

***“In Silico Design of a  
Multi-Epitope Vaccine against  
Lyssavirus Phylogroup II Glycoproteins”***

**By**

Fahmida Fabiha

18136057

**&**

S. M. Salim Shadman

17136037

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirement for the degree of Bachelors in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

December 2022

© 2022. BRAC University

All rights reserved

# Declaration

It is hereby declared that

- The thesis titled “*In Silico* Design of a Multi-Epitope Vaccine against *Lyssavirus* Phylogroup II Glycoproteins” is our own original work completed by us as a prerequisite submission in requirement of the course “Biotech Project” coded as “BTE450” in the Biotechnology program of the Department of Mathematics and Natural Sciences of BRAC University, Dhaka.
- The thesis does not contain materials previously published or written a third party except where this is appropriately cited through full and accurately referencing.
- The thesis does not contain any material which has been accepted or submitted for any other degree.
- I have acknowledged all main sources for help.

Name of Student and Signature:



---

Fahmida Fabiha

18136057



---

S. M. Salim Shadman

17136037

# Approval

The thesis titled “*In Silico* Design of a Multi-Epitope Vaccine against *Lyssavirus* Phylogroup II Glycoproteins” submitted by Fahmida Fabiha (ID: 18136057) and S. M. Salim Shadman (ID: 17136037) has been accepted as satisfactory in partial fulfillment of requirement for the degree of Bachelor of Science in Biotechnology on the 18<sup>th</sup> of December, 2022.

## Examining Committee:

Supervisor:

Rafeed Rahman Turjya

Lecturer, Department of Mathematics and Natural Sciences

BRAC University

Program Director:

Dr. Munima Haque

Associate Professor, Department of Mathematics and Natural Sciences

BRAC University

Departmental Head:

Chairperson, Department of Mathematics and Natural Sciences

BRAC University

# Acknowledgment

We are grateful to Almighty for the blessings upon us which have enables us to come to this far in our education in sound health and allowed us to finish our thesis successfully.

We convey our heartfelt gratitude to our supervisor, Rafeed Rahman Turjya, Lecturer of Department of Mathematics and Natural Sciences, BRAC University, for his continuous support and guidance throughout our work and undergrad journey. He continuously provided encouragement and was always willing and enthusiastic to assist in all the possible ways throughout the research project. This thesis would not have been a success without his valuable advice and feedbacks. We would like to thank him very much for his support and understanding throughout this journey.

Lastly, we would like to thank our parents for always supporting us in our academic endeavors.

# Abstract

The genus of *lyssavirus*, coming from the family of Rhabdoviridae, has been around since the time of 2300 BC. A ~12kb, negative-sense RNA virus, it is known to be one of the lethal viruses ever encountered by mankind. With the advancement in the fields of genetics and bioinformatics, we have been able to classify the genus into 3 phylogroups, phylogroups I, II, & III. Available and newly engineered vaccines target the phylogroup I and III, but no significant vaccine is available for combating the phylogroup II viruses. In this study, we used immunoinformatics based approach to design a multi-epitope-based vaccine that can provide immunity against the phylogroup II lyssaviruses, Lagos Bat Virus, Mokola Bat Virus & Shimoni Bat Virus. We have identified conserved epitopes within the viral glycoprotein sequences, and constructed vaccines containing immunogenic motifs alongside these epitopes. We predicted and optimized the three-dimensional structures of our vaccines, and assessed their capacity to induce immunity. Our designed vaccines are highly antigenic, non-allergenic, and provide wide coverage. They have shown high binding affinity against MHC molecules and induced long-term immunity in immune simulation. We believe that *in silico* design of these vaccines is the first step in preparation against a future spread of phylogroup II *lyssavirus* species.

# Table of Contents

Section No.	Name	Page No.
	Declaration	i
	Approval	ii
	Acknowledgement	iii
	Abstract	iv
	Table of Contents	v-vii
	List of Tables	viii-ix
	List of Figures	x-xii
<b>Chapter 1</b>	<b>Introduction</b>	<b>1</b>
1.1	History	1
1.1.1	Characteristics	1
1.1.1.1	Viral Entry, Spread and Proliferation	1-2
1.1.1.2	Symptoms, Disease Progression, Prevention, and Treatment	2-4
1.2	Genome and Protein	4
1.2.1	Genome	4
1.2.2	Protein	4-7
1.3	Phylogroups of Lyssaviridae	7-12
1.3.1	Phylogroup I	12-13
1.3.2	Phylogroup II	13-14
1.3.3	Other Species	14
1.4	Pathogenicity	15
1.4.1	Factors related to the pathogenicity	16-17
1.5	Available and Designed Vaccines	17-25
1.6	Why Phylogroup II?	25-26
<b>Chapter 2</b>	<b>Methodology</b>	<b>27</b>
2.1	Strain Identification and protein sequence retrieval	27
2.2	Protein sequence alignment	27

2.3	Conserve sequence	27
2.4	MHC I Epitope Prediction	28
2.5	MHC II Epitope Prediction	28-29
2.6	B-Cell Epitope Prediction	29
2.7	Stored identified epitopes in a FASTA format	29
2.8	Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes	29-33
2.9	3D Structure Prediction of the Epitopes	33
2.10	Molecular Docking of the MHC Epitopes	34
2.11	Binding prediction of the MHC peptides to HLA alleles and Population coverage for the filtered MHC epitopes	34-35
2.11.1	Population coverage analysis	35
2.12	Vaccine construction	35-36
2.13	Antigenicity, Allergenicity and Physicochemical properties analysis of each vaccines	36
2.14	2D Structure Prediction of the Constructed Vaccines	36-37
2.15	3D Structure Prediction of the Constructed Vaccines	37-38
2.16	Disulfide engineering of vaccines	38
2.17	Molecular Docking of the Constructed Vaccines against MHC I and MHC II	39
2.18	B-lymphocytic Epitope prediction	39-40
2.19	Immune Simulation Analysis	40
<b>Chapter 3</b>	<b>Results</b>	<b>41</b>
3.1	Result of Strain Identification and Protein Sequence Retrieval	41-44
3.2	Protein sequence alignment and conserved region finding	44-49
3.3	Predicted MHC I Epitopes	49-54
3.4	Predicted MHC II Epitopes	54-55
3.5	Predicted B-Cell Epitopes	56-57
3.6	Stored Epitope Sequences	57-59

3.7	Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes	59-60
3.8	Predicted 3D Structures of Epitopes	60-62
3.9	Molecular Docking Analysis of the MHC Epitopes	62-64
3.10	Binding prediction of the MHC peptides to HLA alleles and Population coverage for the filtered MHC epitopes	64-67
3.10.1	Population coverage analysis of the epitopes and their MHC alleles	67-72
3.11	Multi-Epitope Vaccine Constructs	72-75
3.12	Antigenicity, Allergenicity and Physicochemical properties analysis of each vaccine	75-78
3.13	2D Structure of Vaccine Constructs	78-86
3.14	3D Structure of Vaccine Constructs	86-89
3.15	Disulfide engineering of vaccines	90-94
3.16	Molecular Docking Analysis of the Vaccine Constructs against MHC I and MHC II	94-97
3.17	B-lymphocytic Epitope prediction	97-105
3.18	Immune Simulation of the Vaccines	106-109
<b>Chapter 4</b>	<b>Discussion</b>	<b>110-121</b>
<b>Chapter 5</b>	<b>Conclusion</b>	<b>122</b>
	References	123-133



## List of Tables

Table No.	Name	Page No.
1.1	Current diversity and taxonomy of <i>lyssaviruses</i>	10-12
1.2	Available & Designed Vaccine List	18-26
2.1	List of the Antigenicity, Allergenicity, Toxicity and Human Homology Prediction of conserved epitopes	31-33
3.1	List of the Glycoprotein sequences with their length and their accession number	41-44
3.2	Phylogroup II conserved region from MUSCLE and M-Coffee Alignment	44-49
3.3	Virus MHC I Table based on frequently occurring alleles	49-50
3.4	Mokola Bat Virus MHC I Table based on frequently occurring alleles	50-51
3.5	Shimoni Bat Virus MHC I Table based on frequently occurring alleles	51-52
3.6	Lagos Bat Virus MHC I Table based on top-10 percentile rank	52
3.7	Mokola Bat Virus MHC I Table based on top-10 percentile rank	53
3.8	Shimoni Bat Virus MHC I Table based on top-10 percentile rank	54
3.9	LBV, MKV & SBV combined MHC I Table based on 10 percentile rank	55
3.10	B-cell epitope list	56
3.11	Best Selected B-cell epitope	56
3.12	List of identified epitopes in fasta	57-59
3.13	Result of the Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes analysis	59-60
3.14	Binding Energy calculation for MHC I and MHC II docking human 1QEW and 2G9H	64
3.15	Result of the predicted HLA allele sets for the MHC epitopes	65-67
3.16	Population Coverage Analysis result for African region	68

3.17	Central Africa	68
3.18	East Africa	69
3.19	North Africa	70
3.20	South Africa	70
3.21	West Africa	71
3.22	World	72
3.23	List of Adjuvant	73
3.24	Result of the Antigenicity, Allergenicity and physicochemical properties analysis of constructed vaccines	76-78
3.25	Interpretation of JPred4 signs	81
3.26	Interpretation of Signs used in 2dss tool	83
3.27	Interpretation of results used in 2dss tool	84
3.28	Comparing analysis between SAVES and PDBsum result	88
3.29	The mutated positions in chain A of the V1, V2 and V3	90
3.30	The antigenicity, allergenicity and physicochemical properties analysis of the modified vaccines using the FASTA sequences	92-94
3.31	Cluster model binding affinity score analysis	97
3.32	List of the predicted Continuous epitopes for V1	98
3.33	List of the predicted Continuous epitopes for V2	98-99
3.34	List of the predicted Continuous epitopes for V3	99
3.35	List of the predicted Discontinuous epitopes for V1	99-100
3.36	List of the predicted Discontinuous epitopes for V2	100-101
3.37	List of the predicted Discontinuous epitopes for V3	102

## List of Figures

Figure No.	Name	Page No.
1.1	Genomic structure of rabies	4
1.2	Phylogenetic tree of the <i>lyssavirus</i> phylogroups and their respective species	9
1.3	Pathogenesis of Rabies virus into host	15
1.4	Pathway of Nicotinic acetylcholine receptor (nAChR) from Muscle to Neuron	17
3.1	MHC-II group 1 dataset	61
3.2	MHC-I 4 data sets, (a) set1: Position 122_130, (b) set 2: Position 122_131	61
3.3	B cell epitope data sets, (a) data set 1: Position 122_136, (b) data set 2: Position 122_136	62
3.4	MHC I (4) epitopes docking against 1QEW	63
3.5	MHC II epitope docking against 2G9H	63
3.6	Representation of MHC Class I, Class II and Class Combined coverage of Central Africa	69
3.7	Representation of MHC Class I, Class II and Class Combined coverage of East Africa	69
3.8	Representation of MHC Class I, Class II and Class Combined coverage of North Africa	70
3.9	Representation of MHC Class I and Class Combined coverage of South Africa	71
3.10	Representation of MHC Class I, Class II and Class Combined coverage of West Africa	71
3.11	Representation of MHC Class I, Class II and Class Combined coverage of World	72
3.12	Structural construction of Vaccine 1	74
3.13	Structural construction of Vaccine 2	74

3.14	Structural construction of Vaccine 3	75
3.15	PSIPRED result interpretation for 2d structure prediction	78
3.16	2d structure prediction by PSIPRED for vaccine 1	79
3.17	2d structure prediction by PSIPRED for vaccine 2	80
3.18	2d structure prediction by PSIPRED for vaccine 3	81
3.19	2d structure prediction by Jpred4 for vaccine 1	82
3.20	2d structure prediction by Jpred4 for vaccine 2	82
3.21	2d structure prediction by Jpred4 for vaccine 3	83
3.22	2dss comparison between PSIPRED and JPred4's secondary structure prediction for Vaccine 1	85
3.23	2dss comparison between PSIPRED and JPred4 's secondary structure prediction for Vaccine 2	85
3.24	2dss comparison between PSIPRED and JPred4 's secondary structure prediction for Vaccine 3	86
3.25	Vaccine 1 3d model structure	86
3.26	Vaccine 2 3d model structure	87
3.27	Vaccine 3 3d model structure	87
3.28	Vaccine 1 best model (Model 4 and Model 5)	88
3.29	Vaccine 2 best model (Model 1 and Model 2)	89
3.30	Vaccine 3 best model (Model 3 and Model 4)	89
3.31	Representation of the 3D structure of the disulfide modified vaccines visualization using RCSB PDB - Mol* 3D Viewer (disulfide bonds are indicated with the ball shaped structure)	91
3.32	a) MHC I docking against Vaccine 1 Cluster 2_4    b) MHC 1 docking against Vaccine 2 Cluster 1_1 c) MHC I docking against Vaccine 3 Cluster 5_1	95
3.33	a) MHC II docking against Vaccine 1 Cluster 10_1 b) MHC 2 docking against Vaccine 2 Cluster 4_1 c) MHC II docking against Vaccine 3 Cluster 6_2	96

3.34	2D score charts for three vaccines (V1, V2, V3)	103
3.35	The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine 1	104
3.36	The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine 2	105
3.37	The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine	105
3.38	Vaccine 1_virus, immunoglobulins and immunocomplexes	106
3.39	Vaccine 1_concentration of cytokines and interleukins	106
3.40	Vaccine 2_virus, immunoglobulins and immunocomplexes	107
3.41	Vaccine 2_concentration of cytokines and interleukins	108
3.42	Vaccine 3_virus, immunoglobulins and immunocomplexes	108
3.43	Vaccine 3_concentration of cytokines and interleukins	109
4.1	Vaccine 1, Vaccine 2, Vaccine 3 B-cell population comparison (cells per mm <sup>3</sup> )	115
4.2	Vaccine 1, Vaccine 2, Vaccine 3 TC population per state comparison	116
4.3	Vaccine 1, Vaccine 2, Vaccine 3 TH (Helper T) cell population comparison	118
4.4	Vaccine 1, Vaccine 2, Vaccine 3 TH cell population per state comparison	119
4.5	Vaccine 1, Vaccine 2, Vaccine 3 NK-cell population comparison	120

## **Chapter 1. Introduction**

The genus of *Lyssavirus* comes from the family of *Rhabdoviridae* from the order Mononegavirales, which are consist of a single stranded, negative -sense RNA viruses, which are also known to cause encephalomyelitis commonly known as rabies[1] Presently *Lyssavirus* is known to be composed of 17 viral species and one putative strain[2].

### **1.1 History**

Around 2300 BC in Egypt and in ancient Greece, rabies was first recognized, from the descriptions of Aristotle. In the Avesta (Persia) in the sixth century BC, in the Susrutasamhita (India), in the first century BC, canine based rabies was also noted. Zinke in the year of 1804 observed the infectious nature of the saliva from the infected dogs. Before the year of 1885, there had been no curative treatment or preventive treatment available until Pasteur's discovery. In 1881 Pasteur established the rabies virus's ability to invade and attack the neural tissue. Before understanding the virus properly and having a detailed analysis of the virus's own structure and properties in 1885, Pasteur had worked out on a rabies vaccine. The first patient who received the vaccine was Joseph Meister, who had been attacked by a rabid infected animal. Remlinger and Riffat-Bay in the year of 1903, discovered RABV. Rabies was again observed to make an appearance in the year of 1940s, in Kaliningrad area, from there it spread to Central and Western Europe within a few decades[3]. In Switzerland, during the time of 1978, the very first oral based vaccination for rabies was conducted, for wildlife animals, thus other countries followed up this procedure as well.

#### **1.1.1 Characteristics**

##### **1.1.1.1 Viral Entry, Spread and Proliferation**

Upon entering into the host body, the virus, it slowly migrates towards muscle cells to infect by the nicotinic acetylcholine receptor — and replicates there at a low pace[3]. Meanwhile, the virus tends to remain localized to the inoculation site for variable periods, which may contribute to the variable incubation period characteristic of

rabies[4]. In case of inoculum, with higher titers without the need of initial replication at the muscle cells, the RABV can infect motor endplates [5].Furthermore, RABV uses motor endplates at the neuromuscular junction, to gain entry into the peripheral nervous system (PNS) but exact procedure of this process, needs to contain more data to have a proper conclusion. Furthermore, from PNS to CNS, the RABV travels by microtubule dependent retrograde fast axonal transport[6],[7]. The virus has been observed to travel from neuron to neuron, afterwards, replicating itself and continuing its progression towards the CNS and the brain [8] . **p75NTR**, a type of neuro-trophin receptor is known for promoting comparatively rapid and direct transport of RABV to the CNS[9],[10]. The L protein has been observed manipulating the microtubules for better transportation efficiency[11], meanwhile, the M protein also facilitates the depolymerization of microtubules resulting in improved viral transcription and replication efficiency.

Study suggests that, in humans Retrograde transport occurs at an approximate rate of 50 – 100mm per day, with a species-dependent variation[6],[10]. Evidence based data also indicates that, RABV undergoes, active G protein-dependent anterograde transport in peripheral neurons - such as Dorsal Root Ganglion (DRG) neurons — at a rate three times faster than that of retrograde transport [10, 11]. After entering into the CNS, RABV continuous to spread by retrograde axonal transport, thought to be facilitated by metabotropic glutamate receptor subtype 2, which functions as a cellular entry receptor and is abundant throughout the central nervous system (CNS)[12]. Reaching to the brainstem and ultimately to the brain, the RABV virus finally proliferates and gradually shows clinical symptoms manifest. It spreads to the salivary glands along terminal axons via anterograde transport [13], here it remains to proliferate and then tends travel via saliva for transmitting to a new host. RABV is known for spreading to peripheral, non-neuronal organs anterograde transport, and can be detected in these sites after the onset of clinical symptoms [8], [14].

### **1.1.1.2 Symptoms, Disease Progression, Prevention, and Treatment**

RABV which is known to be transmitted by bats, are commonly presented with tremors and involuntary twitching/jerking (myoclonus), again, RABV which is transmitted by dogs, they exhibit, with classical hydrophobia and aerophobia[15]. Furthermore, between these two types, data shows that, encephalitic rabies form is more common and can be observed in approximately 80% of patients, of which between 50 – 80%

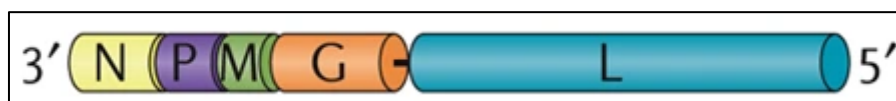
present with the classic symptoms such as aerophobia and hydrophobia – symptoms that are unique to rabies[16], [17]. Encephalitic rabies usually leads to a paralytic form of rabies, which manifests with prominent muscle weakness in the early stages of illness [17]. Rabies symptoms were observed as the cause and effect of large-scale neuronal cell death, but in case of low pathogenicity strains, neuronal apoptosis is the only stimulating factor during the infection[18],[19]. In contrast, neuronal cell dysfunction was thought to be one of the symptoms present during the infection[18],[20],[21],[22], [23], one of the possible reasons for this are due to production of increased Nitric Oxide (NO) by iNOS or inducible nitric oxide synthase in macrophages and neuron .This condition leads to mitochondrial dysfunction and as a result, axonal swelling due to increased levels of NO produced by iNOS[24],[25], thus theoretically explaining the encephalitic symptoms developing in the brain. Rabies-infected patients showed approximately 41% in case of paralytic version of rabies, which is more than that of with encephalitic rabies [15], [26], in spite of this the incubation period for both forms remains the same, which is from 2 weeks to a couple of months although in **one exceptional case** where the incubation period has been documented for more than a year and even to a 8 years[27],[28]. Even though with much advancement in the field of medical sciences, once the clinical symptoms started manifesting, till this day, there is no treatment available to combat the rabies virus. But, treatments which are available in forms of pre- and post-exposure prophylaxis are still limited to certain groups of *Lyssaviruses* which are *Lyssavirus* Phylogroup I[European bat *Lyssavirus*-1 and -2, RABV, Bokeloh bat *Lyssavirus*, Australian bat *Lyssavirus*, DUVV, Aravan *Lyssavirus*, Khujand *Lyssavirus*, Irkut *Lyssavirus*, Taiwan bat *Lyssavirus*, Gannoruwa bat *Lyssavirus* (GBLV)]. Furthermore, based on experimental evidence which suggests that the current available vaccines are not effective against Phylogroup II (LBV, MOKV, SHIBV) or phylogroup 3 *Lyssaviruses* (IKOV, West Caucasian bat *Lyssavirus*, Lleida bat *Lyssavirus*)[27], [29],[30],[31],[32].The viral particles tends to stimulate poorly and restrains the dendritic cells from maturation and activation, due to this a poor categorical immune response is generated which is not sufficient for carrying out the actual clearing processes of the virus from the host system[33]. The P protein is responsible for inhibiting JAK-STAT signalling, thus further inhibiting the interferon (IFN) based autocrine feedback loop; by this mechanism the prevention of maturation of Dendritic Cells is achieved[34].



## 1.2 Genome and Protein

### 1.2.1 GENOME

The non-segmented, negative-sense RNA of *Lyssavirus* consists of around ~12 Kilobases (**Figure 1.1**), but evidence has also highlighted the fact that, variations in the genome length across the genus [35]. The viral genome is known to harbouring based on - two objectives primarily, which are **i**) to generate a **full-length positive sense anti-genome RNA strand**, the precursor RNA molecule for the generation of nascent negative-strand RNA genomes **ii**) the process of the **five viral messenger RNAs** (mRNAs) transcription [36]. According to the NCBI database, there have been 18 completed genome sequences available. The accession number for the genomes are NC\_009527, \_NC\_009528, \_NC\_031988, \_NC\_018629, \_NC\_031955, \_NC\_025408, NC\_020808, \_NC\_003243, \_NC\_025251, \_NC\_025377, \_NC\_020810, \_NC\_020809, NC\_025385, \_NC\_020807, \_NC\_006429, \_NC\_001542, \_NC\_025365, \_NC\_055474.



**Figure 1.1: Genomic structure of rabies (Adapted from [35])**

### 1.2.2 Protein

There are 5 types of protein seen in a *Lyssavirus*, which are:  
**G-protein**

G protein or glycoprotein of the *Lyssavirus* had seen to form a trimeric type spike on the viral partial surface. The N-terminal based domain of G tends to extend outwards on the lipid-based envelope and the C-terminal based domain of G inserts under the virion envelope where it helps with M to produce a complete virus[37].

The viral infiltration to the host cells is based on the interactions of G protein and neuro-specific based receptors. Data suggests that 4 types of receptors had been found in case of RABV [38],[39].The first known binding receptor for RABV was known as **nAChR**[40], which had been observed to co-localize with RABV(CVS) in the region on neuromuscular junctions [41]. Again, nAChR, is responsible for concentrating virus at the sites in proximity to the peripheral nerves, which is responsible for enhancing the viral proliferation from the peripheral nerve to the brain[40]. Furthermore, another receptor for the G, is NCAM which has been proposed by Thoulouze et al[42],data

suggests that, the natural ligand or specific antibodies against NCAM significantly decreased in RABV (CVS strain) infection in vitro, soluble NCAM could neutralise the infectivity of RABV for susceptible cell lines, and rabies mortality was delayed in NCAM -deficient mice [42]. The p75NTR was identified as another ligand for RABV G protein, through the analysis of a cDNA library prepared from a murine neuroblastoma cell line [43]. Using a reverse binding assay, p75NTR interacted with G of certain *Lyssavirus* species, including RABV (wild -type, CVS, and PV strains) and EBLV -2, while no interaction was detected for other studied species. This observation clearly implies the usage of alternative receptor(s) by different *Lyssaviruses* [44] and may justify differences in their pathogenicity and neuro-invasiveness pathway [45],[46],[47]. In vivo studies have indicated the main distribution of p75NTR in the dorsal horn of the spinal cord, due to this, it is suspected that RABV G -p75NTR interaction may play a role in retrograde axonal trafficking of RABV particles in the CNS [48]. Apart from participating in viral entrance, G has been demonstrated to have ability to target the neuronal enzymes by its PDZ -BS, which mimics the PDZ domain of neuronal enzymes. Such interference in infection by the virulent strains of RABV ends with cell survival, while with vaccinal strains ends with neuron death[49], displayed that the G protein of the virulent strain bound to the PDZ domain of MAST2 and inhibited the controlled phosphorylation of PTEN by MAST2. They revealed that the dephosphorylation of PTEN changed its intracellular localization, stability, and activity, leading to altered neuronal homeostasis and neuro-survival [49]. In one study on the network of RABV gene products implicated in rabies using a systems biomedicine approach, authors proposed that G prompted the hyperactivation of PI3K -AKT signalling through the dephosphorylation and redistribution of PTEN. The consequences of the activation and the downstream signalling of AKT could reduce apoptosis or cell survival [50]. On the other hand, G of the vaccinal strain is bound to the PDZ domain of MAST2 and other cellular partners, particularly PTPN4, an anti -apoptotic protein. This interaction suppresses the efficient dephosphorylation of ligand(s) by PTPN4. Therefore, the homeostasis of the infected neuron alters, and apoptosis signalling is triggered [49]. The G protein of RABV (CVS -11, SAD strains) also interacts with SNAP25, a member of the SNARE complex that mediates membrane fusion events. Knockdown of SNAP25 has shown an inhibitory effect on the release of RABV in nerve cells. [51].

## **N protein**

The virion-associated RNA polymerase has N, P and L, which is responsible for the forming of the RNP core, RNP core has polymerase activity. N contributes by suppressing the innate immune response of the host's immune system, thus making the virus enable to replicate and proliferate properly in both brain and CNS. Host defence related genes, IFN and chemokines are suppressed by avoiding the activation of the RIG-I through inhibiting the activation of the IRF-3 pathway [52]. Data indicates that, CVS strain's N protein targets Hsp70 chaperone for binding, it is known for positive regulation of the RABV infection cycle at different stages, such as the transcriptional and/or translational level and/or viral assembly and budding [53]. Upregulation of Hsp70 and its accumulation in NBs, along with its presence in both purified nucleocapsid and virions, have been demonstrated. Downregulation of Hsp70 for elucidating the functional role of N-Hsp70 association using RNAi revealed a decrease in the viral mRNA, proteins, and particles. Moreover, P is recruited by N thus creating N-P complex which brings CCT $\gamma$  to NBs, they act as a main base for virus replication and chaperonin facilitates viral transcription and replication, in general, although more data is needed for this mechanism [54].

### **L protein**

L protein possesses an RdRp activity. By the transcriptase, capping, and polyadenylation activity L protein is responsible for the transcription of viral mRNAs. Replicase enzyme contribute to the replication of the viral genome. For the viral replication and transcription purpose L protein is interacted by its cofactors P, in RNP core and creating the formation of L-P is necessary [55]. Genomic RNA synthesis is caused by binding of L to N protein [56, 57]

### **P protein**

P protein contributes to the central role in viral transcription and replication [55], also, functioning as a host innate immune antagonist. Rpl9, a type of ribosomal protein with translational function, has been proposed to play a role for RABV for escaping the immune responses. P induces the translocation of Rpl9 from the nucleus to the cytoplasm in the first stages of the infection, interacting with this ribosomal component. With the overexpression of L9, replication of RABV decreases, thus, by knocking down the expression of L9, the RABV replication enhances, thus it can be said that L9 interferes with RABV replication in the early stages [58]. Autophagy, is a host defence mechanism by which intracellular pathogens are removed, it has also been found that Incomplete autophagy is another mechanism of immune evasion, which has been

induced by virulent and attenuated RABV (CVS-11, HEP-Flury strains) P and P5 isoform [59], [60]. It has been demonstrated that RABV P interacts with BECN1 and induces incomplete autophagy through activating **BECN1-CASP2-AMPK-MAPK** and **BECN1-CASP2- AMPK-AKT-MTOR** signalling pathways, which enhance the viral replication. The autophagosome, which has engulfed virions, does not fuse with the lysosomes, and lastly, virions escape degradation[60].

