, A. E. (2018). ICTV Virus Taxonomy Profile: Rhabdoviridae.

Wang, P. H., & Xing, L. (2024). The roles of rabies virus structural proteins in immune evasion and implications for vaccine development.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.

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.

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.