### **M protein**

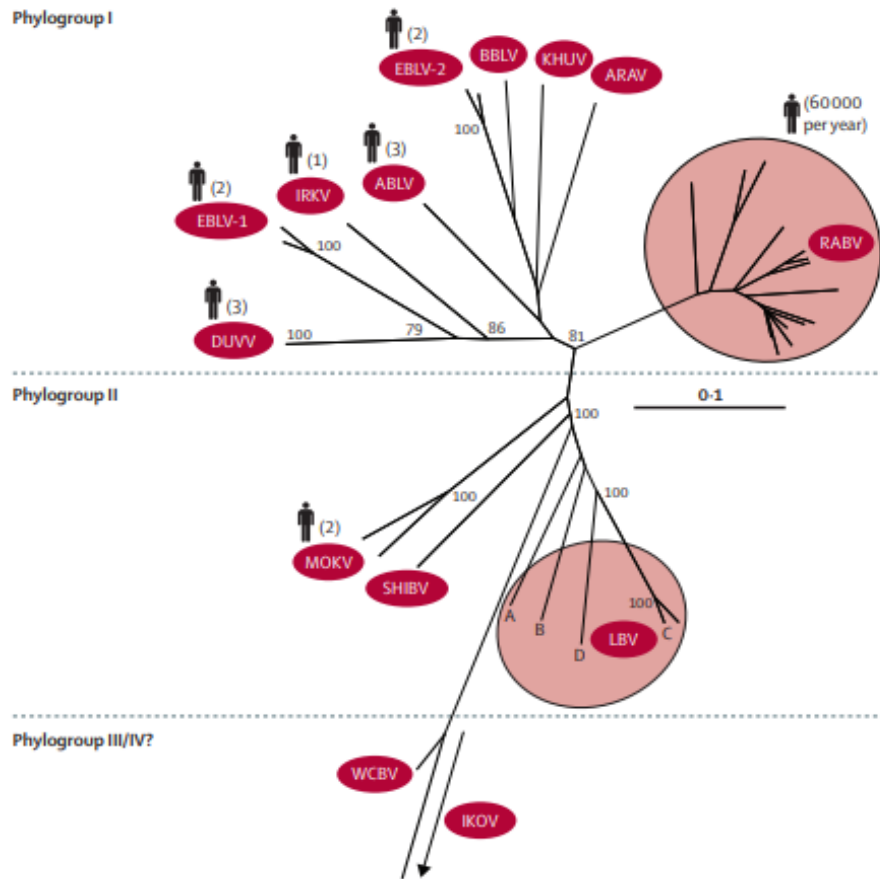
Most abundant and smallest protein in the *Lyssavirus* virion is M protein though it has a multifunctional role [61]. This particular protein is known for assembly/budding and regulating the balance between transcription and replication of the virus through direct or indirect interaction between L and M are the primary functions of M [55]. Furthermore, host gene downregulation can be observed because of M protein [61], apoptosis [23], modulation of host innate immune defence, and virion uncoating [62], [63]. M binds with eIF3h, creating the regulation of the cellular translation initiation. Again, M protein contributes to the low pathogenesis of Mokola virus (a *Lyssavirus* of low pathogenicity) by targeting mitochondria via interaction with the terminal component of the mitochondrial respiratory chain, Cco1. Interaction of M-Cco1 significantly decreases Cco1 activity and ATP level of neurons, due to this, in mitochondrial morphology and function disruption and subsequent apoptosis [23]. M also takes part in the subversion of the host innate immune defence through different mechanisms. NF- $\kappa$ B pathway plays a major role in the regulation of the immune response to infection. Mokola virus targets host mitochondria by binding with the terminal component of the mitochondrial respiratory chain, also known as Cco1. RABV entry into the host cell, the endosome-containing virus becomes acidic, and the conformation of RABV G is changed to stimulate virus-endosome membrane fusion. Then RABV M proteins dissociate and release viral nucleocapsids to cytoplasm[55].

### **1.3 Phylogroups of *Lyssaviridae***

The transmembrane Glycoprotein encoded gene is involved in the virus-host interaction, immunogenicity and pathogenicity which was used to assess the genetic diversity of representative members of the rabies and rabies related *Lyssavirus* genus. By phylogroup analysis seven genotypes were identified and these genotypes were grouped into two major phylogroups (Phylogroup I and Phylogroup II) and another

newly recognized phylogroup III [64, 65]. According to the genotypes, the European bat *Lyssavirus* (EBL) genotypes 5 (EBL1) and 6 (EBL2), the African genotype 4 (Duvenhage virus), and the Australian bat *Lyssavirus* genotype 7 make up Phylogroup I. The global genotype 1 (classic Rabies virus) is also included in phylogroup I. The divergent African genotypes 2 (*Lagos bat virus*) and 3 make up Phylogroup II (Mokola virus) [64]. So, Phylogroup I includes Rabies *Lyssavirus*, Duvenhage *Lyssavirus*, European bat *Lyssavirus* type 1 and 2, Bokeloh bat *Lyssavirus*, Australian bat *Lyssavirus*, Aravan, Khujand, and Irkut *Lyssavirus*. *Lagos bat Lyssavirus*, *Mokola Lyssavirus*, and *Shimoni bat Lyssavirus* are all members of Phylogroup II. Independent phylogroup 3 is made up of the West Caucasian bat *Lyssavirus*, the Ikoma *Lyssavirus*, and the Lleida bat *Lyssavirus*. Significant serological neutralization has been found within phylogroups, while very little cross neutralization has been found between phylogroups. As a result, effective cross-protection against all genetically diverse *Lyssaviruses* may not be provided by rabies virus vaccines. With pre-exposure vaccination and conventional rabies post-exposure prophylaxis, little to no cross-protection was seen against *Lyssaviruses* of phylogroups 2 and 3 [66]. In addition to these three phylogroups, the recent isolation of the Ikoma virus (IKOV) in a clinically rabies African civet in Tanzania's Serengeti National Park has significantly increased the genetic diversity of the *Lyssavirus* genus [65].

Nucleoprotein sequences (405 nucleotides) were aligned with ClustalW and the phylogenetic tree was visualised using TreeView version 3.2. Bootstrap values at relevant nodes are shown. According to the proposed antigenicity of each group of isolates, the viruses are divided into different phylogroups. Several sequences within the phylogeny are unpublished and as such do not have accession numbers. The scale bar represents 0.1 substitutions per nucleotide site. The number of human cases are shown next to silhouettes where reported [67].



**Figure 1.2: Phylogenetic tree of the *Lyssavirus* phylogroups and their respective species. (Adapted from [67])**

**Table 1.1: Current diversity and taxonomy of lyssaviruses ( Adapted from ICTV = International Committee on Taxonomy of Viruses [66])**

Phylogroups	Approved Species (ICTV) <sup>a</sup>	Virus	Potential vector(s)/reservoirs	Distribution
Phylogroup 1	<i>Rabies Lyssavirus</i>	Rabies virus (RABV)	Carnivores (worldwide); bats (Americas)	Worldwide (except several islands)
Phylogroup 2	<i>Lagos bat Lyssavirus</i>	<i>Lagos bat virus</i> (LBV)	Frugivorous bats (Megachiroptera)	Africa

<b>Phylogroups</b>	<b>Approved Species (ICTV)a</b>	<b>Virus</b>	<b>Potential vector(s)/reservoirs</b>	<b>Distribution</b>
			)	
Phylogroup 3	<i>Mokola Lyssavirus</i>	Mokola virus(MOKV)	?	Sub-Saharan Africa
Phylogroup 1	<i>Duvenhage Lyssavirus</i>	Duvenhage virus (DUVV)	Insectivorous bats	Southern Africa
Phylogroup 1	<i>European bat 1 Lyssavirus</i>	European bat 1 <i>Lyssavirus</i> (EBLV-1)	Insectivorous bats (Eptesicus serotinus)	Europe
Phylogroup 1	<i>European bat 2 Lyssavirus</i>	European bat 2 <i>Lyssavirus</i> (EBLV-2)	Insectivorous bats (Myotis daubentonii, M. dasycneme)	Europe
Phylogroup 1	<i>European bat 2 Lyssavirus</i>	Australian bat <i>Lyssavirus</i> (ABLV)	Frugivorous/insectivorous bats (Megachiroptera /Microchiroptera )	Australia

<b>Phylogroups</b>	<b>Approved Species (ICTV)a</b>	<b>Virus</b>	<b>Potential vector(s)/reservoirs</b>	<b>Distribution</b>
Phylogroup 1	<i>Aravan Lyssavirus</i>	Aravan virus (ARAV)	Insectivorous bats (Myotis blythi)	Central Asia
Phylogroup 1	<i>Khujand Lyssavirus</i>	Khujand virus (KHUV)	Insectivorous bats (Myotis mystacinus)	Central Asia
Phylogroup 1	<i>Irkut Lyssavirus</i>	Irkutavirus (IRKV)	Insectivorous bats (Murina leucogaster)	East Siberia
Phylogroup 3	<i>West Caucasian bat Lyssavirus</i>	West Caucasian bat virus (WCBV)	Insectivorous bats (Miniopterus schreibersi)	Caucasian region
Phylogroup 2	<i>Shimoni bat Lyssavirus</i>	Shimoni bat virus (SHBV)	Hipposideros commersoni	East Africa
Phylogroup 1	<i>Bokeloh bat Lyssavirus</i>	Bokeloh bat Lyssavirus (BBLV)	Insectivorous bats (Myotis nattereri)	Europe
Phylogroup 3	<i>Ikoma</i>	Ikoma virus	? (isolated from	Africa



Phylogroups	Approved Species (ICTV)a	Virus	Potential vector(s)/reservoirs	Distribution
	<i>Lyssavirus</i>	(IKOV)	Civettictis civetta)	
Phylogroup 1	<i>Gannoruwa bat Lyssavirus</i>	Gannoruwa bat <i>Lyssavirus</i> (GBLV)	Isolated from Pteropus giganteus	Asia
Phylogroup 3	<i>Lleida bat Lyssavirus</i>	Lleida bat <i>Lyssavirus</i> (LLEBV)	Insectivorous bats (Miniopterus schreibersi)	Europe (Spain)

### 1.3.1 Phylogroup I

According to the data on comparative animal pathogenicity, serology, and genetic distance, the phylogroup I contains RABV, DUVV, EBLV-1, EBLV-2, and ABLV. Members of phylogroup I include the putative species ARAV, KHUV, and IRKV. IRKV is a member of the same clade as DUVV and EBLV-1 in phylogroup I. While ARAV occupies an intermediate position, demonstrating relatedness to both KHUV and EBLV-2 as well as the EBLV-1/DUVV clade, KHUV is mostly related to EBLV-2 based on the absence of serologic cross-reactivity and genetic distances [68]. Phylogroup I also includes the additional prototype *Lyssaviruses* Gannoruwa bat *Lyssavirus* (GBLV), Bokeloh bat *Lyssavirus* (BBLV), Taiwan bat *Lyssavirus* (TWBLV), and Kotalahti bat virus (KBLV) [69]. The United Kingdom (UK) and the Netherlands are the only major countries where EBLV-2 is shown to be connected with Daubenton's bats (*Myotis daubentonii*) and to a lesser extent with pond bats (*Myotis*

dasycneme). It is also present in Finland, Germany, Switzerland, and, more recently, Norway. The expansion of research into bat *Lyssaviruses* has resulted in the discovery of novel *Lyssaviruses* in European bat populations. Bokeloh bat *Lyssavirus* (BBLV), a novel *Lyssavirus*, was reported in a Natterer bat (*Myotis nattereri*) in Germany in 2010. Within six years, BBLV was also isolated five more times in Germany, twice in France (in 2012 and 2013) and once in Poland (in 2016). Except for one instance that was isolated in a common Pipistrelle bat (*Pipistrellus pipistrellus*) in Germany, all cases of BBLV were isolated from Natterer's bats [70]. Due to the widespread human rabies, Phylogroup I is of special epidemiological importance. The World Health Organization estimates that RABV causes 60,000 fatalities annually, with Asia and Africa reporting the bulk of infections. Phylogroup I's genetic diversity exhibits a distinct geographic distribution. In its most basic form, RABV is made up of two main lineages: the first circulates in terrestrial mammals with a global distribution, while the other lineage has only been discovered in the New World. Particularly in North America, where multiple recorded cross-species transmissions occurred, the history of RABV has been intensively examined [65].

### **1.3.2 Phylogroup II**

Phylogroup II of *Lyssavirus* consists of genotype 2 and genotypes 3 species. The genotype 2 includes LagNGA, LagCAR, LagSAF1, and LagSAF2. And genotype 3 includes MokSAF, MokETH, and MokZIM [64]. *Shimoni bat Lyssavirus* (SHIBV), *Mokola Lyssavirus* (MOKV), and *Lagos bat Lyssavirus* (LBV) are all members of Phylogroup II [69]. The second-most diverse clade, Phylogroup II, contains these three species that only occur in Africa. One of the few *Lyssavirus* species that has not been isolated from bats is the Mokola virus (MOKV), which infects mammals in sub-Saharan Africa [65]. Only the intracerebral pathway made the Mokola and *Lagos bat* viruses of phylogroup II pathogenic. Also it could appear to be less dangerous for human and veterinary health due to their widespread distribution in Africa, their decreased pathogenicity in mice, and the small number of human cases and animal epizootics reported thus far. But between 1995 and 1998, the Mokola virus regularly appeared in South Africa, and its reservoir is still a mystery. Furthermore, it should be noted that phylogroup II exhibits high genetic diversity while having very few isolates, which implies even higher levels of diversity in nature. Phylogroup II's higher genetic

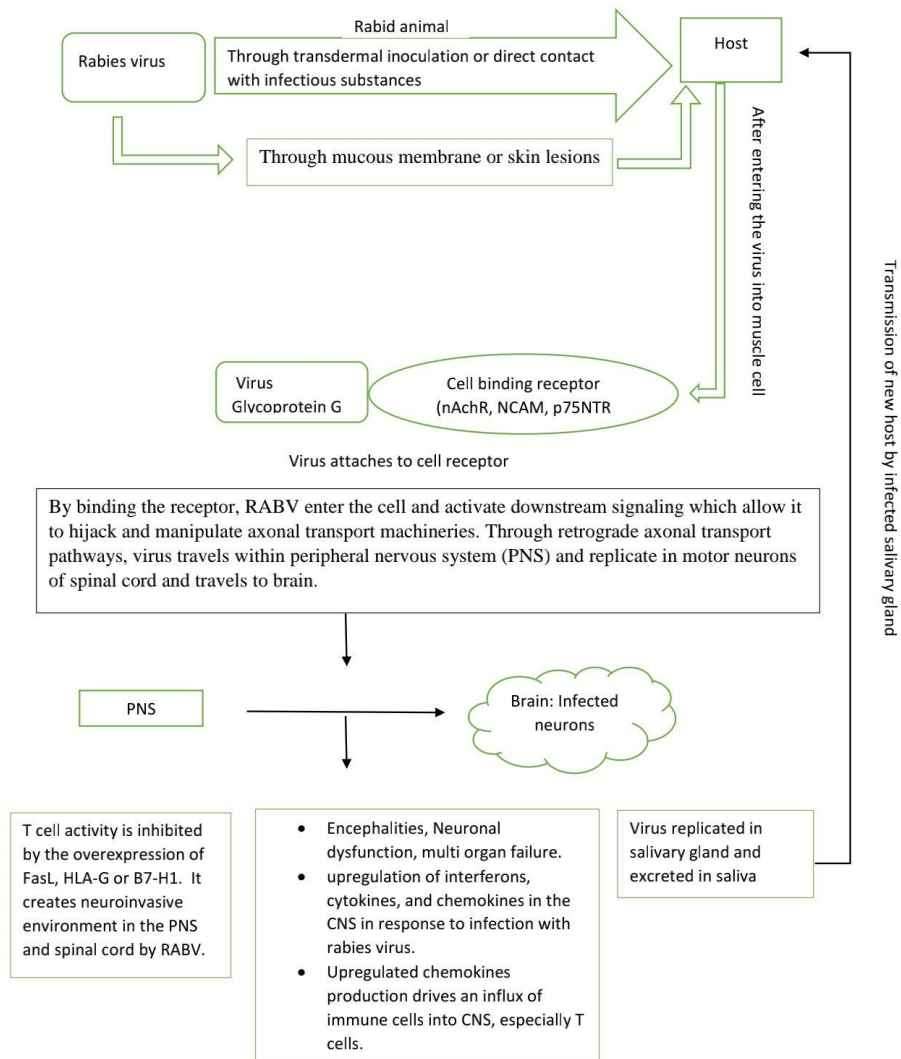
heterogeneity might offer molecular flexibility. On the one hand, it could explain why numerous animals protected against challenge with *Lagos bat virus* (genotype 2) by animal vaccinations (genotype 1: PV and USA7-BT) [64].

### 1.3.3 Other Species

The third group of phylogenetic tree of *Lyssavirus* is Phylogroup III (**Figure 1.2**). This phylogroup containing West Caucasian bat *Lyssavirus* (WCBV), Ikoma *Lyssavirus* (IKOV), and Lleida bat *Lyssavirus* (LLBV) are the three *Lyssaviruses* with the greatest genetic divergence [69]. The geographical range of the genetically diverse West Caucasian bat virus (WCBV), which was initially discovered in southeastern Europe, has since been enlarged as a result of the discovery of WCBV in *Miniopterus* bats in Kenya. The genetic diversity of *Lyssavirus* has significantly increased as a result of the discovery of the Ikoma virus (IKOV) in a clinically rabies African civet in Tanzania's Serengeti National Park. The IKOV, which wasn't found in bats, hasn't been grouped with any of the currently recognized phylogroups.

The inclusion of the IKOV and WCBV in phylogroup III is supported by the possibility that they form a monophyletic group. According to the *Lyssavirus* phylogeny, the species is monophyletic. The same key phylogenetic relationships were visible in both small and large datasets. The IKOV and WCBV's evolutionary connection stood out as an outlier. IKOV was the sister species of WCBV (aLRT 55%, BPP 97%) in the small data set, however in the large data set, IKOV was the sister group of the remaining *Lyssaviruses* (aLRT 72 percent, BPP 97 percent). Despite having genetic differences from phylogroups I and II, the IKOV + WCBV clade induces encephalitis with characteristics like RABV [65].

## 1.4 Pathogenicity:



**Figure 1.3: Pathogenesis of Rabies virus into host.**

### 1.4.1 Factors related to the pathogenicity

#### The binding receptors:

Both the Neuronal Cell Adhesion Molecule (NCAM) and the p75 neurotrophin receptor (p75NTR) have been identified as RABV glycoprotein G binding receptors. Other membrane-associated components have also been implicated in RABV binding [71].

#### a) p75NTR:

- Involved in the retrograde transport of neurotrophic factors, little is known regarding its direct contribution to viral transport.
- RABV may enter the cell by receptor mediated endocytosis following its binding to p75NTR, after it enhances the efficiency of retrograde co transport of RABV-p75NTR complexes.
- Interaction with p75NTR modulates the cellular transport machinery and facilitates movement of RABV to the CNS [71].
- It is a low affinity nerve growth factor receptor, also known as BeX3 and NGFR.
- The function of p75NTR in rabies virus entry is less clear[72]

**b) NCAM:**

- Another potential receptor for rabies virus include neuronal cell adhesion molecule (NCAM, also known as NCAM1).
- Rabies virus enters the neurons using NCAM [72].

**c) Nicotinic acetylcholine receptor (nAChR):**

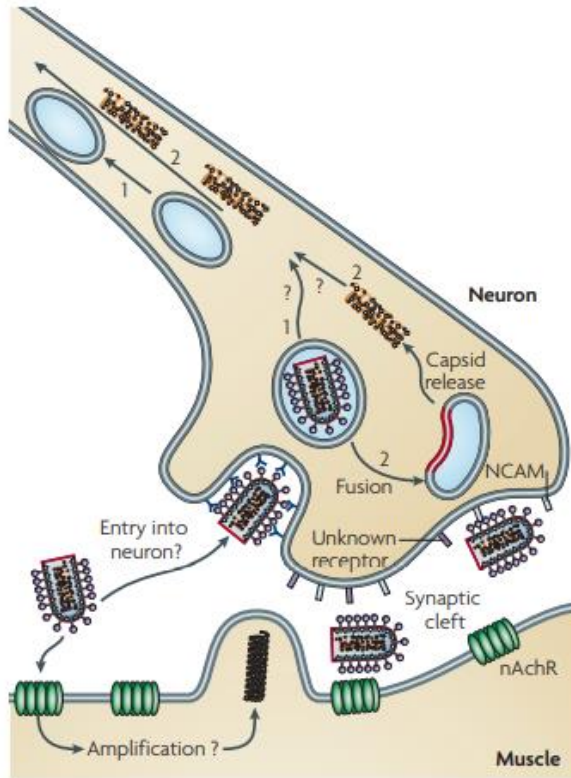
- This is the first identified potential receptor for rabies.
- Located at the postsynaptic muscle membrane and not at presynaptic nerve membrane (**Figure 1.4**). It is unlikely that this receptor is used for the initial entry into motor neurons. Instead, nAChR possibly enriches the rabies virus at the neuromuscular junction or synaptic cleft, which makes it easier to cause infection efficiently to the connected motor neurons.
- Other research suggests that initial rabies virus replicates in muscle cells, that indicated that muscle cells may be infected via nAChRs [72].

**TLR3:**

- Upregulated in rabies encephalitis.
- Negri Bodies formation
- Induction of proinflammatory responses [72].

IFN $\beta$  expression greatly reduces rabies virus pathogenesis and viral replication but not its immunogenicity.

Interferons (IFNs), tumor necrosis factor-alpha (TNF-alpha), and Toll-like receptors (TLRs) are the only innate immune responses that can stimulate B7-H1 production[73].



**Figure 1.4: Pathway of Nicotinic acetylcholine receptor (nAChR) from Muscle to Neuron (Adapted from [72])**

## 1.5 Available and Designed Vaccines

Several types of vaccine are **either** available **or** in trial phase (Table 1.1) to combat against rabies vaccine, some of them are provided below[74],[75],[76],[77].

**Table 1.2: Available & Designed Vaccine List**

Vaccine Type	Vaccine Description	Reference
Alternative Development based vaccine	Rabies Virus glycoprotein has been expressed on the surface of the vaccinia virus	[78]
	Experiments has shown rabies glycoprotein expression on nary pox virus's surface	[79]
	A chimeric <i>Lyssavirus</i> glycoprotein with segments derived from RABV and Mokola virus that provide immunization against more than one <i>Lyssavirus</i>	[80]

Vaccine Type	Vaccine Description	Reference
	DNA vaccination with RABV glycoprotein cloned into a plasmid vector	[81]
Reverse Genetics based approach	An attenuated, fixed strain of RABV, SAD B19, a European derivative of an American SAD strain was recovered from a plasmid-encoded genome by Conzelmann and Schnell.	[82]
Vesicular Stomatitis Virus-based rabies vaccine	Although there have not been proper studies, for VSV being engineered for RABV G protein for immunogenic protection, VSV might be an ideal candidate in case of rabies control	[36]
Parainfluenza Virus Type 5 (PIV5)-Based Rabies Vaccines	rPIV5-RV-G has shown a promising immune response against rabies virus a dose-dependent protection type characteristics were shown when rPIV5-RV-G was introduced into intranasal route (single dose of 106.0 PFU, achieving 100% success rate), into mice (108.0 PFU, with 90% protection success rate) as well	[83]
Newcastle Disease Virus (NDV)-Based Rabies Vaccines	Because of the host range limitation, the NDV is only limited to avian based immune system but not to mammalian immune system, thus making it a safe candidate With the help of recombinant NDV system, study suggest that protection can be generation against SARS, HRSV, and influenza viruses	[84] [85],[86],[87]
Open Reading Frame Virus (ORFV)-Based Rabies Vaccines	apathogenic ORFV strain, which is D1701-V-RabG, has shown that without the replication assistance it can present the RABV G protein on the surface of the infected cells With a single inoculation of 107.0 PFU of D1701-V-RabG, data suggests that mice being protected from	[88]

Vaccine Type	Vaccine Description	Reference
	aggressive strain of rabies strain CVS-11 in high lethal amount	
Vaccinia Virus-Based Rabies Vaccines	Copenhagen strain, a first recombinant poxvirus authorized to use as a vaccine, containing a RABV G gene, which has been incorporated into poxvirus thymidine kinase gene (V-RG)	[89]
	Experiments and data have suggested that bait based orally administered rabies vaccine is the most effective way and best strategy for dealing wildlife and domestic animal, furthermore, V-RG and RABV SAG-2 strains has been used as 2 vaccines	[90],[91]
	V-RG has been associated with severe skin inflammation in humans who have occasional contact with the baits	[92]
	Data suggests that MVA which expresses RABV G gene, is less effective than V-RG	[89]
AcMNPV	By the control of AcMNPV polyhedrin and earlier version of the CMV promoters a recombinant AcMNPV which is expressing G gene of the RABV has been created	[93]
	Mice which were intramuscularly injected with BV-RVG/RVG, they expressed high levels on Virus Neutralizing Antibodies	
Protein Subunit and Peptide Vaccine	In a dosage-based manner, G proteins peptide mimotopes were like antigenic site III of the G protein; they also interact with the human version of anti-RABV IgG while administering with a dosage dependent manner.	[94]
	Due to the high variable of the G protein, there is a chance of peptide vaccine to be compromised, thus another novel approach would be targeting the RABV P protein	[95]



<b>Vaccine Type</b>	<b>Vaccine Description</b>	<b>Reference</b>
Nucleic Acid-Based Rabies Vaccines	Disease where traditional vaccines are ineffective, Nucleic acid-based vaccines has the potential to induce a cellular and humoral immune response	[96],[97],[98] [99]
	Steadiness, low processing cost, easy to develop and scaling up advantages are the main advantages of the nucleic acid-based vaccines	[100]
	Controlled by the SV40 early promoter vector, a rabies only immune response was tested on mice, using a plasmid name pSG5rab.gp, that expressed G protein of rabies, where mice model showed RABV G-specific cytolytic T cells, lymphokine-secreting T helper cells of the Th1 subset and rabies VNAs	[101]
	Against wild-type RABV and to some cases against EBL1 and EBL2, high VNA titers were observed when Beagles were intramuscularly injected by pGPV	[102]
	Mice developed comparable cellular based immune and humoral based immune system than commercial cell culture vaccines after injected with replicon-based rabies DNA vaccine	[103]
	Chemical adjuvants have been proposed to increase the immunogenicity and efficacy of DNA vaccines	[104],[105]
	self-amplifying mRNA vaccines and conventional non-amplifying mRNA vaccines are two types of RNA vaccines which have been constructed based on the auto-replicative capacity of messenger RNA (mRNA)	[106]
Small Interfering RNA (siRNA)-Based Therapy	siRNA or short interfering RNA consisting of double stranded 21–23 bp in length, which is responsible for interfering within the expression of specific genes by degrading mRNA after transcription	[107],[108]

Vaccine Type	Vaccine Description	Reference
	Another possible approach, which has been derived from G protein of RABV, short 29-amino-acid peptide, which has the binding capability with acetylcholine receptor, expressed by neuronal cells, thus allowing trans-vascular delivery of siRNA across the BBB to the brain	[109]
RABV-Specific	RIG a mCAB, targets specific epitopes of G protein has shown success in neutralizing rabies in RABV	[110]
Immunoglobulin in (RIG) Coupled with BBB Permeability-Enhancing Agents	hyperosmotic solution, or cytokine MCP-1, can be used as a backup method, for enhancing BBB permeability resulting in antibodies to go through BBB	[111]
Bi-Specific Antibody (BsAb)-Based Therapy	<p>in the field of biology, BsAb is another new futuristic therapy, where 2 different epitopes of an antigen are aimed to bind</p> <p>BsAb can redirect immune effector cells, which has shown in case of tumor samples, thus enhancing the tumor-killing potentials, furthermore, prototypic form of BsAb is generally made of two linked single-chain fragment variables with one targeting molecule present on immune effector cells, such as the CD3 found on T cells, while the other targets a tumor-specific antigen</p> <p>Although BsAb has not been tested against rabies furthermore, BBB permeability-enhancing agents can cause unwanted neurological complications</p>	[112]

Vaccine Type	Vaccine Description	Reference
	Targeting molecular receptors via acting as molecular Trojan horse to ferry biologics, merging the antibody into the brain via RMT (receptor-mediated transcytosis), here one arm can be used as targeting one of the endogenous BBB receptors while the other targets the RABV G protein. could be a novel process.	[113], [114],[115]
Improve Vaccine	After administering a single dosage, it is expected that PeEP vaccine would produce VNA titers	[116]
Vaccine Suited for PEP	<b>Adjuvanted Rabies Vaccines</b>  Inflammatory response which is important for antigen-driven stimulation for naive B and T cells, adjuvants enhance this inflammatory response.	[117],[118]
	<b>Protein Vaccine</b>  The objective of creating an effective protein-based rabies vaccine is complicated due to virus glycoprotein forming trimers on the virions and furthermore most VNA's tent to bind with the conformation-dependent epitopes.	[119]
	<b>Genetically Modified, Inactivated Rabies Virus</b>  Reverse genetics can be another approach to modify rabies virus, genes coding for phosphoprotein or the matrix protein, deleting them has made the virus a-pathogenic, even for mouse who are immunocompromised	[120],[121]
RNA Vaccine	An RNA vaccine expressing the rabies virus glycoprotein was tested in a phase I dose escalation study in rabies	[122]

Vaccine Type	Vaccine Description	Reference
	virus-seronegative human volunteers between 18–40 years of age	
Viral Vector Vaccine	To initiate adaptive immune responses the inflammatory responses needed to be evoked which can be conducted by PAMP carried by viral vectors	[123]
	For immunization purpose on wild-live animals, a vaccinia virus recombinant, expressing rabies virus glycoprotein also known as VR-G has been authorized, the basic reason is viral vectors, after binding with the cell surface receptors are more efficient than DNA vaccines in cases of immunogenic responses	[124]
	Research further shows that VR-G possess inadvertent infection and reactogenic, ORNAB, oral vaccine was created for foxes, raccoons type animals, which expressed rabies glycoprotein, after observing that orally administering was immunogenic in its target species, another study was conducted with VR-G	[125],[92], [126]
	A replication defective AdC serotype 68 (AdC68) vectors expressing the same antigen was equally effective in mice	[127]
	Replacing the AdC68 E4 open reading frames (ORF) 6 and 7 with those of HAd-V5, the vector has been modified for the purpose of increasing the vector yield on HAd-V5 E1.	[128]
Protein Vaccine	Mammalian expression systems based on human embryonic kidney (HEK) 293, baby hamster kidney (BHK)-21, or Chinese hamster ovary (CHO) cell lines have been tested with varied success; they showed distinct patterns of glycosylation depending on cell substrate and culture conditions.	[129],[130], [131],[132]

<b>Vaccine Type</b>	<b>Vaccine Description</b>	<b>Reference</b>
Adjuvanted rabies Vaccine	Alum has not shown indication of increasing immune response in case of rabies virus vaccine.	[133]
	Relatively better performance has been observed by the ligands for TLR which has been based on Second generation adjuvants	[134]
	PIKA vaccine with the composition of TLR-3-activating adjuvant, polyinosinic-polycytidylic S acid based and Rabipur has finished the phase 2 clinical based trial	[135]
GM Rabies Vaccine	With the help of genetic engineering Rabies can be modified, such as with the attenuated virus where matrix protein and phosphoprotein encoding genes, can be omitted, again, rabies virus with deleted-Matrix protein, is more capable of immunogenic characteristics than an inactivated wild-type virus or deleted phosphoprotein-based rabies.	[136],[121],[137],[138]
Genetic Vaccine	RNA vaccines, DNA vaccines and viral vector vaccines are the subclasses of the Genetic based vaccine, where genetic components are used instead of protein-based antigen. In case of viral vectors which are known to induce B and T cell based immune response whereas DNA and RNA based vaccines are only known for encoding rabies virus-based antigen. In specific cell surface receptors, the viral vectors bind.	[139],[140]
RNA Vaccine	A clinical phase I trial with an RNA rabies vaccine expressing the rabies virus glycoprotein, termed CV7201, tested a three-dose regimen in healthy adult volunteers. The vaccine was well tolerated but protective titers of antibodies were not achieved in all vaccinated individuals, titers declined by one year after vaccination	[122]

Vaccine Type	Vaccine Description	Reference
	and upon an additional boost failed to increase in all participants	
DNA vaccine	Trials in taken in multiple phase I and II with rabies virus has shown success	[141]
Viral Vector Vaccine	The safest and most immunogenic of the different types of viral vector vaccines are based on E1-deleted adenoviruses	[142]
	Because of the deletion of E1, it decreases the ratio of viral antigen's transcription rate, at the same time, without affecting the expression of the transgene product. Cytotoxic characteristics are not expressed by the adenovirus and furthermore, they are seen providing the immune system constant support for a long time by continuously conducting a low-level transcription. Due to this, our immune system maintains stability in case of adenovirus-based vectors.	[143]
	Adenovirus based vectors in cases of primates (except non-human), in titers analysis, have shown presence of antibodies after a single dose.	[144],[145]

## 1.6 Why Phylogroup II?

The phylogroup II was selected as the target since there was insufficient immunogenic protection against it. Additionally, this phylogroup does not participate in any industrialized immunization program. The inactivity of earlier vaccinations against phylogroup II was significantly different according to various investigations and experiments. Also the classical rabies viruses of genotypes 1 or phylogroup I are the most phylogenetically distant from phylogroup II or genotypes 2 (*Lagos Bat Virus*) and 3 (*Mokola Virus*) [146]. The majority of today's commercial rabies vaccinations are

made from the original Pasteur strain, which was identified in 1885, and all of its derivatives, including the Pasteur virus (PV), Pitmann-Moore virus (PM), and challenge virus standard (CVS). These vaccines, particularly HDCV (human diploid cell vaccine) and PVRV (purified vero rabies vaccine) vaccines for human use, have been demonstrated to induce partial to complete protection (20% to 100% depending on the challenge route) against members of phylogroup I *Lyssaviruses* such as EBLV-1, EBLV-2, or BBLV but few/no protection against phylogroup II and III *Lyssaviruses* [70]. The approved rabies vaccinations appear to mostly give protective immunity against phylogroup I *Lyssaviruses* and least potential against IRKV. The level of VNA needed for protection against non-RABV *Lyssaviruses* is unknown. In vivo vaccination-challenge studies have demonstrated that the VNA response produced by RABV vaccines is insufficient to provide protection against challenge for more diverse *Lyssaviruses* LBV, MOKV, and SHIBV are *Lyssaviruses* belonging to phylogroup II. While the current rabies vaccines are secure and, when administered correctly, highly efficient against RABV, they are ineffective against the wildly diverse *Lyssavirus* species, especially those in phylogroup III. Since *Lyssaviruses*' surface glycoprotein G serves as their primary immunogen, research using chimeric G proteins like RABV-MOKV or RABV-EBLV1 has demonstrated that, in the context of an anti-*Lyssavirus* vaccine, it is possible to broaden the cross-neutralization spectrum both within antigenic group I and between antigenic groups I and II [147]. Moreover, genotype 1 virus strains are found in commercially marketed vaccinations. Their range of defense against the various *Lyssavirus* genera varies. The pasteur virus (PV) induces VNABs against European bat *Lyssavirus* (EBL) genotypes 5 (EBL1) and 6 (EBL5), African genotype 4 (Duvenhage virus), and genotype 1 (classic Rabies virus) of the pasteur virus (PV) (EBL2). However, it is unable to prevent against genotypes 2, 3 (phylogroup II). The amino acid sequence of the glycoprotein ectodomain in phylogroups I and II was at least 74% identical, and anti-antiglycoprotein viral antibodies showed cross-neutralization. The lack of cross-neutralization and the low identity between phylogroups explain why the traditional rabies vaccines (phylogroup I) are ineffective against *Lyssaviruses* from phylogroup II [64].

## **Chapter 2. Methodology**

### **2.1 Strain Identification and protein sequence retrieval**

Phylogroup II was selected for the vaccine construction. This phylogroup contains three species: *Lagos bat virus* (LBV), *Mokola virus* (MOKV) and *Shimoni bat virus* (SHBV). There were 20 Glycoprotein (G) sequence of LBV, 23 Glycoprotein sequence of MOKV and Only one unreviewed Glycoprotein sequence of SHBV were retrieved from UniprotKB database [148] on February 2022. All of the protein sequences were around 522 amino acids long, meaning they represented the complete glycoprotein sequence. Among them, 1 UniProtKB reviewed (Swiss-Prot) LBV (Accession no: Q8BDV6), 1 UniProtKB reviewed (Swiss-Prot) MOKV (Accession no: P0C572) and 1 UniProtKB unreviewed (TrEMBL) SHBV (Accession no: D4NRK1) were identified.

### **2.2 Protein sequence alignment**

All 522 amino acid long Glycoprotein sequences of Lagos, Mokola and Shimoni were gathered in a file in a FASTA format. This protein sequence containing file were run in M-Coffee from T-Coffee site [149] for multiple sequence alignment. The results were downloaded and the ALN file with FASTA sequence obtained from M-Coffee result was run in the MEGA X software [150]. Two types of alignments were done in this software: MUSCLE Alignment and ClustalW alignment. Both alignment results were analyzed to find the conserved region among all 44 protein sequences. Found conserved regions were stored in a table.

### **2.3 Conserved sequence**

From the MUSCLE and ClustalW alignment result, a lot of conserved regions were found including 1 to 16 residue long regions. Among them conserved regions  $\geq 9$  amino acids residues were selected from the alignment. 5 conserved regions with  $\geq 9$  amino acids residues were found and they were searched for the overlaps with B cell and T cell epitope prediction.



## 2.4 MHC I Epitope Prediction

MHCI stands for Major Histocompatibility I, which, is a part of adaptive immune system, and plays a role by alerting the Immune system in case of viral infection. IEDB stands for The Immune Epitope Database and Analysis Resource, a reliable tool to analyse MHC I and MHC II's reaction against given epitopes for the purpose of vaccine design and epitope-based studies.

Again, while conducting the MHC I prediction, among the tools in the server, we had selected, NetMHCpan-4.1 [151], [152], which uses Artificial Neural Networks (ANN), in order to predict the binding peptides based on MHC molecules of known sequences. The other parameters that were used were kept in default settings, again while choosing the allele sets, we selected the “frequently occurring alleles” as our aim was to target at least 1% of the human population or allele frequency of 1% or higher.

From PDB, we had selected and isolated conserved sequences for LBV, MOKV (both LBV & MOKV reviewed sequence) and one unreviewed sequence of SBV, furthermore, we used that data for NetMHCpan-4.1 for analysing our results. After that, we tracked the sequences that fall within conserved regions and listed them and after getting the full data on excel file, we had aimed to separate the allele data based on-10 percentile rank, where we selected epitope rank lower than 10, based on frequently used alleles. This process was repeated for all 3 sequences which included LBV, MOKV and SBV. After selecting the particular alleles, we separated them for further analysis

## 2.5 MHC II Epitope Prediction

MHC II stands for Major Histocompatibility II, normally B cells, DC (Dendritic Cells) and Macrophages they are express MHC II molecules, the role of the MHC II is to present antigens from exogenous source after processing them, afterwards, the antigens are sent to CD4(+) cells thus initiating immune response of the host body. MHC II epitope prediction was conducted using the **IEDB recommended 2.22 [153]** [151], this tool utilises, the following methods consisting, Consensus approach, combining NN-align, SMM-align, CombLib and Sturniolo, furthermore, if any corresponding predictor is available for the molecule, otherwise NetMHCIIpan is used. to conduct a prediction. In case of selecting the HLA parameter, we chose to stay with the full HLA reference set as our target. We did this so that we can cover the alleles more frequently observed than others in a given population along with the obtained data that can be maximised

for vaccine development. In case of selecting length, we chose to have the range between **11-18 peptide length**. Other parameters were kept default. sequence of LBV and MOKV and single sequence of SBV were obtained at first, then we had **observed** separately. Furthermore, in order to gain the best results from the MHC-2 allele, we **merged the data of the alleles based on a 10-percentile score** and then observed the pattern of the result.

## **2.6 B-Cell Epitope Prediction**

B cells are, the centre adaptive-humoral immune system also known for being antibody mediated immunity and they mature in bone marrow The binding-based prediction between B cells and epitopes were generated by using the **Bepipred Linear Epitope Prediction 2.0**, [154, 155]. The protein sequences of LBV, MOKV and single sequence of SBV were provided of in a separate manner, then the method process for predicting antibody epitope prediction Bepipred Linear Epitope Prediction 2.0 were used while having a threshold of 0.5. After obtaining the data result from the prediction tool, we observed and checked for all three sequences, and overlaps with conserved sequences.

## **2.7 Stored identified epitopes in FASTA format**

Fasta is a type of TEXT BASED formatting sequence for representing amino-acid or nucleotide-based sequences. Our possible and best candidates for epitopes were saved by creating a fasta based file. Here, header lines were named based on the target. The file containing MHCI, MHCII and B-cell epitopes were contained in a file and used for further tasks

## **2.8 Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes**

In this step, a total 38 predicted conserved epitopes from B cell and T cell epitope prediction were scanned for their Antigenicity, Allergenicity, Toxicity and Homology prediction with human sequences. The Antigenicity of the epitopes were checked with VaxiJen 2.0 server [156]. The server AllerTOP v.2.0 [157] was used for Allergenicity prediction, if the predicted epitopes were allergen or not. Toxicity prediction was done

by ToxinPred server [158] and blastp [159] was done for all the epitope sequences to check if they had any match with any of the human proteins.

Moreover, for constructing a vaccine, it is necessary to check the antigenicity, allergenicity and toxicity because the vaccine components should be highly antigenic and non-allergen to the host (human). Besides, it should be free from any kind of toxic reaction.

The antigenicity determining server VaxiJen 2.0 is an alignment independent tool for protective antigen prediction rather than depending on the sequence alignment approach. The main benefit of this server is, it does not lead the protein to lack obvious sequence similarity. Instead, it allows the antigen classification solely based on the physicochemical properties of proteins with 70% to 89% accuracy. This server set its cut off value to 0.4 for antigen prediction and the target organism for our prediction was used as virus.

VaxiJen 2.0 and AllerTOP v.2.0 both servers are based on auto cross covariance (ACC) transformation of protein sequences into uniform equal-length vectors. The AllerTOP v2.0 server was used to check the allergenicity of the conserved epitopes which has better prediction accuracy of 88.7%. In these tools, all the default parameters were used. After that toxicity of the epitopes were checked by ToxinPred server to check their toxicity. In this step, the homology of the epitopes to the human proteins also determined using blastp (protein-protein BLAST) of BLAST server. In this case, all of the default parameters were used and only Homo sapiens were kept for matching the human proteome with the epitope sequences. Here the default e-value cut off was set at 0.05 and the epitopes didn't show any hit below this cut off point, which indicated non homologous pathogen peptides.

**Table 2.1: List of the Antigenicity, Allergenicity, Toxicity and Human Homology Prediction of conserved epitopes.**

<b>Epitopes</b>	<b>Sequence</b>	<b>Antigenicity Vaxijen2.0(t 0.4)</b>	<b>Allergenicity AllerTop2.0</b>	<b>Toxicity ToxinPred</b>	<b>Human Homology Prediction BLASTp</b>
>89:99- <b>MHCII</b>	<b>TYTNFVGYV TT</b>	<b>0.4367</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>139:149- MHCII	SWLRTVTTT KE	0.3277	Allergen	Non-Toxin	
>89:97-MHCI	TYTNFVGYV	0.3508	Non-Allergen	Non-Toxin	
>89:98-MHCI	<b>TYTNFVGYV T</b>	<b>0.4914</b>	Allergen	Non-Toxin	
>90:98-MHCI	YTNFVGYVT	0.3756	Non-Allergen	Non-Toxin	
>90:99-MHCI	YTNFVGYVT T	0.3355	Non-Allergen	Non-Toxin	
>91:99-MHCI	TNFVGYVTT	0.1993	Allergen	Non-Toxin	
>122:130- <b>MHCI</b>	<b>SGDPRYEES</b>	<b>0.824</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>122:131- <b>MHCI</b>	<b>SGDPRYEES L</b>	<b>0.7134</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>123:131- <b>MHCI</b>	<b>GDPRYEESL</b>	<b>0.7311</b>	Allergen	Non-Toxin	
>123:132- <b>MHCI</b>	<b>GDPRYEESL H</b>	<b>0.6887</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>124:132- <b>MHCI</b>	<b>DPRYEESLH</b>	<b>0.8239</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>124:133- <b>MHCI</b>	<b>DPRYEESLH T</b>	<b>0.5366</b>	Allergen	Non-Toxin	
>126:134- MHCI	RYEESLHTP	-0.3341	Non-Allergen	Non-Toxin	
>126:135-	RYEESLHTPY	-0.0478	Allergen	Non-Toxin	

Epitopes	Sequence	Antigenicity Vaxijen2.0(t 0.4)	Allergenicity AllerTop2.0	Toxicity ToxinPred	Human Homology Prediction BLASTp
MHCI					
>127:135- MHCI	YEESLHTPY	-0.2117	Non-Allergen	Non-Toxin	
>127:136- MHCI	YEESLHTPYP	-0.1804	Allergen	Non-Toxin	
>128:136- MHCI	EESLHTPYP	-0.165	Allergen	Non-Toxin	
>139:147- MHCI	SWLRTVTTT	0.029	Non-Allergen	Non-Toxin	
>139:148- MHCI	SWLRTVTTT K	0.1819	Allergen	Non-Toxin	
>140:148- MHCI	WLRTVTTTK	0.2344	Allergen	Non-Toxin	
>140:149- MHCI	WLRTVTTTK E	0.3896	Allergen	Non-Toxin	
<b>&gt;153:161- MHCI</b>	<b>IISPSIVEM</b>	<b>0.6156</b>	Allergen	Non-Toxin	
>122:136- B_cell_1	SGDPRYEESL HTPYP	0.3189	Allergen	Non-Toxin	
>122:136- B_cell_2	SGDPRYEESL HTPY	0.3338	Non-Allergen	Non-Toxin	
>122:136- B_cell_3	GDPRYEESL HTPYP	0.2997	Non-Allergen	Non-Toxin	
>122:136- B_cell_4	GDPRYEESL HTPY	0.3148	Non-Allergen	Non-Toxin	
>122:136- B_cell_5	DPRYEESLHT PYP	0.3737	Non-Allergen	Non-Toxin	
>122:136- B_cell_6	SGDPRYEESL HTP	0.2011	Non-Allergen	Non-Toxin	

Epitopes	Sequence	Antigenicity Vaxijen2.0(t 0.4)	Allergenicity AllerTop2.0	Toxicity ToxinPred	Human Homology Prediction BLASTp
>122:136- B_cell_7	<b>SGDPRYEES</b> <b>LHT</b>	<b>0.4526</b>	Allergen	Non-Toxin	
>122:136- B_cell_8	PRYEESLHTP YP	-0.0198	Non-Allergen	Non-Toxin	
>122:136- B_cell_9	GDPRYEESL HTP	0.1643	Non-Allergen	Non-Toxin	
>122:136- B_cell_10	DPRYEESLHT PY	0.3962	Non-Allergen	Non-Toxin	
>122:136- B_cell_11	<b>SGDPRYEES</b> <b>LH</b>	<b>0.6802</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>122:136- B_cell_12	RYEESLHTPY P	-0.0324	Allergen	Toxin	
>122:136- B_cell_13	<b>GDPRYEESL</b> <b>HT</b>	<b>0.4372</b>	<b>Non-Allergen</b>	<b>Non-Toxin</b>	<b>No Match</b>
>122:136- B_cell_14	PRYEESLHTP Y	-0.0363	Non-Allergen	Non-Toxin	
>122:136- B_cell_15	DPRYEESLHT P	0.2309	Non-Allergen	Non-Toxin	

## 2.9 3D structures for all these epitopes were generated using PEP-FOLD3

**PEP-FOLD3** [160, 161], utilizes the de-novo based approach in order to predict the peptide structures from given amino-acid sequences. In the PEP-FOLD3 tools we had provided the best predicted epitopes we had selected. After obtaining the results the structural information was stored for further procedures. In order to obtain a clear view, we had used another tool call **Web3dMol** [162, 163].

## 2.10 3D Structure Prediction of the Epitopes

For conducting the MHC 3D structures prediction we had used the HADDOCK tool where we observed the Human MHCI and Human MHCII docking affinity. HADDOCK [164, 165] stands for High Ambiguity Driven protein-protein DOCKing. Based on observation from different papers, we have selected 2 targeted proteins which were, UniProt ID:1QEW((Ref: [UniProt](#))) represented MHC I molecule and UniProt ID:2G9H(Ref: [UniProt](#)) represented MHC II molecule. Before conducting the prediction tool run, all other parameters were kept at default for obtaining the best result possible. After observing the results, top 3 clusters were chosen which had the most potential for further analysis. Each of the 3 clusters that were obtained had 4 PDB structures. PRODIGY[166],[167],[168] stands for “PROtein binDIng enerGY prediction”, a compilation of tools which aims to predict the binding affinity for protein based biological molecules. PRODIGY was used to calculate the binding energy for protein molecules. For this step PDB identification number for our desired molecule was provided into the PRODIGY server while keeping all other parameters as default for the best possible result. Results obtained from the tool were observed. Furthermore, all results were recorded. Thus, 2 separate data were generated based on MHC I-epitope and MHC II-epitope complexes.

## 2.11 Binding prediction of the MHC peptides to HLA alleles and Population coverage for the filtered MHC epitopes

NetMHC - 4.0 [169] and NetMHCII 2.3 [170] server was used for predicting the binding affinity of the MHC epitopes to the HLA alleles. The Major Histocompatibility Complex or MHC is involved in the antigen presentation to T cells. The HLA or Human Leukocyte Antigen classified into HLA class I antigens (HLA-A, HLA-B, and HLA-C) and HLA class II antigens (HLA-DR, HLA-DQ, HLA-DP) [171]. Furthermore, MHC I includes the HLA class I molecules and MHC II includes the HLA class II molecules. For predicting the weak and strong binders to the HLA molecules against MHC I, NetMHC - 4.0 server was used and in the case of MHC II, NetMHCII 2.3 was used. These prediction tools use artificial neuron networks. NetMHC - 4.0 server can predict the binders from 81 different Human MHC alleles including HLA-A, -B, -C and -E as well as from 41 animal (Monkey, Cattle, Pig, and Mouse) alleles. And the NetMHCII 2.3 server predicts binding of peptides to HLA-DR, HLA-DQ, HLA-DP and

mouse MHC class II alleles. On the both server, the epitope sequences were submitted in a FASTA format. For NetMHC-4.0 the threshold for Strong binder was 1%, Weak Binder was 10% and for NetMHCII-2.3 Strong binder was 2%, Weak Binder was 10%. The result defines the strong binding peptides and weak binding peptides for the MHC epitopes.

### **2.11.1 Population coverage analysis**

From the best selected binders to the HLA alleles of MHC epitopes, population coverage analysis was done across the African region and the whole world. The tool [172] was used to predict the portion of individuals to respond to a given epitope set on the basis of HLA genotypic frequencies and on the basis of MHC binding data.

In order to determine the population coverage of the selected epitopes, the predicted epitope sequences with the corresponding HLA alleles (Class I and Class II) were submitted to the population coverage analysis tool of the Immune Epitope Database (IEDB) by setting the parameters to World, East Africa, Central Africa, North Africa, West Africa, and South Africa. The frequency of expression of different HLA types varies in different ethnicities as the MHC molecule is highly polymorphic. Extreme polymorphism restricts the proportion of the human population that may respond to a particular antigen. Thus a peptide which functions as T-cell epitope in a population with certain HLA make up may not be effective in another population with a different HLA allelic distribution. The aim of this analysis was to select promiscuous T-cell epitopes that bind to several alleles of HLA super types for maximal population coverage [173].

### **2.12 Vaccine construction**

After conducting the population coverage, we aimed to construct our vaccine. In our case we had developed 3 best vaccine candidates. Here we have used 3 types of adjuvants for 3 separate vaccines. After creating the adjuvants, we had selected the best epitopes extracted from MHCI, MHCII and B-cell epitopes which were built up by adding linker molecules such as GGGGS, GPGPG, and KK. Before merging the epitopes and adjuvants, a header linker known as EAAAK was inducted into the sequence, we had also generated a PADRE sequence[174] and we inducted the PADRE sequence at the beginning and at the end of our vaccine followed by a GGGGS linker,



thus completing vaccine construction sequence. Final vaccine construction has been stored.

### **2.13 Antigenicity, Allergenicity and Physicochemical properties analysis of each vaccines**

Antigenicity and allergenicity check for constructed vaccines is an important step for developing a vaccine. Antigenicity of the three constructed vaccines (Vaccine1\_L7-L12, Vaccine2\_HBHA and Vaccine3\_beta\_defensin\_3) was checked using VaxiJen 2.0 server to predict the antigenicity of the vaccines. The default cut off value for antigenicity of VaxiJen 2.0 was set at 0.4 and the target organism for our prediction was used as virus. The predicted result “Antigen” is considered as a good result because to generate a better immune response a vaccine should be highly antigenic. The allergenicity prediction tool AllerTOP v.2.0 was used for the constructed vaccines to check if it is safe from any allergic reaction. Here, plain protein sequence of vaccines was submitted to get the result. Moreover, ProtParam server [175] was used to evaluate the physicochemical properties of the constructed vaccines. In this tool only amino acid sequence was submitted to get the result. The server evaluate the Number of amino acids, Molecular weight, Theoretical pI (isoelectric point), Total number of negatively charged residues (Asp + Glu), Total number of positively charged residues (Arg + Lys), Formula, Total number of atoms, Extinction coefficients (M-1 cm-1, at 280 nm), Estimated half-life, Instability index (II), Aliphatic index and Grand average of hydropathicity (GRAVY) of vaccines.

### **2.14 2D Structure Prediction of the Constructed Vaccines**

After constructing our desired vaccine, our aim was to obtain its secondary structural based information, in order to do this, we used **PSIPRED**[176],[177] and **JPred4**[178] bio info tool. JPred4 uses JNet algorithm, and JPred4 also makes predictions of solvent accessibility and coiled-coil regions. Furthermore, JPred4 features higher accuracy, with a blind three-state ( $\alpha$ -helix,  $\beta$ -strand and coil) secondary structure prediction accuracy of 82.0% while solvent accessibility prediction accuracy had been raised to 90% for residues <5% accessible. Three separate vaccine data were provided in both PSIPRED and JPred4 website and their results were stored. Results obtained from the tool were compared to get a proper and accurate visualisation of 2D structure of the

vaccine.

Finally in order to get a clear picture and comparison between JPRED4 and PSIPRED's data we had used another tool named **2dss**[179], [180] for visual comparison of the secondary structures. After obtaining the results we analysed and observed the result and saved the data for further analysis.

## 2.15 3D Structure Prediction of the Constructed Vaccines

After creating 2d structures for our vaccines, we aimed to generate a 3d structure. The 3d structures were generated using the tool called **I-TASSER**[181],[182] I-TASSER stands for **Iterative Threading ASSEmbly Refinement**, which is a type of hierarchical based process to structure-based function annotation and protein structure prediction. Our vaccine data for 3 vaccines were provided to the tool and after running the tool we had obtained our results which were predicted by the I-TASSER, the best model for 3 vaccines were chosen and retrieved. We used RCSB-PDB to observe our vaccines. Since we have already obtained our 3d structure from I-TASSER, now we had aimed to refine and validate the 3d structure by using **Galaxy-Refine**[183],[184] by utilising the next *generation in silico protein design project*, this tool helps to **repack** the **side chain** of the protein and by simulation process, it helps to modify the protein structure. Files that had been stored from I-TASSER result; the same data were provided for **Galaxy-Refine** tool. After providing the data we had obtained **5 results** for **each of the 3 vaccines**, thus a total of **15 models** were obtained. The structural results, which were retrieved from the Galaxy-Refine tool were checked again using **PDBsum**[185, 186]. PDBsum was used both to generate our vaccine proteins 3d model and also the tool provided **Ramachandran plot**. We studied and observed the value of the Ramachandran plot which provided us with necessary numerical data to choose the best protein from the given sequences. Furthermore, we used **SAVES6.0** tool[187], [188],[189],[190],[191] a meta server-based tool where the protein's structural stereochemical quality is being checked by a residue to residue. We studied and observed the value obtained from the Ramachandran plot and data obtained from SAVES too, which provided us with necessary evidential data to choose the best protein from the given sequences. After observing and comparing all the data combined from the PDBsum's Ramachandran plot and SAVES data-based value, we had chosen our best models.

For **Vaccine 1**, we had chosen **model: 4**. Again, For **Vaccine 2**, we had chosen **model 2**. Lastly For **Vaccine 3**, we had chosen **model 3**, with the most stable and reliable result.

## **2.16 Disulfide engineering of vaccines**

Disulfide engineering of the constructed vaccines were carried out by Disulfide by Design server [192]. This server is used to design new disulfide bonds in the protein. Here, the 3D structure of protein files in PDB format were analyzed to predict the residue pairs that are likely to form a disulfide bond if the respective amino acids are mutated to cysteines. The output displays residue pairs having the appropriate geometric characteristics for disulfide formation and provides automated generation of modified PDB files including modeled disulfides [193]. It is important to design disulfide bonds in protein because introducing disulfide bonds into proteins has been used efficiently to improve protein stability, functional characteristics modification and analyzing protein dynamics. Furthermore, this web based server Disulfide by Design has excess ability of functionality, visualization and analyzing capabilities including analysis of the B-factor of protein regions involved in predicted disulfide bonds [194]. In this step, possible disulfide bond positions were analyzed using default parameters in this server. Only the bonds that had a binding energy less than 2.2 kcal/mol were chosen for mutation into C-C bonds because 90% of native disulfide bonds are generally found to have energy value less than 2.2 Kcal/ mol [195]. For vaccine 2, one such bond was left out as it was resulting in discontinuous protein. By disulfide engineering, Vaccine 1 added 5 bonds, Vaccine 2 added 2 bonds, and Vaccine 3 added 1 bond. After that, the mutated PDBs were downloaded for each vaccine, and then processed into standard PDB files using UCSF chimera [196]. After disulfide modification, the mutated vaccines were considered as the final vaccine candidates.

In the end of this step, the disulfide modified vaccines were used for further steps: a) The 3D structure of the disulfide modified vaccines visualization using RCSB PDB - Mol\* 3D Viewer [197], b) The FASTA format sequences generated using Chimera and c) antigenicity, allergenicity, and physicochemical properties prediction using VaxiJen v2.0, AllerTOP v.2.0 and ProtParam server respectively.

## **2.17 Molecular Docking of the Constructed Vaccines against MHC I and MHC II**

Docking was conducted using the **HADDOCK**. Protein molecules selected for 3d structures were 1QEW and 2G9H. In case of 1QEW, we had selected Chain A - HLA-A 0201 - MHCI, here From Chain A, Amino acids 1-182 were selected as target, as they correspond to the alpha-1 and alpha-2 domains of the peptide-binding cleft (Ref: [UniProt](#)) Again, in case of 2G9H, we chose Chain B - HLA-DRA1 (MHCII), here From Chain B, Amino acids 1-95 were selected as targets, as they correspond to the beta-1 domain of the peptide-binding cleft. (Ref: [UniProt](#)). Simulation was run aimed to dock against Vaccine proteins with MHC I epitopes, similarly with MHC II epitopes as well. While conducting this we kept all the parameters default to generate best probable results. Clusters we obtained from simulated docking; we chose the best 3 clusters. The 3 clusters also had 4 separate pdb structures thus a total of 12 results were observed. The binding affinity was calculated using the PRODIGY tool. We observed and selected our best results based on comparing the values of the binding energy later merging them, thus obtaining 6 results with low binding affinity. After observing the binding energy results, they were recorded in an excel file via **xls** format. Finally, files were saved for further analysis.

## **2.18 B-lymphocytic Epitope prediction**

The target of constructing a vaccine is to stimulate the memory of the adaptive immune system so that it responds immediately to the next antigen exposure [198]. And a class of all the adaptive immune systems, the humoral immune system is mediated by antibodies which are produced by B lymphocyte cells. Any kind of immune system including humoral immune systems are stimulated by recognition of antigen or epitope to the receptor [198]. As a result a vaccine should have the most effective conformational B cell epitopes for providing better immunity. ElliPro [199] is the best tool for predicting the B cell epitopes. Each PDB file of the vaccines were run in this tool by keeping the default parameter (Minimum score: 0.5, Maximum distance: 6 Angstrom). This tool uses estimation of protein structure as an ellipsoid, calculation of the residual protrusion index (PI), and grouping of residues based on PI values for prediction. And this tool is based on the combination of the Thornton concept, the

MODELER program, and Jmol viewer. The 3D structure of continuous and discontinuous epitopes can be viewed by the Jmol viewer [198].

Epitopes recognized by B cells can be classified into two types: continuous and discontinuous epitopes. Continuous epitopes (also referred to as linear or sequential epitopes) are short peptide fragments (about 15 amino acids in size) of an antigen protein that are specifically identified by certain antibodies. Discontinuous epitopes consist of amino acid residues that are not sequential in their primary structure but involve a folding mechanism that forms into a region that is close together. However, the folding mechanism increases the complexity of epitope prediction; the classification is not rigid because several continuous epitopes could form certain conformations that are recognized by antibodies and discontinuous epitopes can also contain several sequential linear peptide sequences. Because of their complexity, the prediction of B-cell epitopes is often less accurate than the simpler prediction of T-cell epitopes.

## **2.19 Immune Simulation Analysis**

In order to conduct our vaccine's immune based simulation for understanding the effectiveness and immunogenicity of our vaccine, we used the tool called,

**C-IMMSIM [200, 201]**. The C-IMMSIM tool's server uses Position-Specific Scoring Matrix (PSSM) in order to generate a simulation based real life immune interaction. While running the tool, we kept the parameters default in order to generate the best possible result for simulation. Here the number of steps to conduct the simulation were kept at 2000 steps. Furthermore, for the class I HLA-A epitope selection, we chose A0101 and A2402, For the class I HLA-B epitope selection, we chose B3501 and B0702. Lastly for class II, HLA-DRB epitope molecules we chose DRB1\_0701 and DRB1\_1501. For our vaccines to work with the best possible effect we chose 3 doses of injection. Number of antigens in all 3 injections were kept 1000. We also had chosen 3 timesteps where 1st time step was on day 1, for 1st injection. 2nd time step was on day 7, for 2nd injection. 3rd time step was on day 14, for 3rd injection. Then we generated graph for immune response simulation for all three vaccines with a 3-dose regimen (1, 7, and 14 days). Results obtained from the graph were analysed, observed and kept in a file.

## Chapter 3. Results

### 3.1 Result of Strain Identification and Protein Sequence Retrieval

The full-length sequence of *Lagos bat virus* (LBV), *Mokola virus* (MOKV) and *Shimoni bat virus* (SHBV) of phylogroup II were 522 amino acids long. On UniprotKB database 20 Glycoprotein (G) sequence of *Lagos bat virus*, 23 G protein sequence of *Mokola virus* and 1 unreviewed G protein sequence of *Shimoni bat virus* were retrieved from the UniprotKB database (**Table 3.1**). The protein sequences were retrieved in a FASTA format.

**Table 3.1: List of the Glycoprotein sequences with their length and their accession number**

Serial Number	Name of the virus	Accession number	Status	Protein name	Length (Amino acid)
1	LBV	Q91C26	unreviewed	Glycoprotein	522
2	LBV	B3SN97	unreviewed	Glycoprotein	522
3	LBV	B2XXV7	unreviewed	Glycoprotein	522
4	LBV	Q8BDV6	<b>reviewed</b>	Glycoprotein	522
5	LBV	B3SN95	unreviewed	Glycoprotein	522
6	LBV	B3SN93	unreviewed	Glycoprotein	522
7	LBV	A0A565DIL6	unreviewed	Glycoprotein	522
8	LBV	B3SNA1	unreviewed	Glycoprotein	522
9	LBV	K9K092	unreviewed	Glycoprotein	522

<b>Serial Number</b>	<b>Name of the virus</b>	<b>Accession number</b>	<b>Status</b>	<b>Protein name</b>	<b>Length (Amino acid)</b>
10	LBV	B2C4C7	unreviewed	Glycoprotein	522
11	LBV	B3SNA0	unreviewed	Glycoprotein	522
12	LBV	B3SNA4	unreviewed	Glycoprotein	522
13	LBV	B3SNA3	unreviewed	Glycoprotein	522
14	LBV	B3SN94	unreviewed	Glycoprotein	522
15	LBV	B3SN98	unreviewed	Glycoprotein	522
16	LBV	D4NRK6	unreviewed	Glycoprotein	522
17	LBV	K7TAY6	unreviewed	Glycoprotein	522
18	LBV	E5FR83	unreviewed	Glycoprotein	522
19	LBV	B3SN96	unreviewed	Glycoprotein	522
20	LBV	Q91C27	unreviewed	Glycoprotein	522
21	MOKV	Q89507	unreviewed	Glycoprotein	522
22	MOKV	P0C572	<b>reviewed</b>	Glycoprotein	522
23	MOKV	S4S290	unreviewed	Glycoprotein	522
24	MOKV	B2XXZ2	unreviewed	Glycoprotein	522

<b>Serial Number</b>	<b>Name of the virus</b>	<b>Accession number</b>	<b>Status</b>	<b>Protein name</b>	<b>Length (Amino acid)</b>
25	MOKV	S5DTC1	unreviewed	Glycoprotein	522
26	MOKV	S5DTB7	unreviewed	Glycoprotein	522
27	MOKV	B2XXZ7	unreviewed	Glycoprotein	522
28	MOKV	S5DTB2	unreviewed	Glycoprotein	522
29	MOKV	S5DTA7	unreviewed	Glycoprotein	522
30	MOKV	S4S295	unreviewed	Glycoprotein	522
31	MOKV	S4S2B6	unreviewed	Glycoprotein	522
32	MOKV	A0A1L2C207	unreviewed	Glycoprotein	522
33	MOKV	R9Q7P2	unreviewed	Glycoprotein	522
34	MOKV	Q83543	unreviewed	Glycoprotein	522
35	MOKV	E5FR84	unreviewed	Glycoprotein	522
26	MOKV	R9Q9C4	unreviewed	Glycoprotein	522
37	MOKV	A0A1L2C213	unreviewed	Glycoprotein	522
38	MOKV	F2VIE2	unreviewed	Glycoprotein	522
39	MOKV	S4S266	unreviewed	Glycoprotein	522



Serial Number	Name of the virus	Accession number	Status	Protein name	Length (Amino acid)
40	MOKV	U5L2V9	unreviewed	Glycoprotein	522
41	MOKV	S4S268	unreviewed	Glycoprotein	522
42	MOKV	R9Q7B7	unreviewed	Glycoprotein	522
43	MOKV	R9Q7B6	unreviewed	Glycoprotein	522
44	SHBV	D4NRK1	<b>unreviewed</b>	Glycoprotein	522

### 3.2 Protein sequence alignment and conserved region finding

From multiple sequence alignment of M-Coffee results, several conserved sequences have been found (**Table 3.2**) and analyzed. Conserved sequence analysis revealed that only 5 regions were conserved among all the 44 strains which are equal and longer than 9 amino acid residues. These 5 conserved regions (**Table 3.2**) were searched for the B cell and T cell epitope prediction.

**Table 3.2: Phylogroup II conserved region from MUSCLE and M-Coffee Alignment Total 44 sequence (522aa) was aligned (including reviewed and unreviewed): The selected 5 conserved regions were highlighted with BOLD format.**

Serial	Conserved region sequence	Residues	AA Sequence
1	1	1	M
2	17	1	S
3	21-28	8	FPLYTIPE
4	33-37	5	WTPID

<b>Serial</b>	<b>Conserved region sequence</b>	<b>Residues</b>	<b>AA Sequence</b>
5	40-41	2	HL
6	43-46	4	CPNN
7	49	1	S
8	53-54	2	GC
9	62	1	Y
10	65	1	L
11	68	1	G
12	70	1	L
13	72-75	4	HQKV
14	77-83	7	GFTCTGV
15	86-87	2	EA
<b>16</b>	<b>89-99</b>	<b>11</b>	<b>TYTNFVG YVTT</b>
17	101-106	6	FKRKHF
18	108-109	2	PT
19	112-116	5	ACRDA
20	119-120	2	WK
<b>21</b>	<b>122-137</b>	<b>16</b>	<b>SGDPRYEESLHTPYPD</b>
<b>22</b>	<b>139-149</b>	<b>11</b>	<b>SWLRTVTTTKE</b>
23	151	1	L
<b>24</b>	<b>153-162</b>	<b>10</b>	<b>IISPSIVEMD</b>
25	164	1	Y
26	166	1	R
27	168-174	7	LHSPMFP

<b>Serial</b>	<b>Conserved region sequence</b>	<b>Residues</b>	<b>AA Sequence</b>
28	176	1	G
29	182	1	Y
30	184	1	S
31	186-188	3	PSC
32	190-195	6	TNHDYT
33	197-199	3	WLP
34	201	1	D
35	208-211	4	CDIF
36	216	1	G
37	218	1	K
38	220-223	4	MNGS
39	226-228	3	CGF
40	230-233	4	DERG
41	235-236	2	YR
42	240	1	G
43	242-248	7	CKLTLCG
44	250-251	2	PG
45	253-254	2	RL
46	256-259	4	DGTW
47	261	1	S
48	265	1	P
49	271	1	C
50	273	1	P

<b>Serial</b>	<b>Conserved region sequence</b>	<b>Residues</b>	<b>AA Sequence</b>
51	275-276	2	QL
52	278	1	N
53	280-281	2	HN
54	283	1	R
55	285-286	2	DE
56	288-292	5	EHLIV
57	294	1	D
58	299-300	2	RE
59	302-308	7	CLDTLET
60	311-312	2	MS
61	314	1	S
62	316-319	4	SFRR
<b>63</b>	<b>321-329</b>	<b>9</b>	<b>SHFRKLVPG</b>
64	331-335	5	GKAYT
65	338-341	4	NGSL
66	345-346	2	NV
67	348	1	Y
68	351-352	2	VD
69	354	1	W
70	357-360	4	ILPS
71	362-365	4	GCLK
72	370	1	C
73	376-377	2	GV

<b>Serial</b>	<b>Conserved region sequence</b>	<b>Residues</b>	<b>AA Sequence</b>
74	379-381	3	FNG
75	383	1	I
76	387-388	2	DG
77	391-397	7	LIPEMQS
78	401-403	3	KQH
79	405-406	2	DL
80	408-409	2	KA
81	412-414	3	FPL
82	416-419	4	HPLI
83	425	1	F
84	428-429	2	DG
85	431-432	2	AD
86	434-435	2	FV
87	438-441	4	HMPD
88	444	1	K
89	447	1	S
90	450	1	D
91	452	1	G
92	456	1	W
93	463	1	G
94	472	1	L
95	474-475	2	CL
96	479-480	2	CC

Serial	Conserved region sequence	Residues	AA Sequence
97	509	1	V
98	511-512	2	SS
99	514-515	2	ES
100	517	1	K
101	522	1	V

### 3.3: Predicted MHCI Epitopes

Using NetMHCpan-4.1 tool, predictions were generated for MHC I binding epitopes. Separated prediction results for, 3 strains (MOKV, SBV and LBV) of rabies virus from Phylogroup II along with frequently occurring alleles (**Table 3.3**), (**Table 3.4**), (**Table 3.5**), were generated for, with the help of Artificial Neural Networks (ANN) method of NetMHCpan-4.1 tool.

**Table 3.3: Virus MHCI Table based on frequently occurring alleles**

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-B*57:01	1	24	33	10	YTIPEKIGPW	0.985811	0.02
HLA-B*57:01	1	189	197	9	ATNHDYTLW	0.984491	0.02
HLA-B*58:01	1	189	197	9	ATNHDYTLW	0.981122	0.01
HLA-B*57:01	1	111	119	9	LACRDYHW	0.980632	0.03
HLA-B*58:01	1	111	119	9	LACRDYHW	0.979647	0.01
HLA-A*11:01	1	58	66	9	SVFSYVELK	0.973828	0.01

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-B*58:01	1	24	33	10	YTIPEKIGPW	0.960138	0.03
HLA-B*57:01	1	315	323	9	ISFRRLSHF	0.956117	0.06
HLA-B*15:01	1	253	262	10	RLFDGTWISF	0.950643	0.01
HLA-B*15:01	1	340	348	9	SLMETNVHY	0.948703	0.01
HLA-A*32:01	1	253	262	10	RLFDGTWISF	0.939308	0.01

**Table 3.4: Mokola Bat Virus MHC I Table based on frequently occurring alleles**

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-B*57:01	1	262	270	9	FTKPDVHVW	0.997186	0.01
HLA-B*57:01	1	24	33	10	YTIPEKIEKW	0.99525	0.01
HLA-B*58:01	1	262	270	9	FTKPDVHVW	0.994926	0.01
HLA-B*57:01	1	261	270	10	SFTKPDVHVW	0.993333	0.01
HLA-B*58:01	1	24	33	10	YTIPEKIEKW	0.993152	0.01
HLA-A*26:01	1	88	96	9	ETYTNFVGY	0.987696	0.01
HLA-B*35:01	1	173	182	10	FPSGVCSNVY	0.980475	0.01
HLA-B*57:01	1	111	119	9	AACRDAYNW	0.973196	0.04

<b>Allele</b>	<b>Seq_num</b>	<b>Start</b>	<b>End</b>	<b>Length</b>	<b>Peptide</b>	<b>Score</b>	<b>Percentile Rank</b>
HLA-A*02:01	1	283	291	9	RLDEIEHLI	0.96611	0.02
HLA-B*15:01	1	253	262	10	RLFDGTWVSF	0.962106	0.01
HLA-B*58:01	1	111	119	9	AACRDAYNW	0.960732	0.03

**Table 3.5: Shimoni Bat Virus MHC I Table based on frequently occurring alleles**

<b>Allele</b>	<b>Seq_num</b>	<b>Start</b>	<b>End</b>	<b>Length</b>	<b>Peptide</b>	<b>Score</b>	<b>Percentile Rank</b>
HLA-B*57:01	1	262	270	9	ITRPEIVMW	0.998588	0.01
HLA-B*58:01	1	262	270	9	ITRPEIVMW	0.99198	0.01
HLA-B*57:01	1	261	270	10	SITRPEIVMW	0.990753	0.01
HLA-B*57:01	1	24	33	10	YTIPEKIGPW	0.985811	0.02
HLA-B*57:01	1	189	197	9	LTNHDYTIW	0.982352	0.03
HLA-A*68:01	1	58	66	9	STFSYIELR	0.981597	0.01
HLA-B*58:01	1	111	119	9	SACRDAYHW	0.981432	0.01
HLA-B*57:01	1	111	119	9	SACRDAYHW	0.98009	0.03
HLA-B*58:01	1	189	197	9	LTNHDYTIW	0.974645	0.01
HLA-B*57:01	1	110	119	10	ASACRDAYHW	0.96163	0.05



Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-B*58:01	1	24	33	10	YTIPEKIGPW	0.960138	0.03

Furthermore, another selection was done on the first database that we had obtained, this selection was based on conserved regions and on-10 percentile-based rank on the allele that were provided for maximum efficiency generating a vaccine

**Table 3.6: *Lagos Bat Virus* MHC I Table based on top-10 percentile rank**

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-A*24:02	1	89	97	9	TYTNFVG YV	0.112993	0.79
HLA-A*23:01	1	89	97	9	TYTNFVG YV	0.056121	1.2
HLA-A*68:02	1	89	97	9	TYTNFVG YV	0.024693	3.1
HLA-A*33:01	1	89	97	9	TYTNFVG YV	0.006556	6
HLA-B*51:01	1	89	97	9	TYTNFVG YV	0.004698	9
HLA-A*26:01	1	89	97	9	TYTNFVG YV	0.001469	9
HLA-A*68:02	1	89	98	10	TYTNFVG YVT	0.003531	8.5
HLA-A*24:02	1	89	98	10	TYTNFVG YVT	0.000709	9.3
HLA-A*23:01	1	89	98	10	TYTNFVG YVT	0.000637	9.9

**Most common** peptides we had obtained for *Lagos Bat* virus (Table 3.6) were **TYTNFVG YV**, **TYTNFVG YVT**, **YTNFVG YVT**, **YTNFVG YVTT**, **SGDPRYEESL**, **GDPRIEESL**, **DPRYEESLH**, **DPRYEESLHT**, **RYEESLHTP**,

**RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTVTTT, SWLRTVTTTK, WLRTVTTTK, WLRTVTTTKE, IISPSIVEM** Uncommon peptides (**found only once**) we had obtained for *Lagos Bat* virus are **TNFVGYVTT, SGDPRYEES, GDPRYEESLH**

**Table 3.7: Mokola Bat Virus MHC I Table based on top-10 percentile rank**

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-A*24:02	1	89	97	9	TYTNFVGYYV	0.112993	0.79
HLA-A*23:01	1	89	97	9	TYTNFVGYYV	0.056121	1.2
HLA-A*68:02	1	89	97	9	TYTNFVGYYV	0.024693	3.1
HLA-A*33:01	1	89	97	9	TYTNFVGYYV	0.006556	6
HLA-B*51:01	1	89	97	9	TYTNFVGYYV	0.004698	9
HLA-A*26:01	1	89	97	9	TYTNFVGYYV	0.001469	9
HLA-A*68:02	1	89	98	10	TYTNFVGYYVT	0.003531	8.5
HLA-A*24:02	1	89	98	10	TYTNFVGYYVT	0.000709	9.3
HLA-A*23:01	1	89	98	10	TYTNFVGYYVT	0.000637	9.9

**Most common** peptides we had obtained for Mokola Virus (**Table 3.7**) are **TYTNFVGYYV, TYTNFVGYYVT, YTNFVGYYVT, YTNFVGYYVT, SGDPRYEESL, GDPRYEESL, DPRYEESLH, DPRYEESLHT, RYEESLHTP, RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTVTTT, SWLRTVTTTK, WLRTVTTTK, IISPSIVEM**. **Uncommon (found only once)** peptides we had obtained for, Mokola Virus are **TNFVGYVTT, SGDPRYEES, GDPRYEESLH, WLRTVTTTKE**

**Table 3.8: Shimoni Bat Virus MHC I Table based on top-10 percentile rank**

Allele	Seq_num	Start	End	Length	Peptide	Score	Percentile Rank
HLA-A*24:02	1	89	97	9	TYTNFVGYV	0.112993	0.79
HLA-A*23:01	1	89	97	9	TYTNFVGYV	0.056121	1.2
HLA-A*68:02	1	89	97	9	TYTNFVGYV	0.024693	3.1
HLA-A*33:01	1	89	97	9	TYTNFVGYV	0.006556	6
HLA-B*51:01	1	89	97	9	TYTNFVGYV	0.004698	9
HLA-A*26:01	1	89	97	9	TYTNFVGYV	0.001469	9
HLA-A*68:02	1	89	98	10	TYTNFVGYVT	0.003531	8.5
HLA-A*24:02	1	89	98	10	TYTNFVGYVT	0.000709	9.3
HLA-A*23:01	1	89	98	10	TYTNFVGYVT	0.000637	9.9

**Most common** peptides (Table 3.8) we had obtained for Shimoni Virus are **TYTNFVGYV, TYTNFVGYVT, YTNFVGYVT, YTNFVGYVTT, SGDPRYEESL, GDPRYEESL, DPRYEESLH, DPRYEESLHT, RYEESLHTP, RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTVTTT, SWLRTVTTTK, WLRTVTTTK, IISPSIVEM. Uncommon (found only once)** peptides we had obtained for Shimoni Virus are **TNFVGYVTT, SGDPRYEES, GDPRYEESLH, WLRTVTTTKE**

### 3.4 Predicted MHCII Epitopes

A variety of recommended tools were used for generating the best possible prediction for MHC II binding prediction. In case of HLA while using default setting, of 12-18 amino acid sequence, we were having complications with our results, to avoid this shortcoming we readjusted our amino acid from default 12-18 amino acid sequence to 11-18 amino acid sequence, thus making a broad range HLA set which provided us with a satisfactory prediction. Furthermore, rest of the parameters were kept in default settings. Separated prediction results for, 3 strains (MOKV, SBV and LBV) of rabies virus from phylogroup-II were generated. Our **main objective** was to retrieve and analyse the **best percentile rank** which **needed** to be **less than score of 10**, from 3 sets of MHC-2 binding epitope database, we observed that the **results were exactly same for each of 3 sets**, thus we combined and merge 3 database into **a single database** for better visualization (**Table 3.9**), which would be used for generating best vaccine candidates.

**Table 3.9: LBV, MKV & SBV combined MHC II Table based on 10 percentile rank**

Allele	Seq_num	Start	End	Length	Peptide	Percentile_Rank	Adjusted_Rank
HLA-DRB1*15:01	1	89	99	11	TYTNFVGYVTT	8.8	50.28
HLA-DRB1*07:01	1	89	99	11	TYTNFVGYVTT	9.8	55.99
HLA-DRB1*04:01	1	139	149	11	SWLRTVTTTKE	3	17.14
HLA-DRB1*04:05	1	139	149	11	SWLRTVTTTKE	4.7	26.85
HLA-DRB1*08:02	1	139	149	11	SWLRTVTTTKE	7	39.99

In the combined database that we had generated it consisted of **5 similar results** (**Table 3.9**). The allele sequence was (**HLA-DRB1\*15:01, HLA-DRB1\*07:01, HLA-DRB1\*04:01, HLA-DRB1\*04:05, HLA-DRB1\*08:02**). Furthermore, we had found **2**

**common** types of **peptides** in the 5 allele groups, which were **TYTNFVGYVTT**, size -11 (start seq-89, End seq-99) and **SWLRTVTTTKE**, size -11 (start seq-139, End seq-149)

### 3.5 Predicted B-Cell Epitopes

We observed the B-Cell epitopes that we had primarily obtained (Table 3.10).

**Table 3.10: B-cell epitope list**

No	Start	End	Peptide	Length
1	28	38	EKIEKWTPIDM	11
2	41	75	LSCPNNLLSEEEGCNAESSFTYFELKSGYLAHQKV	35
3	98	115	TTTFKRKHFRPTVAACRD	18
4	117	136	YNWKVSGDPRYEESLHTPYP	20
5	144	145	VT	2
6	159	173	VEMDIYGRTLHSPMF	15
7	175	175	S	1
8	178	191	CSNVYPSVPSCETN	14
9	201	208	DPSLSLVC	8
10	213	223	SSNGKKAMNGS	11
11	232	239	RGFYRSLK	8
12	268	268	H	1
13	272	272	T	1
14	276	291	LINIHNDRLDEIEHLI	16
15	297	297	K	1
16	299	303	REECL	5
17	310	318	LMSQSVSFR	9
18	320	331	LSHFRKLVPGYG	12
19	353	377	KWADILPSKGCLKVGQCMEPVKGV	25
20	391	457	LIPEMQSEQLKQHMDLLKAAVFPLRHPLISREAVFK KDGDADD FVDLHMPDVHKSVDVDLGLPHWG	67
21	483	518	VRRRRSGRATQEIPLSFPSAPVPRAKVSSWESYKG	36

After obtaining the data result from the prediction tool, we checked for all three sequences, and overlaps with conserved sequences. Based on the overall data provided by the prediction tool, we concluded that

**Table 3.11: Best Selected B-cell epitope**

<b>4</b>	<b>117</b>	<b>136</b>	<b>YNWKVSGDPRYEESLHTPYP</b>	<b>20</b>
----------	------------	------------	-----------------------------	-----------

Amino acid **sequence 121-137**, falls under the **conserved region and sequence 117-136** amino acid, had been identified as Epitope and B cell epitope falls under the area of 122-136 amino acid sequence (**Table 3.11**). Furthermore, a combination of total 15 epitope amino acids were generated, by analyzing the 122-136 range sequence, where the epitopes were sized between 11 to 15 aa.

### 3.6 Stored Epitope Sequences

**Table 3.12: List of identified epitopes in fasta**

<b>SL</b>	<b>Range</b>	<b>Epitopes</b>	<b>Peptide sequence</b>
1	89:99	MHCII	TYTNFVGYVTT
2	139:149	MHCII	SWLRTVTTTKE
3	89:97	MHCI	TYTNFVGYV
4	89:98	MHCI	TYTNFVGYVT
5	90:98	MHCI	YTNFVGYVT
6	90:99	MHCI	YTNFVGYVTT
7	91:99	MHCI	TNFVGYVTT
8	122:130	MHCI	SGDPRYEES
9	122:131	MHCI	SGDPRYEESL
10	123:131	MHCI	GDPREYEEESL
11	123:132	MHCI	GDPREYEEESLH
12	124:132	MHCI	DPREYEEESLH

<b>SL</b>	<b>Range</b>	<b>Epitopes</b>	<b>Peptide sequence</b>
13	124:133	MHCI	DPRYEESLHT
14	126:134	MHCI	RYEESLHTP
15	126:135	MHCI	RYEESLHTPY
16	127:135	MHCI	YEESLHTPY
17	127:136	MHCI	YEESLHTPYP
18	128:136	MHCI	EESLHTPYP
19	139:147	MHCI	SWLRTVTTT
20	139:148	MHCI	SWLRTVTTTK
21	140:148	MHCI	WLRTVTTTK
22	140:149	MHCI	WLRTVTTTKE
23	153:161	MHCI	IISPSIVEM
24	122:136	B_cell_1	SGDPRYEESLHTPYP
25	122:136	B_cell_2	SGDPRYEESLHTPY
26	122:136	B_cell_3	GDPRYEESLHTPYP
27	122:136	B_cell_4	GDPRYEESLHTPY
28	122:136	B_cell_5	DPRYEESLHTPYP
29	122:136	B_cell_6	SGDPRYEESLHTP
30	122:136	B_cell_7	SGDPRYEESLHT
31	122:136	B_cell_8	PRYEESLHTPYP
32	122:136	B_cell_9	GDPRYEESLHTP

SL	Range	Epitopes	Peptide sequence
33	122:136	B_cell_10	DPRYEESLHTPY
34	122:136	B_cell_11	SGDPRYEESLH
35	122:136	B_cell_12	RYEESLHTPYP
36	122:136	B_cell_13	GDPRYEESLHT
37	122:136	B_cell_14	PRYEESLHTPY
38	122:136	B_cell_15	DPRYEESLHTP

Our results provided us with **2** MHC-II epitopes, **21** MHC-I epitopes and **15** B cell epitopes (**Table 3.12**).

### **3.7 Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes**

From the Antigenicity, Allergenicity, Toxicity prediction of the 38 conserved epitopes, 12 epitopes showed antigenicity, 22 epitopes were non-allergen, only 1 epitope were non-toxic epitopes and the remaining 37 epitopes were toxic. And comparing the homology prediction result with human proteome, 7 epitopes (**Table 3.13**) were considered as the best epitopes for vaccine construction. Among the 7 epitopes, 4 were MHC I predicted epitopes, 1 were MHC II predicted and 2 were B-cell predicted epitopes. These 7 epitopes fulfilled all the required criteria for further steps. Moreover, these filtered epitopes were stored in a FASTA file.

**Table 3.13: Result of the Antigenicity, Allergenicity, Toxicity, and Human Homology prediction of conserved epitopes analysis**

Serial	Epitopes	Sequence
1	>89:99-MHCII	TYTNFVGYVTT



Serial	Epitopes	Sequence
2	>122:130-MHCI	SGDPRYEES
3	>122:131-MHCI	SGDPRYEESL
4	>123:132-MHCI	GDPRYEESLH
5	>124:132-MHCI	DPRYEESLH
6	>122:136-B_cell_11	SGDPRYEESLH
7	>122:136-B_cell_13	GDPRYEESLHT

### 3.8 Predicted 3D Structures of Epitopes

Data obtained using PEP-FOLD 3 result, we observed the data using Web3dMol, we found that the data sets can be divided into three groups, **Group 1(MHC-II epitopes, consisting of 1 set of result), Group 2(MHC-I epitopes, consisting of 4 sets of result) & Group 3(B-cell epitopes, consisting of 2 sets of result).**

#### Group 1

For **single MHC-II epitope (position: 89-99), 1 data set were obtained** from PEP-FOLD 3

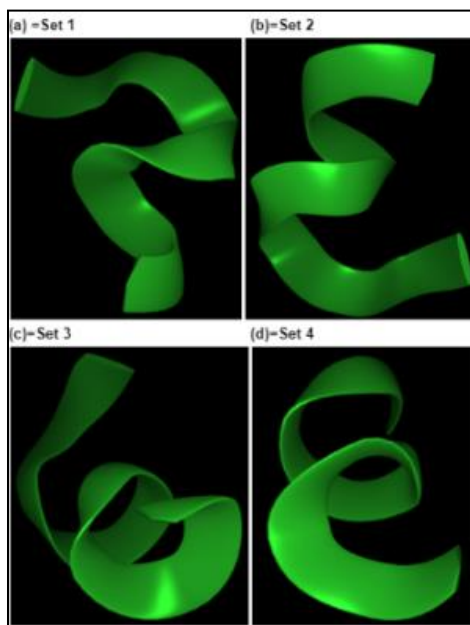


**Figure 3.1: MHC-II, group 1 dataset**

and Web3dMol. Common **amino** acid residues that we had observed (**Figure 3.1**) are **THR, TYR, ASN, PHE, VAL, GLY**

**Group 2**

For **MHC-I** related epitopes, there were, **4 datasets** we had obtained.

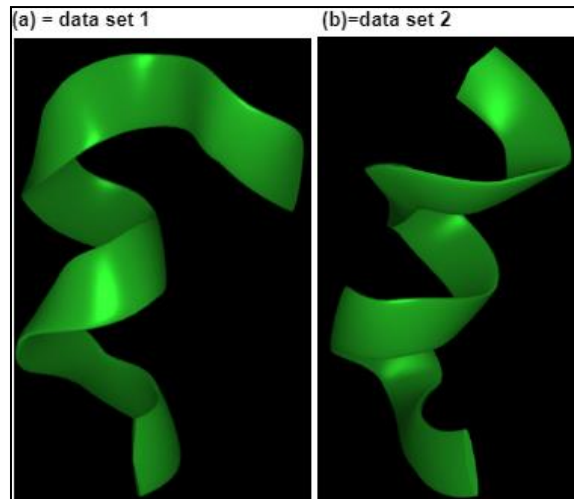


**Figure 3.2: MHC-I, 4 data sets, (a) set 1: Position 122\_130, (b) set 2: Position 122\_131, (c) set 3: Position 123\_132, (d) set 4: Position 124\_132**

Common **amino** acid residues that we had observed (**Figure 3.2**) from PEP-FOLD 3 tool for **4 epitopes** for **MHC-I** are **SER, GLY, ASP, PRO, ARG, TYR, GLU**

**Group 3**

**B-cell epitopes** related dataset, **2 data sets** were obtained



**Figure 3.3: B cell epitope data sets, (a) data set 1\_B-cell\_11: Position 122\_136, (b) data set 2\_B-cell\_13: Position 122\_136**

**Common amino acid residues** that we had observed from PEP-FOLD 3 tool for 2 epitopes (Figure 3.3) for B-cell are **SER, GLY, ASP, PRO, ARG, TYR, GLU, LEU, HIS**

### 3.9 Molecular Docking Analysis of the MHC Epitopes

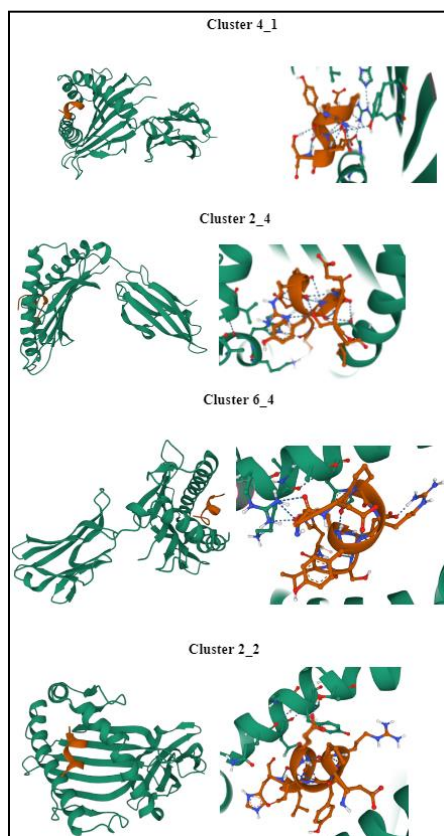
Observing and analyzing among the best candidates we had chosen 2 targeted proteins, One protein was

1QEY: Chain A HLA-A 0201 - Refers to a functional molecule for MHC1. When it binds with the B2M/beta 2 microglobulin, it presents primarily viral and tumor-obtained based peptides on antigen-presenting cells for recognition by alpha-beta T cell receptor (TCR) on HLA-A-restricted CD8-positive T cells. Furthermore, HLA-A 02:01 which was representative of MHC-1, can be found as a major allele in the human population. Again, from Chain A variant, Amino acid sequence 1-182 were selected as target, the reason was as they correspond to the alpha-1 and alpha-2 domains of the peptide-binding cleft (Ref: [UniProt](#)).

Another protein for MHC-II, that we had chosen was 2G9H: Chain B - HLA-DRA1 (MHCII): Refers to the antigen-presenting major histocompatibility complex class II (MHCII) molecule's beta chain. While binding with the HLA-DRA's alpha chain, it shows the antigenic peptides on professional antigen presenting cells (APCs) for recognition by alpha-beta T cell receptor (TCR) on HLA-DRB1-restricted CD4-positive T cells. From Chain B, Amino acids 1-95 were selected as target, as they correspond to the beta-1 domain of the peptide-binding cleft. (Ref: [UniProt](#)).

### MHCI epitopes docking against 1QEW\_A

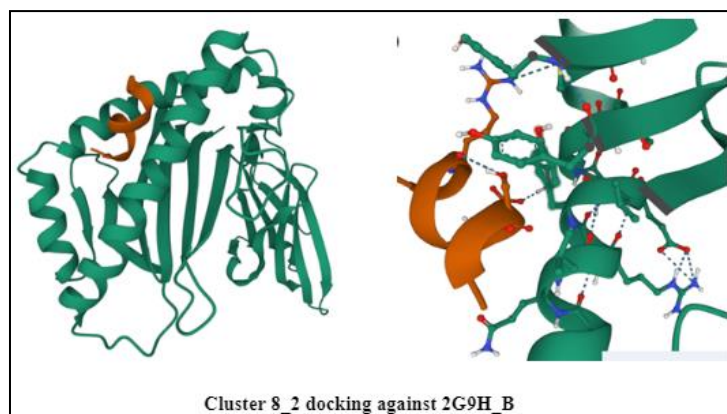
We had obtained **4 cluster** results (**Figure 3.4**) for docking with MHC-1, they were Cluster 4\_1, Cluster 2\_4, Cluster 6\_4, Cluster 2\_2



**Figure 3.4: MHC 1 (4) epitopes docking against 1QEW**

### MHC II epitope docking against 2G9H\_B

We had obtained **1 cluster** result (**Figure 3.5**) for docking with MHC-2, which was Cluster 8\_2.



**Figure 3.5: MHC II epitope docking against 2G9H**

Among these **4 clusters** (**3 models** from MHC-1 and **1 model** from MHC-2) after comparative analysis, the best **3** were chosen (**2 models** from MHC-1 and **1 model** from

MHC-2). Again, each of these clusters had 4 PDB based structures According to the protocol, we observed our best result, which was based on the “lowest energy that was needed for binding affinity”. Furthermore, the data suggests that our clusters had undergone stable binding with targeted proteins. An excel file was created and we kept store data.

**Table 3.14: Binding Energy calculation for MHC I and MHC II docking human 1QEW and 2G9H**

Epitope	Position	Type	Cluster No	Model No	Binding Energy
MHC-I	122_130	1QEW	4	1	-10.7
MHC-I	122_131	1QEW	2	4	-11.1
MHC-I	123_132	1QEW	6	4	-11.2
MHC-I	124_132	1QEW	2	2	-10.2
MHC-II	89_99	2G9H	8	2	-10.4

Based on the data, we had obtained, **4** clusters of **MHCI** (Table 3.14), for the **position** of **122\_130**, the **binding energy** was **-10.7**, for the **position** of **122\_131**, the **binding energy** was **-11.1**, for the **position** of **123\_132**, the **binding energy** was **-11.2**, lastly, for the **position** of **124\_132**, the **binding energy** was **-10.2**. Again, for the **1** cluster of **MHC II** (Table 3.14), which had the **position** of **89\_99**, it had the **binding energy** of **-10.4**.

### **3.10 Binding prediction of the MHC peptides to HLA alleles and Population coverage for the filtered MHC epitopes**

The NetMHC-4.0 and NetMHCII-2.3 server output indicated the weak binding prediction and strong binding prediction. A Comprehensive set of HLA alleles were identified for all MHC epitopes, which were stored in a file (Table 3.15).

**Table 3.15: Result of the predicted HLA allele sets for the MHC epitopes**

<b>Epitope Name</b>	<b>Epitope Sequence</b>	<b>Main Prediction</b>	<b>Extended Prediction Strong</b>	<b>Extended Prediction Weak</b>	<b>Submission</b>
MHCII: 89-99	TYTNF VGYVT T	HLA-DRB1*15:01, HLA-DRB1*07:01		HLA-DRB1*15:01, HLA-DRB1*16:02, HLA-DQA1*01:04/DQB1*05:03, HLA-DQA1*02:01/DQB1*03:03, HLA-DQA1*02:01/DQB1*04:02, HLA-DQA1*03:01/DQB1*03:01	HLA-DRB1*15:01, HLA-DRB1*07:01, HLA-DRB1*16:02, HLA-DQA1*01:04/DQB1*05:03, HLA-DQA1*02:01/DQB1*03:03, HLA-DQA1*02:01/DQB1*04:02, HLA-DQA1*03:01/DQB1*03:01
MHCI:1 22-130	SGDPR YEES	HLA-A*01:01		HLA-B*08:02, HLA-C*05:01, HLA-C*08:02	HLA-A*01:01, HLA-B*08:02, HLA-C*05:01, HLA-C*08:02
MHCI:1 22-131	SGDPR YEESL	HLA-B*08:01, HLA-B*07:02, HLA-A*01:01, HLA-A*24:02, HLA-B*40:01	HLA-B*35:03, HLA-B*42:01	HLA-B*08:01, HLA-B*07:02, HLA-A*02:17, HLA-A*24:03, HLA-B*07:02, HLA-B*08:01, HLA-B*08:02, HLA-B*14:02, HLA-B*27:20, HLA-B*38:01, HLA-B*40:02, HLA-B*40:13, HLA-B*48:01,	HLA-B*08:01, HLA-B*07:02, HLA-A*01:01, HLA-A*24:02, HLA-B*40:01, HLA-B*35:03, HLA-B*42:01, HLA-A*02:17, HLA-A*24:03, HLA-B*07:02, HLA-B*08:01, HLA-B*08:02, HLA-B*14:02,

<b>Epitope Name</b>	<b>Epitope Sequence</b>	<b>Main Prediction</b>	<b>Extended Prediction Strong</b>	<b>Extended Prediction Weak</b>	<b>Submission</b>
				B*83:01, HLA- C*05:01, HLA- C*08:02, HLA- C*14:02, HLA- E*01:01	HLA-B*27:20, HLA-B*38:01, HLA-B*40:02, HLA-B*40:13, HLA-B*48:01, HLA-B*83:01, HLA-C*05:01, HLA-C*08:02, HLA-C*14:02, HLA-E*01:01
MHCI:1 23-132	GDPRY EESLH	HLA- B*07:02	HLA- B*35:03, HLA- B*42:01	HLA-B*07:02, HLA- A*24:03, HLA- B*08:01, HLA- B*14:02, HLA- B*27:20, HLA- B*35:01, HLA- B*38:01, HLA- B*40:02, HLA- B*40:13, HLA- B*48:01, HLA- B*83:01, HLA- C*06:02, HLA- C*14:02, HLA- E*01:01	HLA-B*07:02, HLA-B*35:03, HLA-B*42:01, HLA-A*24:03, HLA-B*08:01, HLA-B*14:02, HLA-B*27:20, HLA-B*35:01, HLA-B*38:01, HLA-B*40:02, HLA-B*40:13, HLA-B*48:01, HLA-B*83:01, HLA-C*06:02, HLA-C*14:02, HLA-E*01:01
MHCI:1 24-132	DPRYE ESLH	HLA- B*35:01, HLA- B*53:01,	HLA- B*35:03, HLA- B*42:01	HLA-B*35:01, HLA- B*07:02, HLA- B*08:01, HLA- B*83:01, HLA-	HLA-B*35:01, HLA-B*53:01, HLA-B*07:02, HLA-B*35:03,

<b>Epitope Name</b>	<b>Epitope Sequence</b>	<b>Main Prediction</b>	<b>Extended Prediction Strong</b>	<b>Extended Prediction Weak</b>	<b>Submission</b>
		HLA-B*07:02		C*06:02	HLA-B*42:01, HLA-B*07:02, HLA-B*08:01, HLA-B*83:01, HLA-C*06:02

### 3.10.1 Population coverage analysis of the epitopes and their MHC alleles

The population coverage analysis tool of the Immune Epitope Database (IEDB) computes the projected population coverage, average number of epitope hits/ HLA combinations recognised by the populations and minimum number of epitope hits/ HLA combinations recognised by 90% of the population (PC90) [173]. The analysis showed that (**Table 3.16**) in the world population, MHC class I, MHC class II and the MHC class I and class II combination covered about 94.43%, 36.36% and 96.45%, respectively. The highest coverage has been shown in the world population (**Table 3.22, Figure 3.11**). And among the African region (Central Africa, East Africa, North Africa, South Africa and West Africa), The highest pervasiveness of MHC class I epitopes was found in North Africa which is 74.83% (**Table 3.19, Figure 3.8**) and the lowest was found in East Africa which covered 68.68% (**Table 3.18, Figure 3.7**). Besides, North Africa and East Africa have shown the highest MHC class II coverage (35.34%) (**Table 3.19, Figure 3.8**) and lowest MHC class II coverage (11.08%) (**Table 3.18, Figure 3.7**) respectively.

Moreover, the highest combining coverage prediction of MHC class I and class II were found in North Africa (83.73%) (**Table 3.19, Figure 3.8**) and the lowest combining coverage was found in East Africa (72.15%) (**Table 3.18, Figure 3.7**). The overall coverage result of analysis and individual region's population coverage of the epitopes with their MHC alleles are visualized in the following table (**Table 3.16**) and charts.



**Table 3.16: Population Coverage Analysis result for African region**

population / area	Class I			Class II			Class combined		
	coverage a	average hit b	pc90 c	coverage a	average hit b	pc90 c	coverage a	average hit b	pc90 c
Central Africa	70.32 %	2.09	0.34	16.48%	0.17	0.12	75.22%	2.26	0.4
<b>East Africa</b>	<b>68.68 %</b>	2.07	0.32	<b>11.08 %</b>	0.11	0.11	<b>72.15 %</b>	2.18	0.36
<b>North Africa</b>	<b>74.83 %</b>	2.38	0.4	<b>35.34 %</b>	0.37	0.15	<b>83.73 %</b>	2.75	0.61
South Africa	73.79 %	2.55	0.38	0.00%	0	0	73.79%	2.55	0.38
West Africa	69.94 %	1.98	0.33	15.85%	0.16	0.12	74.71%	2.14	0.4
<b>World</b>	<b>94.43 %</b>	4.02	2.08	<b>36.36 %</b>	0.39	0.16	<b>96.45 %</b>	4.4	2.19
<b>Average</b>	<b>75.33</b>	<b>2.52</b>	<b>0.64</b>	<b>19.18</b>	<b>0.2</b>	<b>0.11</b>	<b>79.34</b>	<b>2.71</b>	<b>0.72</b>
<b>Strandard deviation</b>	<b>8.81</b>	<b>0.7</b>	<b>0.64</b>	<b>12.96</b>	<b>0.14</b>	<b>0.05</b>	<b>8.49</b>	<b>0.78</b>	<b>0.66</b>

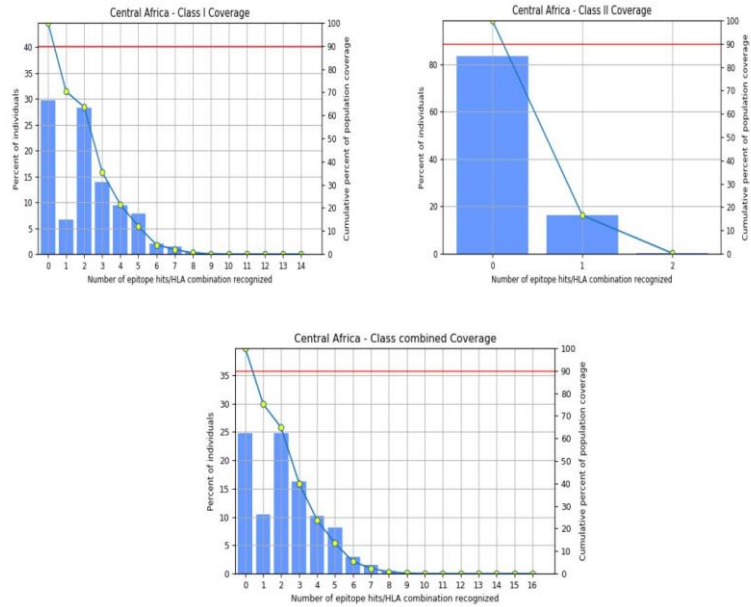
a= projected population coverage

b= average number of epitope hits / HLA combinations recognized by the population

c= minimum number of epitope hits / HLA combinations recognized by 90% of the population [172].

**Table 3.17: Central Africa**

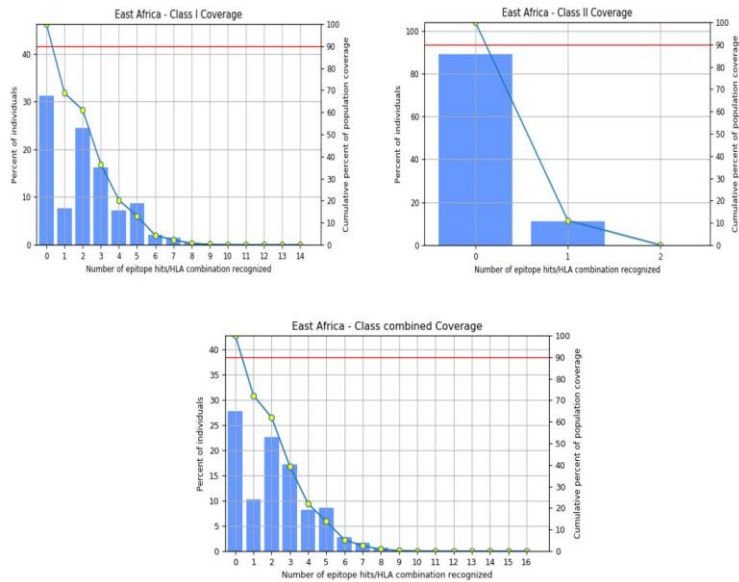
MHC class	Coverage	Average hit	PC90
I	70.32%	2.09	0.34
II	16.48%	0.17	0.12
Combined	75.22%	2.26	0.4



**Figure 3.6: Representation of MHC Class I, Class II and Class Combined coverage of Central Africa**

**Table 3.18: East Africa**

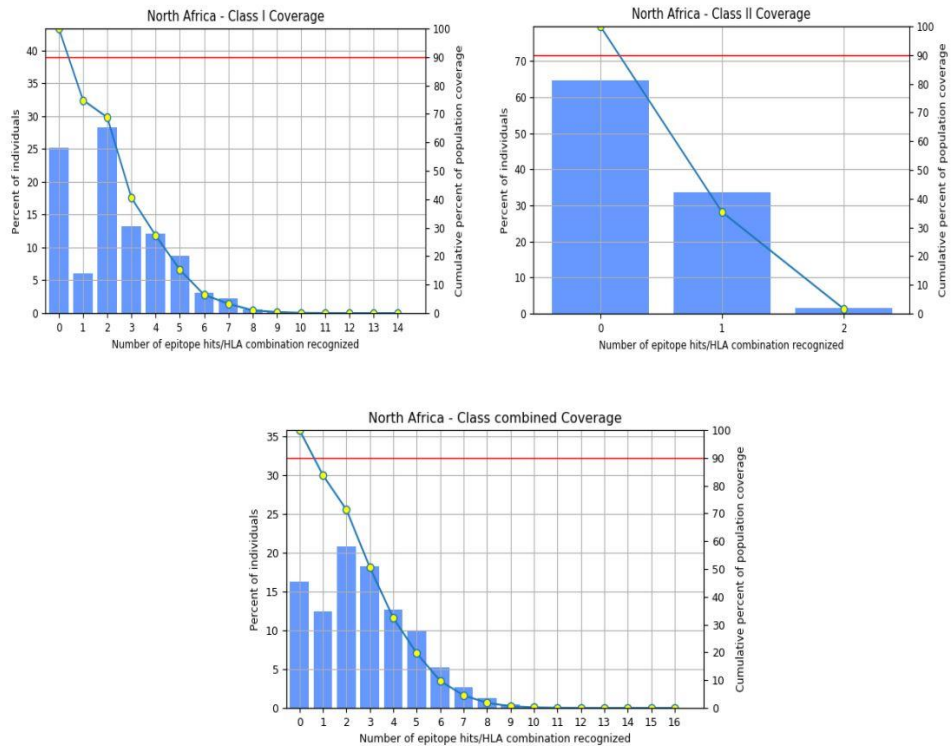
MHC class	Coverage	Average hit	PC90
I	68.68%	2.07	0.32
II	11.08%	0.11	0.11
Combined	72.15%	2.18	0.36



**Figure 3.7: Representation of MHC Class I, Class II and Class Combined coverage of East Africa**

**Table 3.19: North Africa**

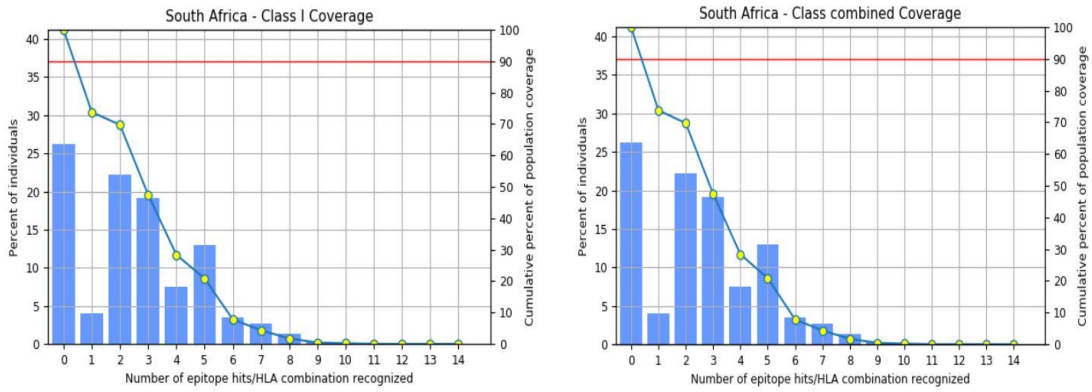
MHC class	Coverage	Average hit	PC90
I	74.83%	2.38	0.4
II	35.34%	0.37	0.15
Combined	83.73%	2.75	0.61



**Figure 3.8: Representation of MHC Class I, Class II and Class Combined coverage of North Africa.**

**Table 3.20: South Africa**

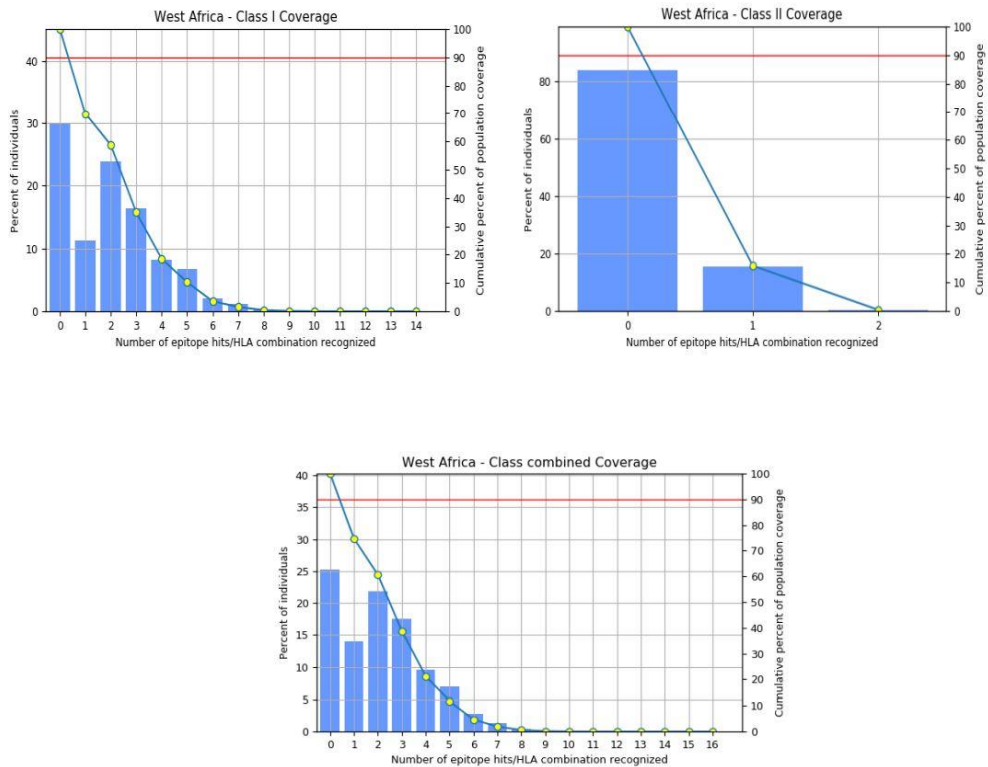
MHC class	Coverage	Average hit	PC90
I	73.79%	2.55	0.38
II	0.00%	0	0
Combined	73.79%	2.55	0.38



**Figure 3.9: Representation of MHC Class I and Class Combined coverage of South Africa**

**Table 3.21: West Africa**

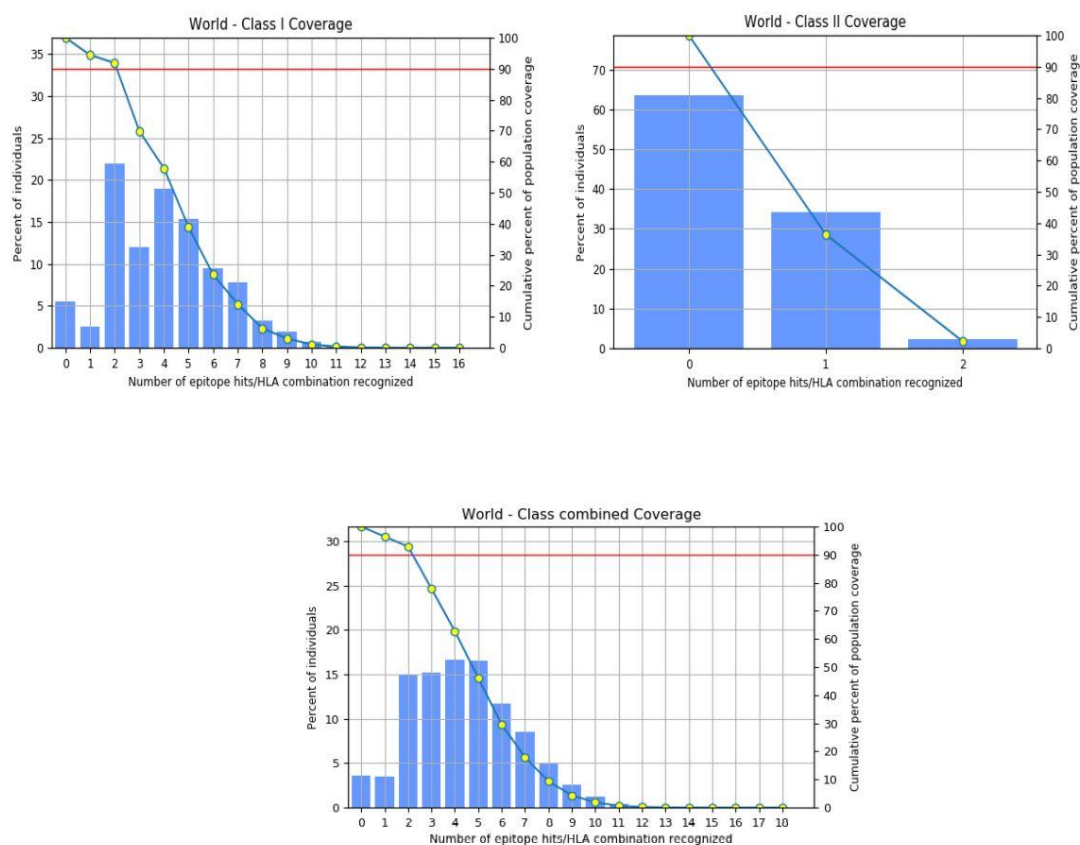
MHC class	Coverage	Average hit	PC90
I	69.94%	1.98	0.33
II	15.85%	0.16	0.12
Combined	74.71%	2.14	0.4



**Figure 3.10: Representation of MHC Class I, Class II and Class Combined coverage of West Africa**

**Table 3.22: World**

MHC class	Coverage	Average hit	PC90
I	94.43%	4.02	2.08
II	36.36%	0.39	0.16
Combined	96.45%	4.4	2.19



**Figure 3.11: Representation of MHC Class I, Class II and Class Combined coverage of World**

### 3.11 Multi-epitope Vaccine Constructs

We generated three best possible vaccine candidates for our rabies Phylogroup II viruses, termed as **Vaccine 1**, **Vaccine 2** and **Vaccine 3**.

The main components are

**a) Adjuvants (Table 3.23)** (V1: L7/L12 ribosomal protein, V2: HBHA protein, V3: beta-defensin-3).

**Table 3.23: List of Adjuvant**

Adjuvant name	Adjuvant sequence
L7/L12 ribosomal protein	MAKLSTDELLDAFKEMTLLELSDFVKKFEETFVETAAA PVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKI GVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAA DEAKAKLEAAGATVTVK
HBHA protein	MAENPNIDDLAPLLAALGAADLALATVNDLIANLRER AEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFTT EELRKA AEGYLEAATNRYNELVERGEAALQRLRSQTA FEDASARAEGYVDQAVELTQEALGTVASQTRAVGERA AKLVGIEL
beta-defensin-3	GIINTLQKYCYCRVGGRCVLSCLPKKEEQIGKCSTRGR KCCRRKK

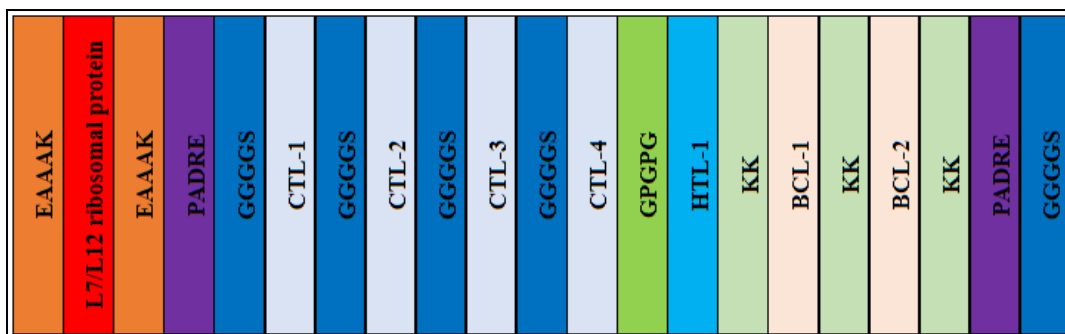
Adjuvants are known to stimulate and enhance the vaccines longevity, stability, immunogenicity and antigenicity[202],[203].

**b) MHC I, MHC II and B cell epitopes** work as specific binding sites of the antigen to work and properly function.

**c) PADRE** sequence (AKFVAAWTLKAAA) also known as the **pan HLA-DR epitope sequence**; PADRE sequence activates the **helper CD4+ T Cells [204]**.

**d) Linkers** function as junctional molecules, linkers in our vaccine were selected as **EAAAK, GGGGS, GPGPG**, and **KK** linkers. **EAAAK** acts as effective separation of the domains of bifunctional fusion protein [205]. **GGGGS** is a type of flexible linker which has proven to be efficient in conferring resistance to protease enzymes[206]. **GPGPG** aids in the immune processing and presentation furthermore, preventing the generation of the junctional epitopes [207]. **KK**, also known as bi-lysine, aims to preserve the independent immunological functions of the epitopes of a vaccine [208]

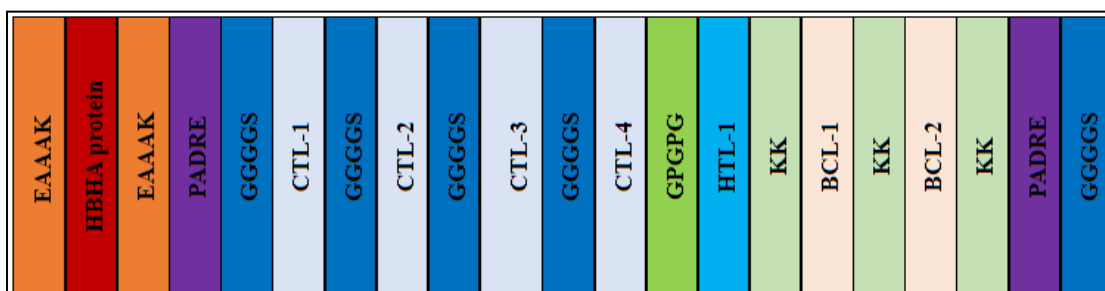
**Vaccine 1(Figure 3.12)** is constructed with the **L7/L12 ribosomal protein** (*M. tuberculosis*, accession number: *AGV15514.1*) as **main adjuvant** which were held by **2 EAAAK** linkers in



**Figure 3.12: Structural construction of Vaccine 1**

front and in back of the sequence, then **PADRE** sequences was been added and 3 **MHC-1 epitopes** were inducted, the MHC-1 epitopes were separated with the addition of the **GGGGS** linkers in between, then **MHC-2 epitopes** were inducted and they were also separated with the **GPGPG** linkers in between, again, **B cell epitopes** were also inducted and they were separate with **KK** linkers in between as well, finally, at the end of the vaccine construction another **PADRE** sequence along with **GGGGS** linker were added [202].

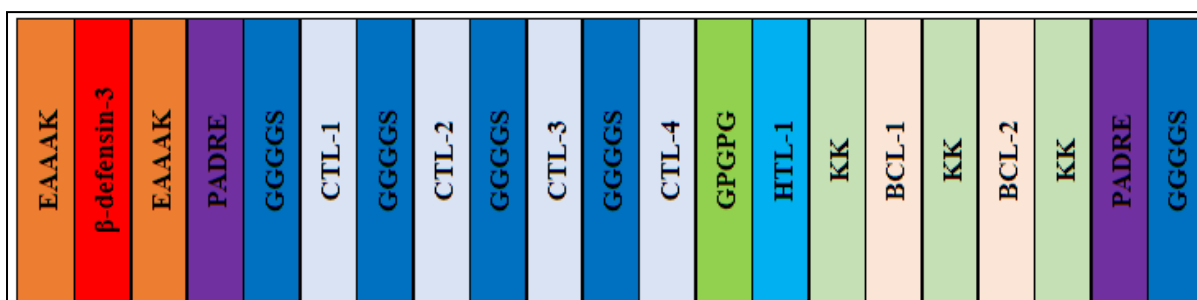
**Vaccine 2**(**Figure 3.13**) was constructed with the **HBHA protein** (*M. tuberculosis*, accession number: *AGV15514.1*), and was also constructed the same as vaccine 1,



**Figure 3.13: Structural construction of Vaccine 2**

keeping the same PADRE sequence, MHC I, MHC II epitopes and B cell epitopes, linkers.

**Vaccine 3** Again, (*UniProt accession number: Q5U7J2*), was constructed with the **beta-defensin-3**[203]



**Figure 3.14: Structural construction of Vaccine 3**

and the same components were used to construct it as well (**Figure 3.14**). Finally, after we had constructed our vaccine, we stored the necessary information for checking further aspects of the vaccines function.

### **3.12 Antigenicity, Allergenicity and Physicochemical properties analysis of each vaccine**

All of the three vaccines (Vaccine 1, vaccine 2 and vaccine 3) has been found to be as probable antigen with the scores of 0.5314, 0.5685 and 0.6151 respectively in the VaxiJen 2.0 server (**Table 3.24**). Also, they appeared as non-allergen in the AllerTOP v.2.0. Therefore, from ProtParam analysis, the Vaccine 1 and 2 have comparatively similar isoelectric point or Theoretical PI which is 5.51 and 5.46 respectively, whereas vaccine 3 have 9.5 Theoretical PI value. Besides, the extinction coefficient of Vaccine 3 was higher than Vaccine 1 and 2 and it has the lowest GRAVY. Furthermore, these three vaccines were found to be stable and they had similar estimated half-life.



**Table 3.24: Result of the Antigenicity, Allergenicity and physicochemical properties analysis of constructed vaccines**

		Name of the vaccines		
		Vaccine1_L7-L12	Vaccine2_HBHA	Vaccine3_beta_defensin_3
	<b>Antigenicity</b> [Vaxijen v2.0] Threshold 0.4 Target organism: Virus	Probable <b>Antigen</b> 0.5314	Probable <b>Antigen</b> 0.5685	Probable <b>Antigen</b> 0.6151
	<b>Allergenicity</b> [AllerTop v.2.0]	Non-Allergen	Non-Allergen	Non-Allergen
<b>Physicochemical Properties [ProtParam]</b>	<b>Number of amino acids</b>	308	337	241
	<b>Molecular weight</b>	31537.15	35725.31	25058
	<b>Theoretical pI</b>	5.51	5.46	9.5
	<b>Total number of negatively charged residues (Asp + Glu)</b>	47	53	25
	<b>Total number of positively charged residues (Arg + Lys)</b>	41	46	41

	Name of the vaccines		
	Vaccine1_L7-L12	Vaccine2_HBHA	Vaccine3_beta_defensin_3
<b>Formula</b>	C1379H2198N382 O455S4	C1537H2450N460 O516S3	C1079H1718N3 30O343S8
<b>Total number of atoms</b>	4418	4966	3478
<b>Extinction coefficients</b> M-1 cm-1, at 280 nm	Ext. coefficient 23045 Abs 0.1% (=1 g/l) 0.731, assuming all pairs of Cys residues form cystines Ext. coefficient 22920 Abs 0.1% (=1 g/l) 0.727, assuming all Cys residues are reduced	Ext. coefficient 27515 Abs 0.1% (=1 g/l) 0.770, assuming all pairs of Cys residues form cystines Ext. coefficient 27390 Abs 0.1% (=1 g/l) 0.767, assuming all Cys residues are reduced	Ext. coefficient 31900 Abs 0.1% (=1 g/l) 1.273, assuming all pairs of Cys residues form cystines Ext. coefficient 31400 Abs 0.1% (=1 g/l) 1.253, assuming all Cys residues are reduced
<b>Estimated half-life</b>	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours (Escherichia coli, in vivo)	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours (Escherichia coli, in vivo)	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours (Escherichia coli, in vivo)

Physicochemical Properties [ProtParam]

		Name of the vaccines		
		Vaccine1_L7-L12	Vaccine2_HBHA	Vaccine3_beta_defensin_3
Physicochemical Properties [ProtParam]	Instability index II	25.24 Stable	36.02 Stable	29.50 Stable
	Aliphatic index	66.07	63.32	47.55
	Grand average of hydropathicity (GRAVY)	-0.481	-0.743	-0.8

### 3.13 2D Structures of Vaccine Constructs

Analysis run by the Psipred tool revealed Helix, Coil and Strand Regions based on 2D structure for the three best vaccine candidates. PSIPRED uses two feed-forward neural networks, by performing an analysis on output obtained from position Specific Iterated BLAST(PSI-BLAST) (Figure 3.15).



Figure 3.15: PSIPRED result interpretation for 2d structure prediction



## Vaccine 2

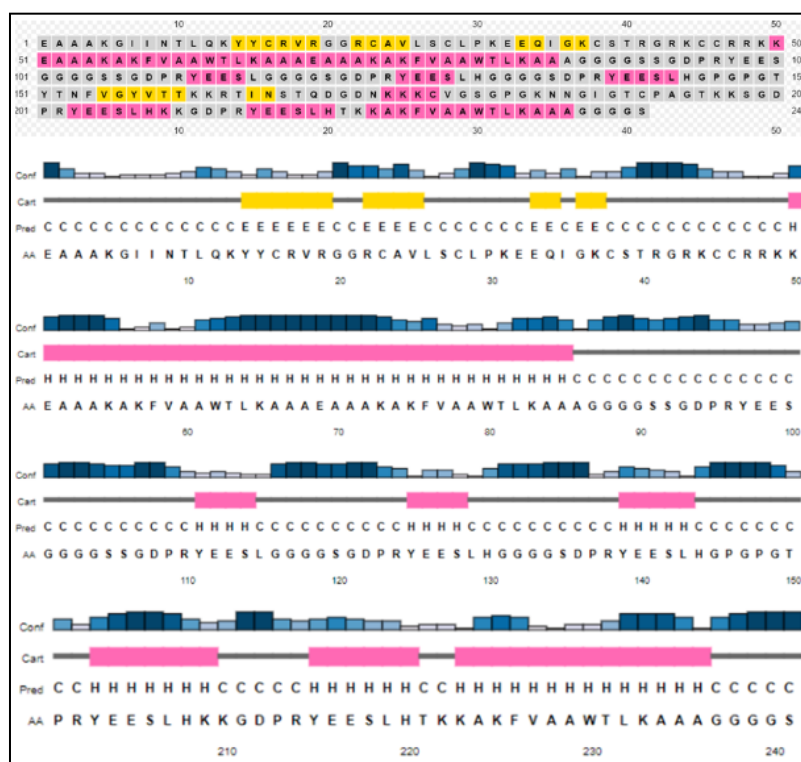
Vaccine 2 consisted of **337 amino acid** (Figure 3.17) sequences, a PSIPRED based result for Vaccine 2 was provided.



Figure 3.17: 2d structure prediction by PSIPRED for vaccine 2

### Vaccine 3

Vaccine 3 consisted of **241 amino acid (Figure 3.18)** sequences, a PSIPRED based result for Vaccine 3 was provided.



**Figure 3.18: 2d structure prediction by PSIPRED for vaccine 3**

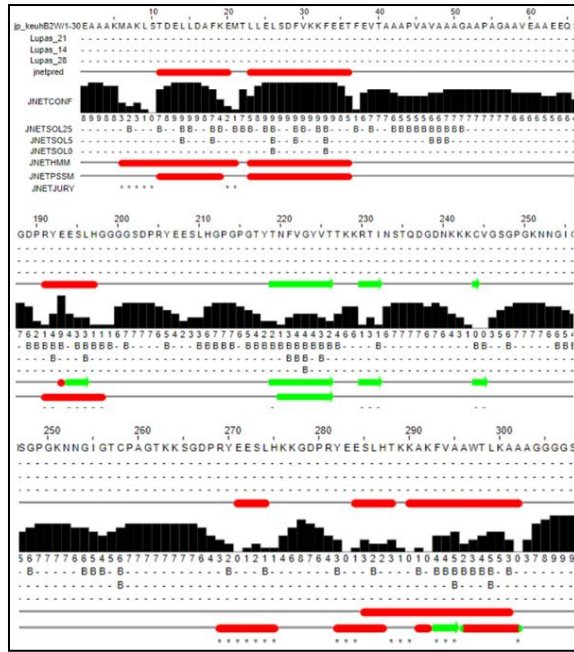
Another tool named **JPred4** was used to again determine the 2D structure of our best vaccine proteins, we used this tool to further analyze our proteins Helix, Coil and Extracellular Regions (**Table 3.25**)

**Table 3.25: Interpretation of JPred4 signs**

Signs	Interpretation
—	Coil
—	Helix
→	Strand

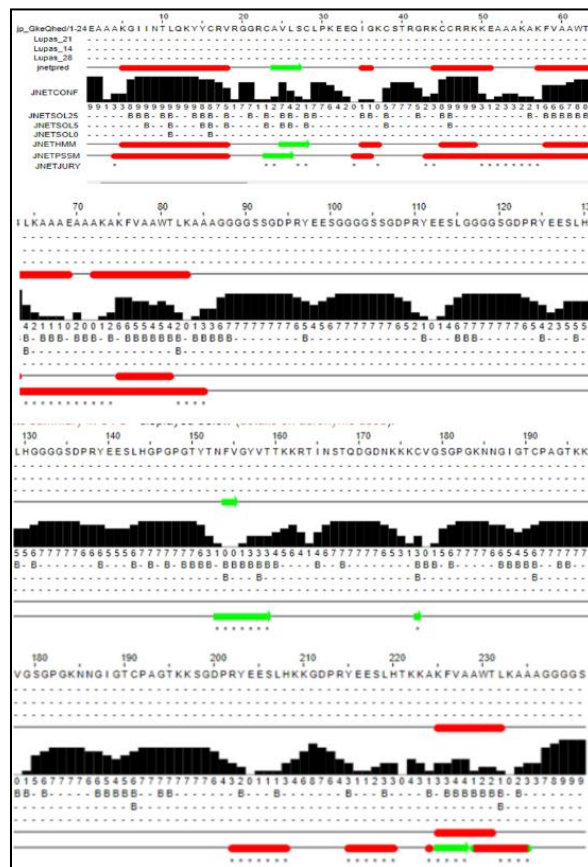
### Vaccine 1

In case of vaccine 1(**Figure 3.19**), there are several Coil, Helix (red colour) and Strand (green colour) were predicted by the JPRED4 tool



**Figure 3.19: 2d structure prediction by Jpred4 for vaccine 1**  
**Vaccine 2**

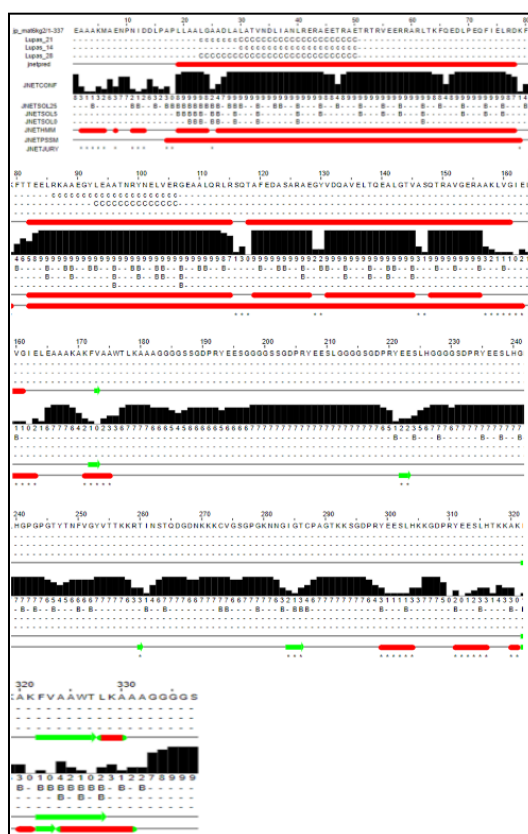
In case of vaccine 2 (Figure 3.20), several Coil, Helix (red colour) and Strand (green colour) regions were predicted by the tool



**Figure 3.20: 2d structure prediction by Jpred4 for vaccine 2**

### Vaccine 3

In case of vaccine 3(**Figure 3.21**), regions of Strand (green color), Coil and Helix (red color) were predicted by JPRED4






**Figure 3.21: 2d structure prediction by Jpred4 for vaccine 3**

Another comparison was conducted between PSIPRED and JPRED4 data using 2dss tool.

#### Comparison using 2dss tool

We obtained 3 types of figures (**Table 3.26**) from 2dSS, which are





**Table 3.26: Interpretation of Signs used in 2dss tool**

Signs	Interpretation
	Helix
	Coil
	Strand



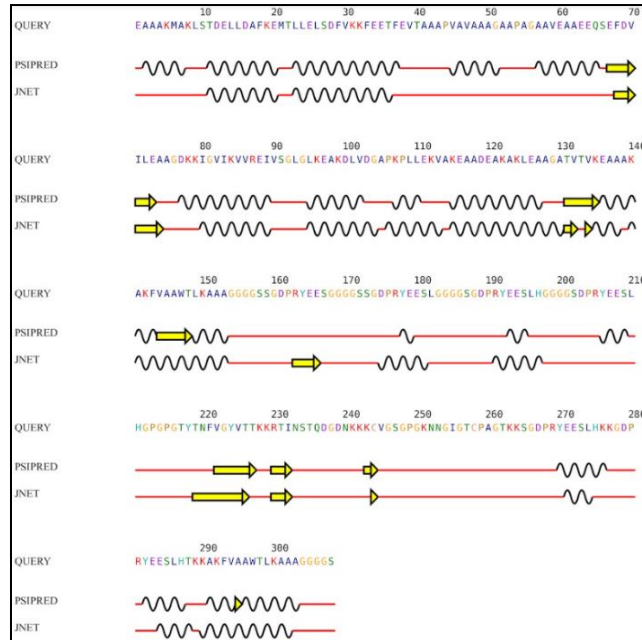
Again, based on the picture provided by 2dss (Table 3.27), the results that we had observed can be interpreted into 2 types of results such as,

**Table 3.27: Interpretation of results used in 2dss tool**

Picture	Interpretation
<p>QUERY  <span style="margin-left: 100px;">10</span>      <span style="margin-left: 100px;">20</span>  EAAAKMAENPNIDDLPA PLLAALGAAI</p> <p>PSIPRED  </p> <p>JPRED4  </p>	<p>JPRED4 and PSIPRED results had predicted similar 2d structures protein in the given query sequence.</p>
<p>QUERY  <span style="margin-left: 100px;">10</span>  EAAAKGIINTLQKYYCRVRI</p> <p>PSIPRED  </p> <p>JPRED4  </p>	<p>JPRED4 and PSIPRED results had predicted two different 2d structures of proteins in the given query sequence.</p> <p>For example, in the given query sequence, where <b>PSIPRED</b> was predicting a <b>possible Coil region</b> from <b>1-13th</b> amino acid residue, with the same query sequence, <b>JPRED4</b> was predicting a <b>possible Helix region</b> from starting from the <b>5-18th</b> amino acid residue</p>

**For Vaccine 1**

In case of vaccine 1, we had observed that there are several regions where PSIPRED and JPRED had predicted similar results.

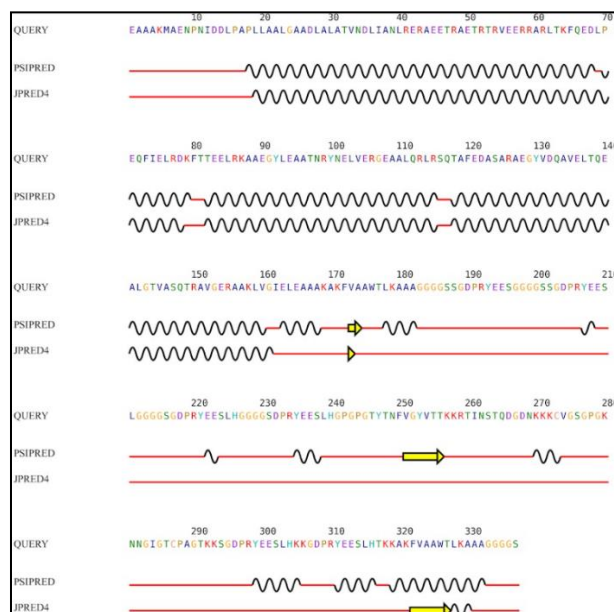


**Figure 3.22: 2dss comparison between PSIPRED and JPred4's secondary structure prediction for Vaccine 1**

At the same time, we also observed regions where there had been mismatch results of predictions from both PSIPRED and JPRED4(Figure 3.22).

**For Vaccine 2**

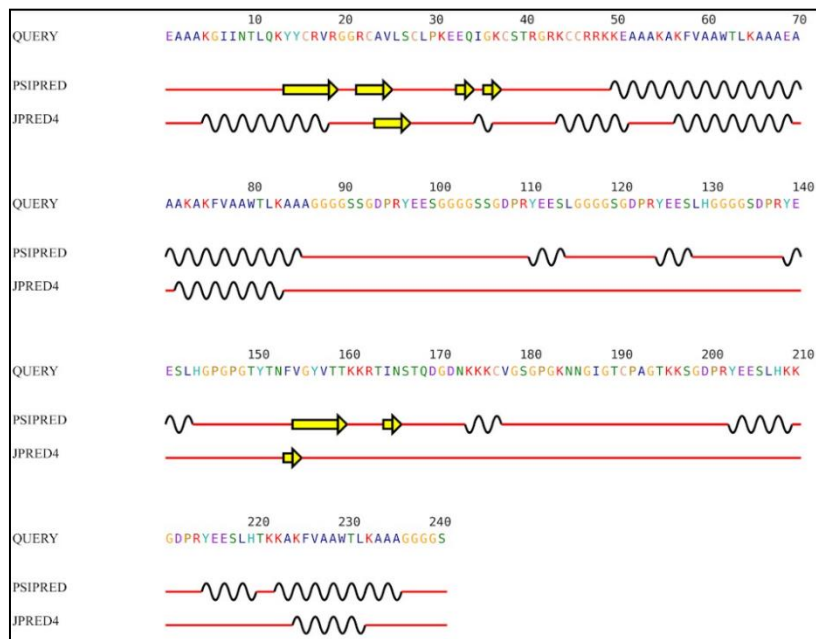
In the case of Vaccine 2, we observe that both tools had predicted almost similar results (Figure 3.23) although there are some regions where tools made different predictions.



**Figure 3.23: 2dss comparison between PSIPRED and JPred4 's secondary structure prediction for Vaccine 2**

### **For Vaccine 3**

In the case of Vaccine 3(**Figure 3.24**), we again observed there are match and mismatch areas practiced by both of the tools

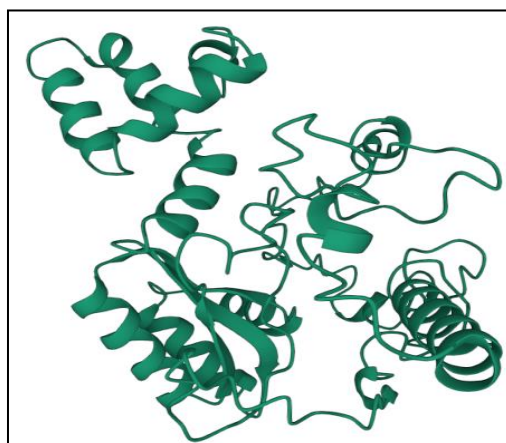


**Figure 3.24: 2ds comparison between PSIPRED and JPred4 's secondary structure prediction for Vaccine 3**

### **3.14 3D Structure of Vaccine Constructs**

3d structures generated from the I-TASSER provided data. for 3 of our vaccines.

#### **Vaccine 1**

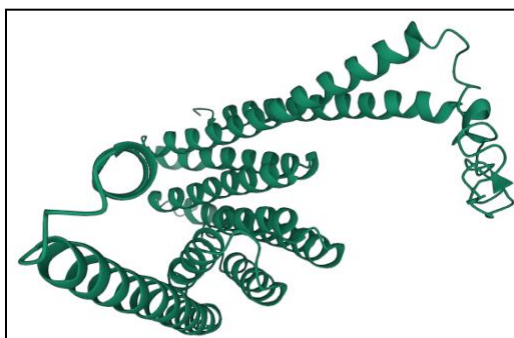


**Figure 3.25: Vaccine 1 3d model structure**

In case of vaccine 1(**Figure 3.25**), we observed multiple amino acids regions in the peptide chain, which were identified as GLU (Glutamic acid), ALA, LYS, MET, LEU, SER, THR(Threonine), ASP, PHE(Phenylalanine), VAL, PRO, ILE, ARG, TYR,

HIS(Histidine), GLY, ASN(Asparagine), PHE(Phenylalanine), GLN(Glutamine), CYS, TRP.

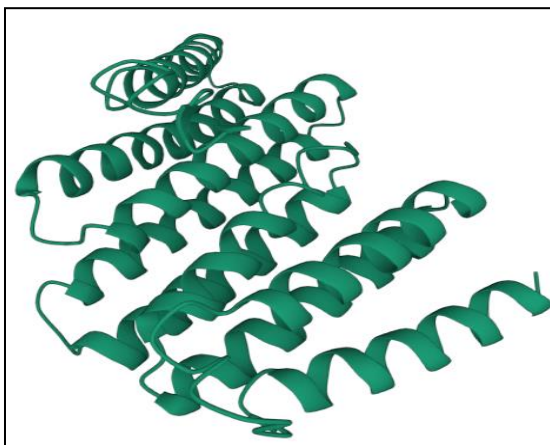
### Vaccine 2



**Figure 3.26: Vaccine 2 3d model structure**

For Vaccine 2(**Figure 3.26**) we identified some regions as well, which are GLU, ALA, LYS, MET, ASN, ILE, PRO, ASP, LEU, GLY, THR, VAL, ARG, THR, PHE, GLN, SER, TRP, HIS

### Vaccine 3



**Figure 3.27: Vaccine 3 3d model structure**

Amino acids by which vaccine 3(**Figure 3.27**) was consist of were GLU, ALA, LYS, GLY, ILE, ASN, THR, LEU, GLN, TYR, CYS, ARG, VAL, SER, PRO, ILE, HIS, PHE, TRP.

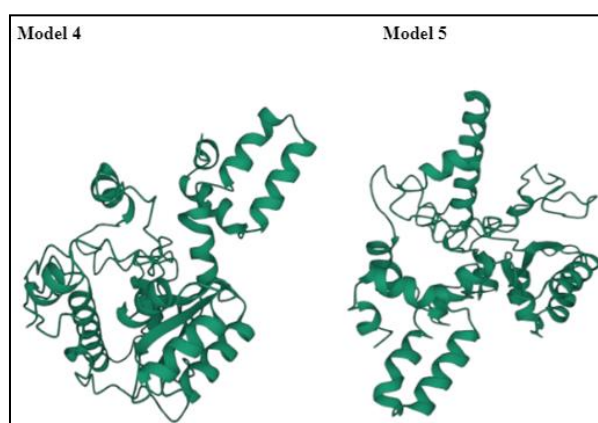
From the results of **Galaxy Refine**, after repacking the side chains of the 3d structures, we have been provided 5 models (Model 1, Model 2, Model 3, Model 4, Model 5) for each of the vaccines, thus obtaining 15 results altogether. In order to obtain the best result data, we observed our vaccine models through SAVES and PDBsum generated a combined database (**Table 3.28**), and retrieved the best vaccine model from the SAVES and PDBsum data we conducted a compared analysis in order to detect the best model

from the 15 models that we had obtained earlier

**Table 3.28: Comparing analysis between SAVES and PDBsum result**

	IT_V1	IT_V2	IT_V3
Main SAVES Pass	0.5	1.5	1.5
Main PDBsum Score	-0.84 (2.4%)	-0.56 (3.2%)	-0.49 (3.2%)
Model 1 SAVES Pass	1	1.5	1.5
Model 1 PDBsum Score	-0.17 (2%)	0.00 (1.4%)	0.04 (1.6%)
Model 2 SAVES Pass	1.5	1.5	1.5
Model 2 PDBsum Score	-0.16 (2.8%)	0.02 (1.4%)	0.04 (2.1%)
Model 3 SAVES Pass	1	1	1.5
Model 3 PDBsum Score	-0.19 (2.4%)	0.01 (1.4%)	0.06 (1.6%)
Model 4 SAVES Pass	2.5	1	1.5
Model 4 PDBsum Score	-0.18 (2.4%)	0.02 (1.4%)	0.06 (2.1%)
Model 5 SAVES Pass	2	1	1
Model 5 PDBsum Score	-0.18 (2.8%)	-0.02 (1.8%)	0.07 (1.6%)

### Best models for Vaccine 1

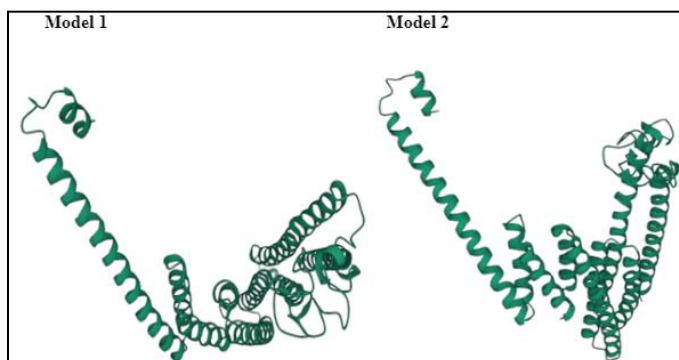


**Figure 3.28: Vaccine 1 best model (Model 4 and Model 5)**

In the case of vaccine 1, we selected 2 best candidates (Figure 3.28), which were model 4 and model 5. According to SAVES database model 4 had a score of 2.5 and model 5 had a score of 2 furthermore, based on PDBsum Score model 4 had a score of -0.18 (2.4%) and model 5 had a score of -0.18 (2.8%). Thus Model 4 was selected as the best model for Vaccine 1.

### **Best model for Vaccine 2**

**In the case of vaccine 2 (Figure 3.29), we selected 2 best candidates, which were model 1 and model 2.**

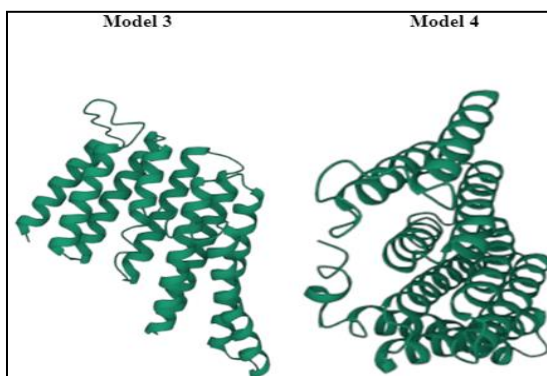


**Figure 3.29: Vaccine 2 best model (Model 1 and Model 2)**

Based on, **SAVES database** model 1 had a score of **1.5** similarly model 2 had a score of **1.5**, furthermore, based on **PDBsum Score**, model 1 has a score of **0.00 (1.4%)** and model 2 has a score of **0.02 (1.4%)**. Thus **Model 2** was selected as the best model for **Vaccine 2**.

### **Best model for Vaccine 3**

**In the case of vaccine 3, we selected 2 best candidates (Figure 3.30), which were, model 3 and model 4.**



**Figure 3.30: Vaccine 3 best model (Model 3 and Model 4)**

According to **SAVES database** model 3 had a score of **1.5** similarly model 4 had a score of **1.5**, furthermore, based on **PDBsum Score**, model 3 has a score of **0.06 (1.6%)** and model 4 has a score of **0.06 (2.1%)**. Thus **Model 3** was selected for as the best model for **Vaccine 3**.

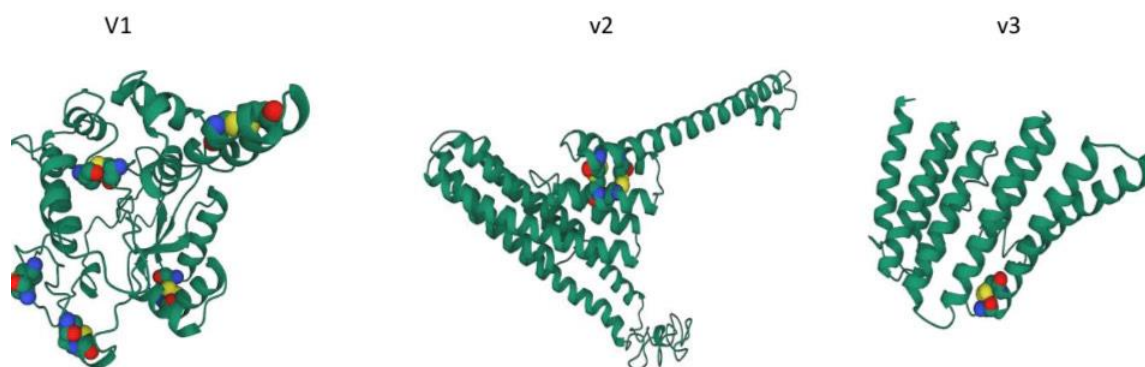
### 3.15 Disulfide engineering of vaccines

By disulfide engineering of the three constructed vaccines: V1, V2 and V2, possible disulfide positions in the amino acid residues (with binding energy less than 2.2 kcal/mol) were analyzed and selected for designing the disulfide bonds. In V2, one bond was left out from designing because it was resulting in discontinuous protein. After protein disulfide engineering of the vaccines, it has been found that V1 added 5 bonds, V2 added 2 bonds, and V3 added 1 bond (**Table 3.29**). And, visualization of the 3D structure of these three mutated vaccines by RCSB PDB-Mol\* 3D Viewer indicated that disulfide bond formation was undergone into 5 pairs of the V1, 2 pairs of V2 and 1 pair of V3 (**Figure 3.31**). The mutated PDBs were downloaded for each vaccine, and then processed into standard PDB files using UCSF chimera and stored the files for further steps.

**Table 3.29: The mutated positions in chain A of the V1, V2 and V3**

Name of the vaccines (Disulfide modified)	Number of bonds	Mutated positions
V1	5 pairs	1. A  CYS 33 - A  CYS 45 2. A  CYS 78 - A  CYS 129 3. A  CYS 146 - A  CYS 257 4. A  CYS 190 - A  CYS 306 5. A  CYS 204 - A  CYS 206
V2	2 pairs	1. A  CYS 55 - A  CYS 70 2. A  CYS 88 - A  CYS 91
V3	1 pair	1. A  CYS 174 - A  CYS 177

- a) The 3D structures of disulfide modified vaccines with the disulfide bonds were retrieved (**Figure 3.31**).



**Figure 3.31: Representation of the 3D structure of the disulfide modified vaccines visualization using RCSB PDB - Mol\* 3D Viewer (disulfide bonds are indicated with the ball shaped structure)**

- b) As these modified vaccines are considered as the final vaccine candidates, they were subjected to their properties analysis. For this, the FASTA format sequences for all three vaccines were generated using Chimera and stored in a file.
- c) All of the three vaccines (V1, V2 and V3) has been found to be as probable antigen with the scores of 0.5413, 0.5644 and 0.5893 respectively in the VaxiJen 2.0 server where threshold was set at 0.4 and the organism was selected as virus. Also, they appeared as non-allergen in the AllerTOP v.2.0. Therefore, from ProtParam analysis, the Vaccine 1 and 2 have comparatively similar isoelectric point or Theoretical PI which is 5.4 and 5.46 respectively, whereas vaccine 3 have 9.42 Theoretical PI value. Besides, the extinction coefficient of Vaccine 3 was higher than Vaccine 1 and 2 and it has the lowest GRAVY. Furthermore, these three vaccines were found to be stable and they had similar estimated half-life (**Table 3.30**).



**Table 3.30: The antigenicity, allergenicity and physicochemical properties analysis of the modified vaccines using the FASTA sequences**

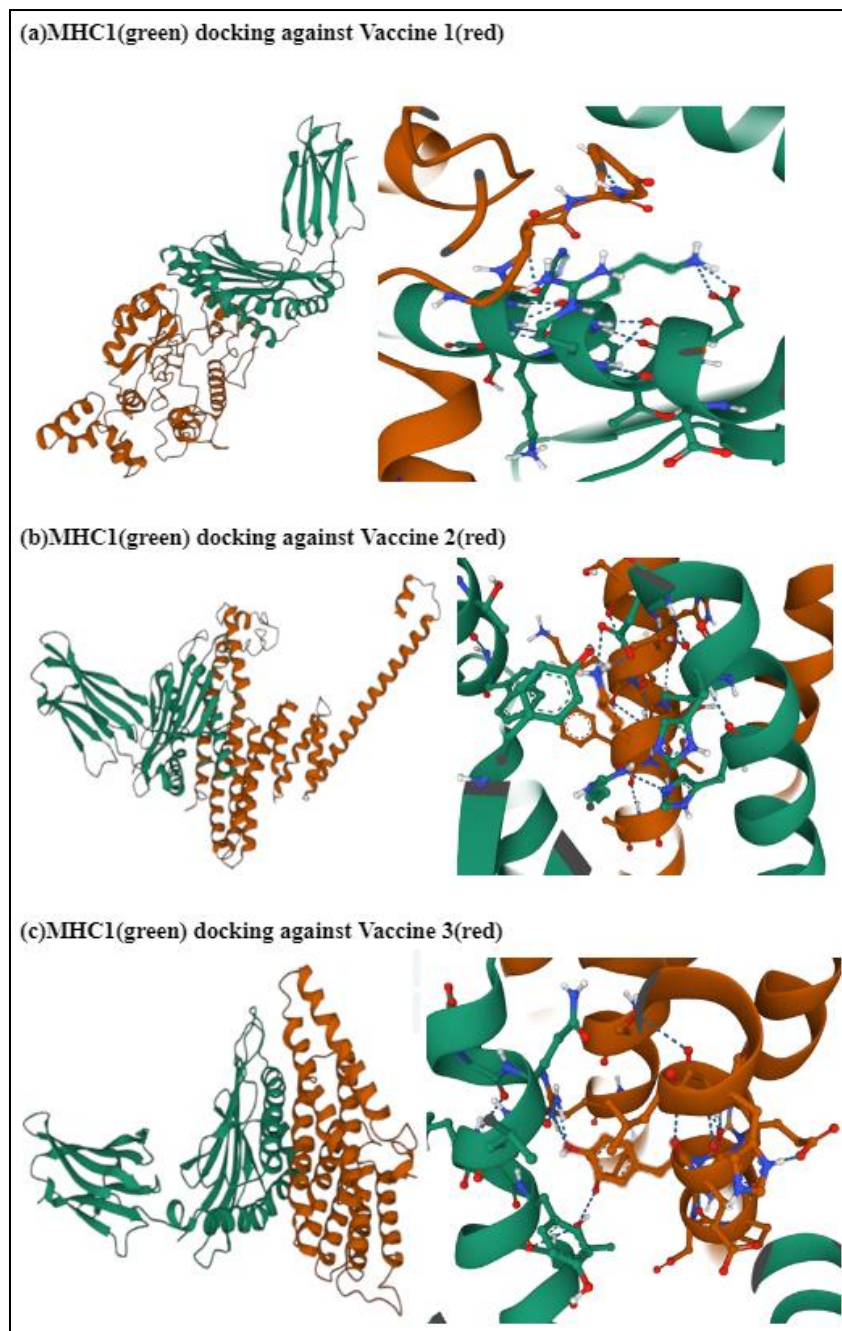
		Name of the DS modified vaccines		
		Modified V1 chainA_1-308	Modified V1 chainA_1-337	Modified V1 chainA_1-241
	<b>Antigenicity</b> [Vaxijen v2.0] Threshold 0.4 Target organism: Virus	Probable <b>Antigen</b> 0.5413	Probable <b>Antigen</b> 0.5644	Probable <b>Antigen</b> 0.5893
	<b>Allergenicity</b> [AllerTop v.2.0]	Non-Allergen Nearest Protein: UniProtKB accession number Q75I13	Non-Allergen Nearest Protein: UniProtKB accession number Q75I13	Non-Allergen Nearest Protein: UniProtKB accession number Q75I13
<b>Physicochemical Properties [ProtParam]</b>	<b>Number of amino acids</b>	308	337	241
	<b>Molecular weight</b>	31550.36	35813.49	25032.96
	<b>Theoretical pI</b>	5.4	5.46	9.42
	<b>Total number of negatively charged residues (Asp + Glu)</b>	47	53	25
	<b>Total number of positively charged residues (Arg +</b>	40	46	40

	Name of the DS modified vaccines		
	Modified V1 chainA_1-308	Modified V1 chainA_1-337	Modified V1 chainA_1-241
Lys)			
<b>Formula</b>	C1359H2177N38 1O453S14	C1534H2446N46 0O516S7	C1076H1711N32 9O343S9
<b>Total number of atoms</b>	4384	4963	3468
<b>Extinction coefficients</b> $M^{-1} cm^{-1}$ , at 280 nm	Ext. coefficient 22180 Abs 0.1% (=1 g/l) 0.703, assuming all pairs of Cys residues form cystines Ext. coefficient 21430 Abs 0.1% (=1 g/l) 0.679, assuming all Cys residues are reduced	Ext. coefficient 27765 Abs 0.1% (=1 g/l) 0.775, assuming all pairs of Cys residues form cystines Ext. coefficient 27390 Abs 0.1% (=1 g/l) 0.765, assuming all Cys residues are reduced	Ext. coefficient 31900 Abs 0.1% (=1 g/l) 1.274, assuming all pairs of Cys residues form cystines Ext. coefficient 31400 Abs 0.1% (=1 g/l) 1.254, assuming all Cys residues are reduced
<b>Estimated half-life</b>	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours ( <i>Escherichia coli</i> ,	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours ( <i>Escherichia coli</i> ,	1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours ( <i>Escherichia coli</i> ,

	Name of the DS modified vaccines		
	Modified V1 chainA_1-308	Modified V1 chainA_1-337	Modified V1 chainA_1-241
	in vivo)	in vivo)	in vivo)
<b>Instability index II</b>	29.00 Stable	36.49 Stable	28.40 Stable
<b>Aliphatic index</b>	64.81	62.17	47.55
<b>Grand average of hydropathicity (GRAVY)</b>	-0.396	-0.725	-0.773

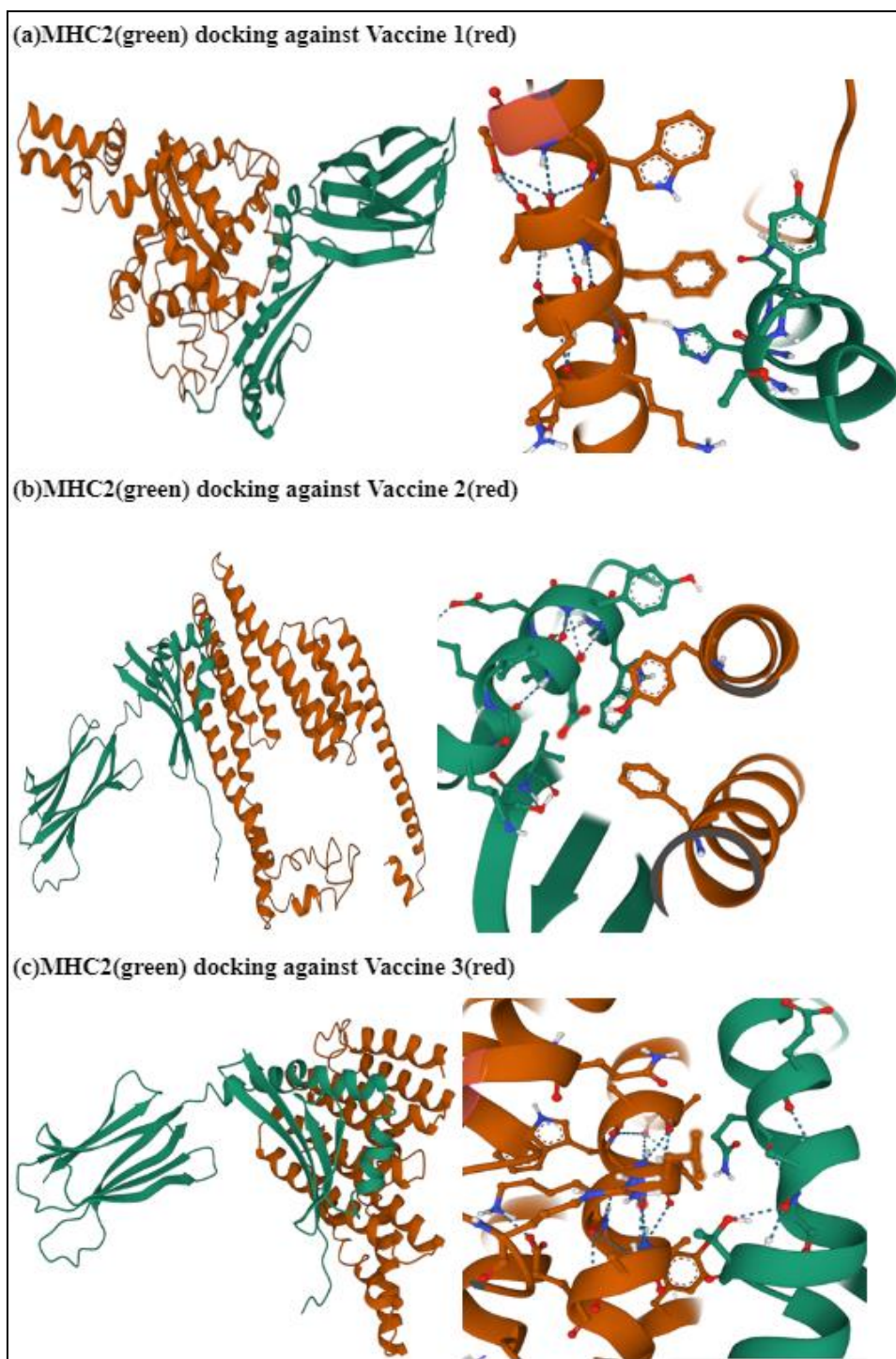
### 3.16 Molecular Docking Analysis of the Vaccine Constructs against MHC I and MHC II

In case of **MHC-I's** 1QEW (**Figure 3.32**), we observed 3 clusters for each of 3 vaccines docked, which were **V1\_cluster 2\_4**, **V2\_cluster 1\_1** and **V3\_cluster 5\_1**.



**Figure 3.32: a) MHC 1 docking against Vaccine 1 Cluster 2\_4    b) MHC 1 docking against Vaccine 2 Cluster 1\_1    c) MHC 1 docking against Vaccine 3 Cluster 5\_1**

Again, for **MHC-II 2G9H (Figure 3.33)**, we observed 3 clusters for each of 3 vaccines which were V1\_cluster 10\_1, V2\_cluster 4\_1 and V3\_cluster 6\_2.



**Figure 3.33: a) MHC II docking against Vaccine 1 Cluster 10\_1 b) MHC II docking against Vaccine 2 Cluster 4\_1 c) MHC II docking against Vaccine 3 Cluster 6\_2**

We observed and merged the binding affinity for 6 clusters models, based on the data set provided by PRODIGY (Table 3.31) score,

**Table 3.31: Cluster model binding affinity score analysis**

Epitopes	Types	Cluster No	Model No	Binding Energy
MHC-I	1QEW vs V1	2	4	-14.5
MHC-I	1QEW vs V2	1	1	-15.3
MHC-I	1QEW vs V3	5	1	-16.5
MHC-II	2G9H vs V1	10	1	-15.8
MHC-II	2G9H vs V2	4	1	-13.6
MHC-II	2G9H vs V3	6	2	-15.3

As we know the lower the binding energy the better the bonding would take place, thus, based on scores (**Table 3.31**), we analyzed that, For **MHCI** epitope in case of **1QEW**, we had obtained 3 results, for Vaccine 1, our **cluster number** was **2** and **model number** was **4** and the binding energy score was **-14.5**. Again, for Vaccine 2, **our cluster number** was **1** and **model number** was **1** and the **binding energy score** was **-15.3**. Lastly for Vaccine 3, the best suitable **cluster number** was **5** and **best model** was **model no 1**, while having a **-16.5 binding energy** score. Similarly, for **MHC II**(**Table 3.31**) epitope in case of **2G9H**, we had obtained 3 results, for Vaccine 1, our **cluster number** was **10** and model number was 1 and the **binding energy score** was **-15.8**. Again, for Vaccine 2, our **cluster number** was **4** and model number was **1** and the **binding energy score** was **-13.6**. Lastly for Vaccine 3, the best suitable **cluster number** was **6** and best **model** was **model no 2**, while having a **-15.3 binding energy score**.

### 3.17 B-lymphocytic Epitope Prediction

The continuous and discontinuous B- lymphocytic epitopes were predicted from the vaccine's 3D structure and generated by the ElliPro tool. Continuous (**Figure 3.34**) and Discontinuous epitopes were found from all of the three vaccines (**Figure 3.35, Figure 3.36, Figure 3.37**).

From vaccine 1, a total 308 residues were found. Among them 12 continuous (score varies from 0.543 to 0.845 and length varies from 8 to 30 amino acids) (**Table 3.32, Figure 3.34**) and 7 discontinuous (score varies from 0.54 to 0.756 and length varies from 3 to 44 amino acids) (**Table 3.35, Figure 3.35**) epitopes were predicted.

From vaccine 2, a total 337 residues were found. Among them 6 continuous (score varies from 0.546 to 0.834 and length varies from 10 to 61 amino acids) (**Table 3.33, Figure 3.34**) and 4 discontinuous (score varies from 0.61 to 0.862 and length varies from 8 to 61 amino acids) (**Table 3.36, Figure 3.36**) epitopes were predicted.

From vaccine 3, a total 241 residues were found. Among them 7 continuous (score varies from 0.562 to 0.82 and length varies from 7 to 34 amino acids) (**Table 3.34, Figure 3.34**) and 4 discontinuous (score varies from 0.649 to 0.767 and length varies from 18 to 45 amino acids) (**Table 3.37, Figure 3.37**) epitopes were predicted.

**Table 3.32: List of the predicted Continuous epitopes for V1**

	No	Start	End	Sequence	Length	Score
V1	1	24	47	LELSDFVKKCEETFEVTAAAPCAV	24	0.845
	2	1	13	EAAAKMAKLSTDE	13	0.751
	3	202	215	SDCRCEESLHGPGP	14	0.739
	4	263	292	KKSGDPRYEESLHKKGDPRYEESL HTKKAK	30	0.736
	5	183	195	GGGGSGDCRYEES	13	0.673
	6	75	88	AGDCKIGVIKVVRE	14	0.667
	7	90	96	VSGLGLK	7	0.643
	8	246	254	GSGPGKNNG	9	0.644
	9	123	130	AKLEAACA	8	0.602
	10	304	308	GCCGS	5	0.579
	11	156	176	GGSSGDPRYEESGGGGSSGDP	21	0.576
	12	228	235	KKRTINST	8	0.543

**Table 3.33: List of the predicted Continuous epitopes for V2**

	No	Start	End	Sequence	Length	Score
V2	1	1	46	EAAAKMAENPNIDDLAPLLA ALGAADLALATVNDLIANLRE RAEE	46	0.834
	2	247	307	YTNFVGIVTTKKRTINSTQDG DNKKKCVGSGPGKNNGIGTCP	61	0.761

No	Start	End	Sequence	Length	Score
			AGTKKSGDPRYEESLHKKG		
3	147	163	SQTRAVGERAAKLVGIE	17	0.665
4	200	223	GSSGDPRYEESLGGGGSGDPR YEE	24	0.665
5	328	337	LKAAAGGGGS	10	0.633
6	49	60	AETRTRCEERRA	12	0.56

**Table 3.34: List of the predicted Continuous epitopes for V3**

No	Start	End	Sequence	Length	Score	
V3	1	193	216	AGTKKSGDPRYEESLHK KGDPRYE	24	0.82
	2	113	123	ESLGGGGSGDP	11	0.756
	3	1	34	EAAAKGIINTLQKYYCR VRGGRCVLSCLPKEEQ	34	0.742
	4	140	150	EESLHGPGPGT	11	0.66
	5	226	241	FVAAWTLKAAAGGGGS	16	0.641
	6	171	179	GDNCKKCVG	9	0.627
	7	68	74	AEAAAKA	7	0.562

**Table 3.35: List of the predicted Discontinuous epitopes for V1**

V1			
No	Residues	No. of Residues	Score
1	A:E1, A:A2, A:A3, A:A4, A:K5, A:M6, A:A7, A:K8, A:L9, A:S10, A:T11, A:D12, A:E13, A:D16, A:A17, A:E20, A:M21, A:T22, A:L24, A:E25, A:S27, A:D28, A:F29, A:V30, A:K31, A:K32, A:C33, A:E34, A:E35, A:T36, A:F37, A:E38, A:V39, A:T40, A:A41, A:A42, A:A43, A:P44, A:C45, A:A46, A:V47, A:A49, A:A50, A:G51	44	0.756
2	A:V245, A:G246, A:S247, A:G248, A:P249, A:G250, A:K251, A:N252, A:N253, A:G254, A:T262, A:S265,	35	0.72



V1			
No	Residues	No. of Residues	Score
	A:G266, A:D267, A:P268, A:R269, A:Y270, A:E272, A:S273, A:L274, A:K276, A:G278, A:D279, A:P280, A:R281, A:Y282, A:E283, A:E284, A:S285, A:L286, A:H287, A:T288, A:K289, A:K290, A:K292		
3	A:A75, A:G76, A:D77, A:C78, A:K79, A:I80, A:G81, A:I83, A:K84, A:V85, A:R87, A:E88, A:V90, A:S91, A:G92, A:L93, A:G94, A:L95, A:K96, A:K99, A:A123, A:K124, A:E126, A:A127, A:A128, A:C129, A:A130	27	0.668
4	A:G183, A:G184, A:G185, A:G186, A:S187, A:G188, A:D189, A:C190, A:R191, A:Y192, A:E193, A:E194, A:S195, A:L196, A:G198, A:G199, A:G200, A:G201, A:S202, A:D203, A:C204, A:R205, A:C206, A:E207, A:E208, A:S209, A:L210, A:H211, A:G212, A:P213, A:G214, A:P215, A:G216, A:G304, A:G305, A:C306, A:G307, A:S308	38	0.641
5	A:G155, A:G156, A:G157, A:S158, A:S159, A:G160, A:D161, A:P162, A:R163, A:E165, A:E166, A:S167, A:G168, A:G169, A:G170, A:G171, A:S172, A:S173, A:G174, A:D175, A:P176	21	0.579
6	A:I232, A:N233, A:S234, A:T235, A:G238, A:D239, A:K242	7	0.577
7	A:T227, A:K229, A:T231	3	0.54

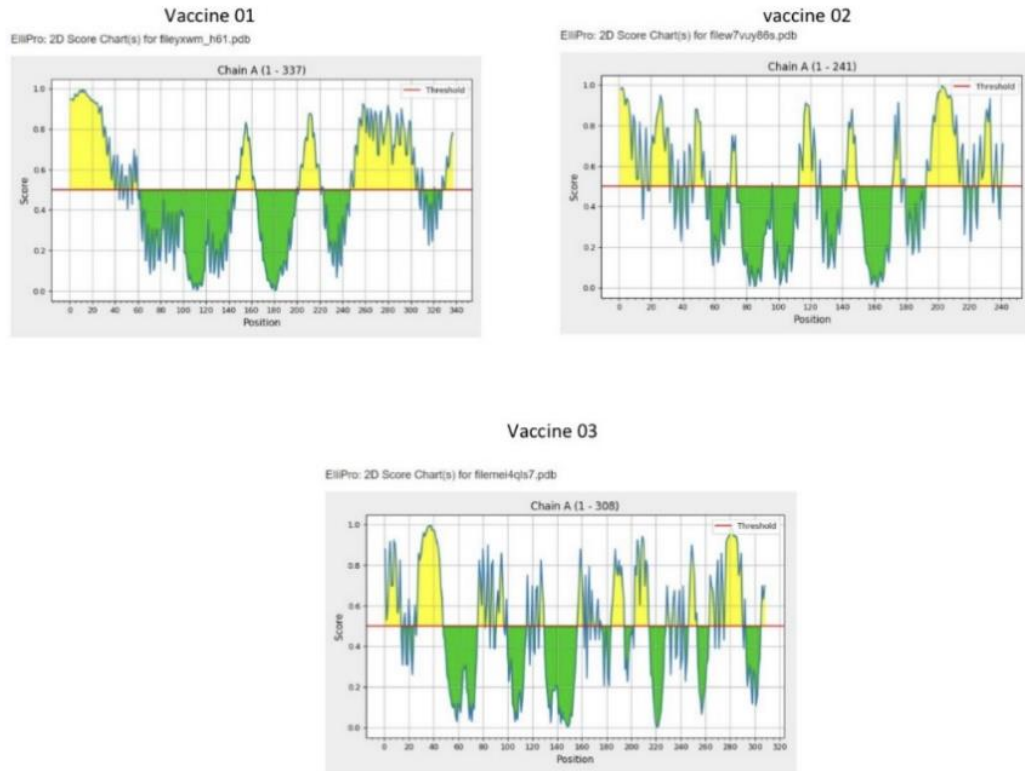
**Table 3.36: List of the predicted Discontinuous epitopes for V2**

V2			
No	Residues	No. of Residues	Score
1	A:E1, A:A2, A:A3, A:A4, A:K5, A:M6, A:A7, A:E8, A:N9, A:P10, A:N11, A:I12, A:D13, A:D14, A:L15, A:P16, A:A17, A:P18, A:L19, A:L20, A:A21, A:A22, A:L23,	42	0.862

V2			
No	Residues	No. of Residues	Score
	A:G24, A:A25, A:A26, A:D27, A:L28, A:A29, A:L30, A:A31, A:T32, A:V33, A:N34, A:D35, A:L36, A:I37, A:A38, A:N39, A:R41, A:E42, A:R43		
2	A:Y247, A:T248, A:N249, A:F250, A:V251, A:G252, A:Y253, A:V254, A:T255, A:T256, A:K257, A:R259, A:T260, A:I261, A:N262, A:S263, A:T264, A:Q265, A:D266, A:G267, A:D268, A:N269, A:K270, A:K271, A:K272, A:C273, A:V274, A:G275, A:S276, A:G277, A:P278, A:G279, A:K280, A:N281, A:N282, A:G283, A:I284, A:G285, A:T286, A:C287, A:P288, A:A289, A:G290, A:T291, A:K292, A:K293, A:S294, A:G295, A:D296, A:P297, A:R298, A:Y299, A:E300, A:E301, A:S302, A:L303, A:H304, A:K306, A:G307, A:R310, A:Y311	61	0.757
3	A:S147, A:Q148, A:T149, A:R150, A:A151, A:V152, A:G153, A:E154, A:R155, A:A156, A:A157, A:K158, A:L159, A:V160, A:G161, A:I162, A:E163, A:G200, A:S202, A:G203, A:D204, A:P205, A:R206, A:Y207, A:E208, A:E209, A:S210, A:L211, A:G212, A:G213, A:G214, A:G215, A:S216, A:G217, A:D218, A:P219, A:R220, A:E222, A:E223, A:L328, A:K329, A:A331, A:A332, A:G333, A:G334, A:G335, A:G336, A:S337	48	0.67
4	A:A49, A:E50, A:R52, A:T53, A:R54, A:E56, A:E57, A:R59	8	0.61

**Table 3.37: List of the predicted Discontinuous epitopes for V3**

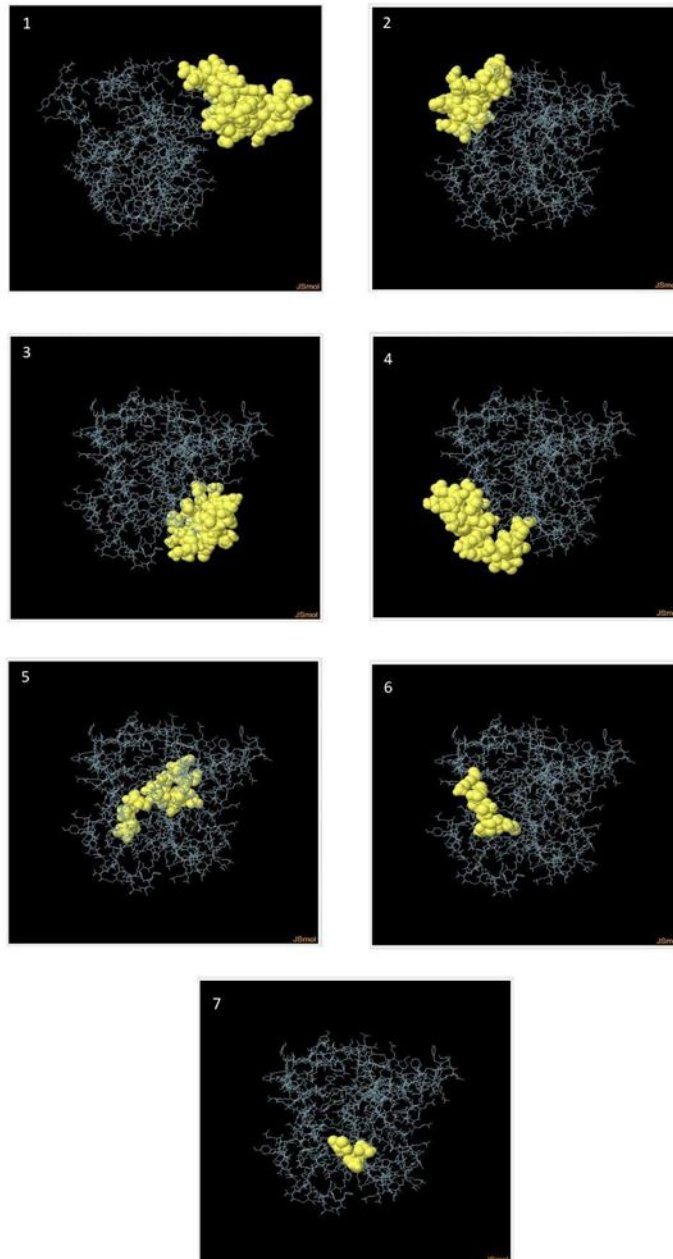
V3			
No	Residues	No. of Residues	Score
1	A:E140, A:E141, A:L143, A:H144, A:G145, A:P146, A:G147, A:P148, A:G149, A:T150, A:A193, A:G194, A:T195, A:K196, A:K197, A:S198, A:G199, A:D200, A:P201, A:R202, A:Y203, A:E205, A:S206, A:L207, A:H208, A:K209, A:K210, A:G211, A:D212, A:P213, A:R214, A:Y215, A:E216, A:L219	34	0.767
2	A:E1, A:A2, A:A3, A:A4, A:K5, A:G6, A:I7, A:I8, A:N9, A:T10, A:L11, A:Q12, A:K13, A:R41, A:K44, A:C45, A:R47, A:R48, A:K49, A:E51, A:A52, A:A53, A:A54, A:K57	24	0.758
3	A:Y14, A:C16, A:R17, A:V18, A:R19, A:G20, A:G21, A:R22, A:C23, A:A24, A:V25, A:L26, A:S27, A:C28, A:L29, A:P30, A:K31, A:E33, A:Q34, A:K37, A:A68, A:A70, A:A71, A:A72, A:K73, A:A74, A:E113, A:S114, A:L115, A:G116, A:G117, A:G118, A:G119, A:S120, A:G121, A:D122, A:P123, A:E126, A:G171, A:D172, A:N173, A:C174, A:K175, A:V178, A:G179	45	0.7
4	A:K176, A:K223, A:F226, A:V227, A:A228, A:A229, A:W230, A:T231, A:L232, A:K233, A:A234, A:A235, A:A236, A:G237, A:G238, A:G239, A:G240, A:S241	18	0.649



**Figure 3.34: 2D score charts for three vaccines (V1, V2, V3)**

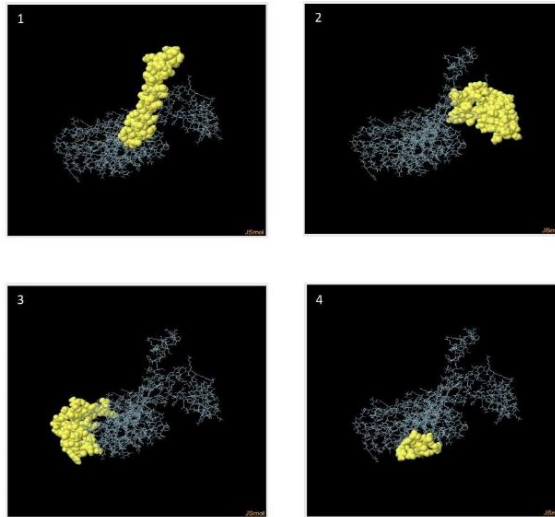
**JSmol-Rendered PDB Structure of predicted Discontinuous epitopes for three vaccines:**

**Vaccine 1**



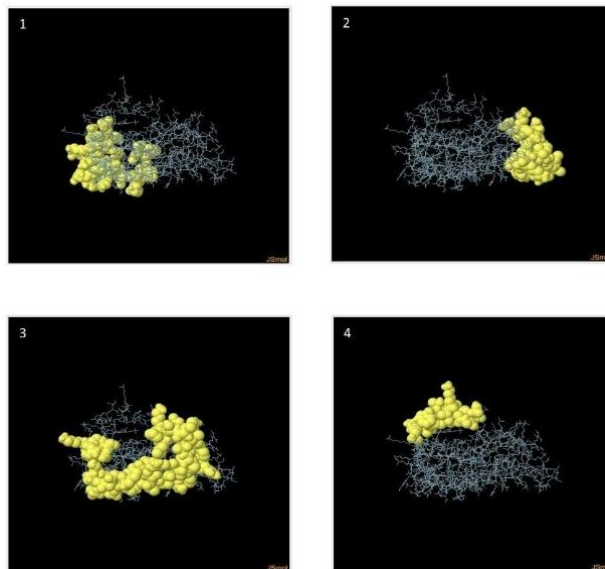
**Figure 3.35: The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine 1**

### Vaccine 2



**Figure 3.36: The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine 2**

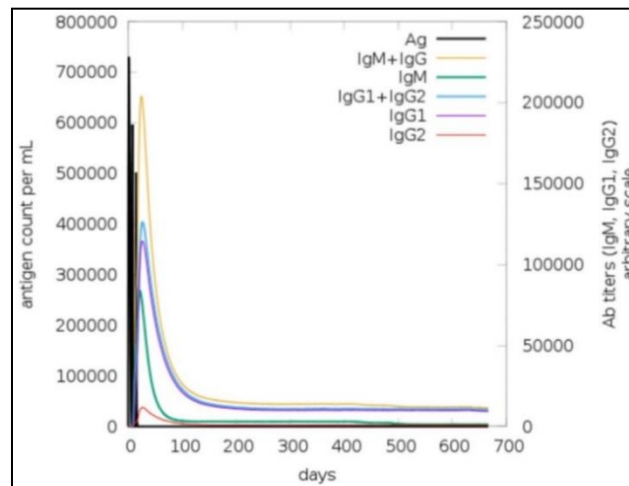
### Vaccine 3



**Figure 3.37: The predicted Discontinuous B-cell epitopes (yellow colored ball-shaped structures) in Vaccine 3**

### 3.18 Immune Simulation of the Vaccines

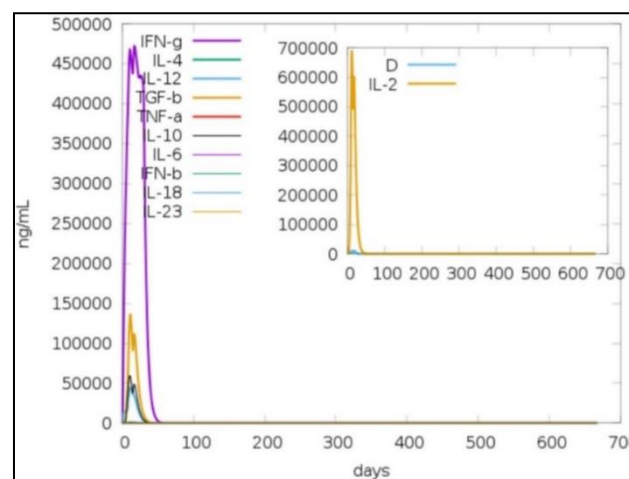
#### Vaccine 1:



**Figure 3.38: Vaccine 1\_virus, immunoglobulins and immunocomplexes**

Based on C-IMMSIM simulation for our vaccine 1 Ab titters (**Figure 3.38**), we observed that our vaccine 1 would stay in the system for a period of **600-700** days. From **day 0- 100** the **IgM+ IgG** count was between **600000-700000** *antigen count per ml*, referring to the peak point of our graph, after which the number had been seen to be reduced gradually over time. Again **IgG1 + IgG2** (**Figure 3.38**) ratio was around **400000** *antigen count per ml* from day 0-100, then gradual reduction in number was seen over time.

#### Graph for Vaccine-1 Concentration of cytokines and interleukins

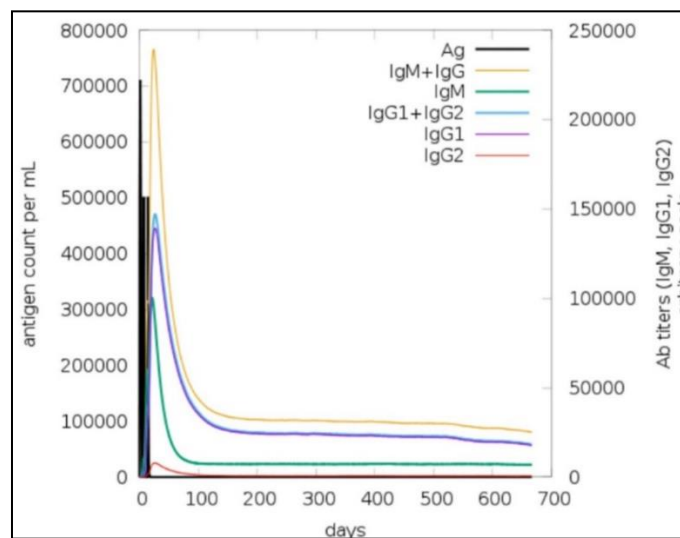


**Figure 3.39: Vaccine 1\_concentration of cytokines and interleukins**

In case of Vaccine 1(**Figure 3.39**), we again analyzed another graph for Concentration of cytokines and interleukins. Here we observed the curve for IFN-g growth ratio, as

**IFN-g/ Type-2** interferon is a type of **cytokine** which plays vital role in the **innate and adaptive immunity** against different types of pathogens, based on the graph ranging from **day 0 to day 100**, the ratio of IFN-g is between **450000-500000 ng/ml**. Again, the graph further showed us, **Inset plot showing danger signal together with leukocyte growth factor IL-2, a type of cytokine** which assists in **regulating the WBC** and a type of **signaling molecule** in immune system, We concluded that the danger signal(**D-in the graph, Figure 3.39**) based on the graph was **very minimal** ranging **between 0-100 days**, furthermore, in case of **IL-2**, the ratio was in between **600000-700000 ng/ml, indicating peak point**, within the range of **0-100 days** then gradually **decreasing in a steady rate**.

**Vaccine 2:**



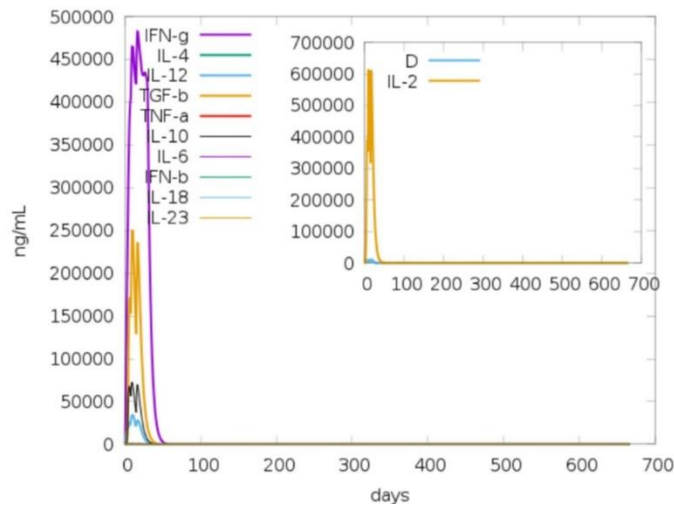
**Figure 3.40: Vaccine 2\_virus, immunoglobulins and immunocomplexes**

Based on C-IMMSIM simulation for our vaccine 2 Ab titters (**Figure 3.40**), we observed that our vaccine 1 would stay in the system for a period of **600-700 days** overall. From **day 0- 100** the **IgM+ IgG count** was between **700000-800000** antigen count per ml, referring to the peak point of our graph, after which the number had been seen to be reduced gradually over time.

Again, **IgG1 + IgG2(Figure 3.40)** ratio was **400000-500000** antigen count per ml from day **0-100** then gradual reduction in number were seen over time at the range between **0-100000** antigen count per ml between day **600-700**



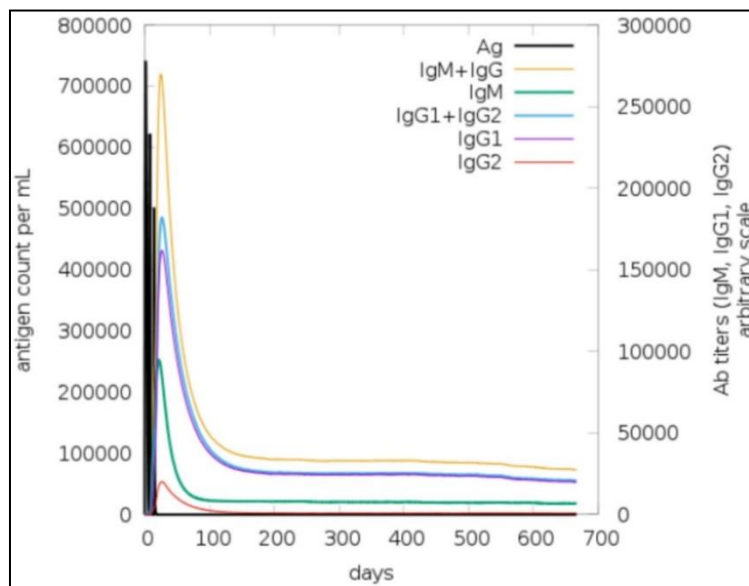
**Graph for Vaccine-2 Concentration of cytokines and interleukins**



**Figure 3.41: Vaccine 2\_concentration of cytokines and interleukins**

In the case of Vaccine 2(Figure 3.41), we observed the curve for IFN-g growth ratio, from **day 0 to day 100**, the ratio of IFN-g is between **450000-500000 ng/ml**. Again, the graph further showed us, **inset plot showing danger signal together with leukocyte growth factor IL-2**, we observed that the danger signal (**D-in the graph**, Figure 3.41) based on the graph was **very minimal** ranging **between 0-100 days**, furthermore, in case of **IL-2**, the ratio was **a little over 600000 ng/ml**, within the range of **0-100 days**, then gradually **decreasing in a steady rate**.

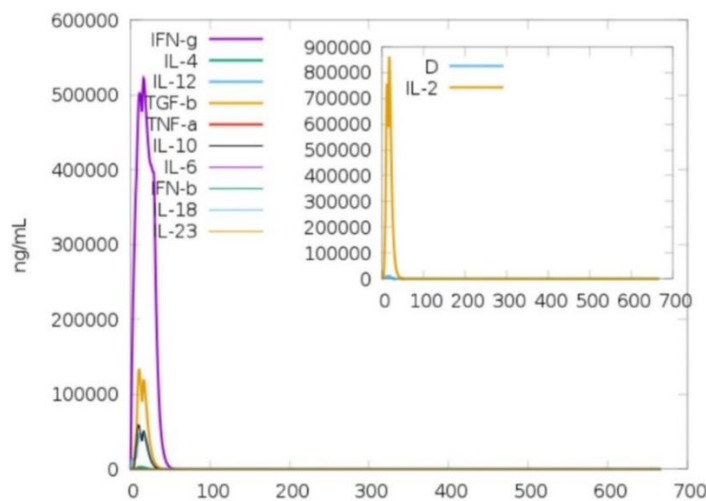
**Vaccine 3:**



**Figure 3.42: Vaccine 3\_virus, immunoglobulins and immunocomplexes**

Based on C-IMMSIM simulation for our vaccine 3 Ab titers (**Figure 3.42**), we observed that our vaccine 1 would stay in the system for a period of **600-700** days overall. From **day 0- 100** the **IgM+ IgG count** was **a little over 700000 antigen count per ml**, referring to the peak point of our graph, after which the number had been seen to be reduced gradually over time. Again **IgG1 + IgG2 (Figure 3.42)** ratio was **400000-500000 antigen count per ml** from day **0-100** then gradual reduction in number were seen over time at the range between **0-100000 antigen count per ml** between day **600-700**.

**Graph for Vaccine-3 Concentration of cytokines and interleukins**



**Figure 3.43: Vaccine 3\_concentration of cytokines and interleukins**

In the case of Vaccine 3(**Figure 3.43**), we observed the curve for IFN-g growth ratio, from **day 0 to day 100**, the ratio of IFN-g is **a little over 500000 ng/ml**. Again, the graph further showed us, **inset plot showing danger signal together with leukocyte growth factor IL-2**, we observed that the danger signal (**D-in the graph, Figure 3.43**) based on the graph was **very minimal** ranging **between 0-100 days**, furthermore, in case of **IL-2**, the ratio was **between 800000-900000 ng/ml**, within the range of **0-100** days, then gradually **decreasing in a steady rate**.

## Chapter 4. Discussion

A more focused approach in a new paradigm of vaccine creation that challenges the conventional approach and promises to be more effective is being made possible by advancements in computer science and technology, genetics and immunology, and the emergence of the new science of bioinformatics [209]. In this study, we constructed vaccines against the phylogroup II of Rabies Virus or *Lyssavirus* genus using several bioinformatics tools.

The vaccine construction comprised several steps. The targeted Glycoprotein sequences of LBV, MOKV and SHBV strain with 522 amino acid length were identified and retrieved from the UniprotKB database. All of the protein sequences were aligned using M-Coffee and conserved region were found by visualizing the aligned sequences in MEGA X software. By analysing the length of the conserved sequences, 5 conserved regions with  $\geq 9$  amino acids residues were selected to predict the overlaps with the top 10 B cell and T cell epitopes with the percentile rank less than 10. Our MHC-I epitopes based on 10 percentiles rank based observation revealed there were allele epitopes which had Common repetition 3 strains of Phylogroup-II rabies viruses (MOKV, SBV and LBV). Peptides such as **TYTNFVG YV, TYTNFVG YVT, YTNFVG YVT, YTNFVG YVTT, SGDP RYEESL, GDP RYEESL, DP RYEESLH, DP RYEESLHT, RYEESLHTP, RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTV TTT, SWLRTV TTTK, WLRTV TTTK, WLRTV TTTKE, IISPSIVEM** came repeatedly for Lagos MHC-I epitopes. For Mokola Bat Virus. For Mokola on a 10-percentile rank, we had seen **TYTNFVG YV, TYTNFVG YVT, YTNFVG YVT, YTNFVG YVTT, SGDP RYEESL, GDP RYEESL, DP RYEESLH, DP RYEESLHT, RYEESLHTP, RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTV TTT, SWLRTV TTTK, WLRTV TTTK, IISPSIVEM** peptides were in a repetitive manner. For Shimoni Bat Virus, on a 10-percentile rank, we had seen **TYTNFVG YV, TYTNFVG YVT, YTNFVG YVT, YTNFVG YVTT, SGDP RYEESL, GDP RYEESL, DP RYEESLH, DP RYEESLHT, RYEESLHTP, RYEESLHTPY, YEESLHTPY, YEESLHTPYP, EESLHTPYP, SWLRTV TTT, SWLRTV TTTK, WLRTV TTTK, IISPSIVEM** peptides were in a repetitive manner for MHC-I epitopes.

Our MHC-II epitopes based on 10 percentiles rank based observation revealed the 2 peptides which had Common repetition for 3 strains of Phylogroup II rabies viruses, they peptides were **TYTNFVGYVTT** (size -11) and **SWLRTVTTTKE** (size -11). While observing our B cell epitopes we had identified, **YNWKVSGDPRYEESLHTPYP**, as the best candidate for B cell epitope, while having the length of 20 amino acids. Furthermore, the length of the B cell epitope fell under the sequence of 117-136 amino acid length which also overlapped with the conserved region length of 121-137 amino acid range. By a combined analysis, we had generated 15 amino acids between 122 to 136 region based on 11 to 15 amino acids. A total of **2** MHC-II epitopes, **21** MHC-I epitopes and **15** B cell epitopes were obtained all together as best candidates. The antigenicity, allergenicity, toxicity and human homology of the predicted conserved epitopes were determined using bioinformatics approaches. Among all of the 38 conserved epitopes, 7 epitopes were considered as the best epitopes for vaccine construction. These 7 epitopes were filtered by analysing the antigenicity, allergenicity, toxicity and comparing the homology prediction result with the human proteome (**Table 3.13**). These epitopes were found to be highly antigenic, non-allergen, nontoxic and non-homologous to the human proteome. The epitopes must be highly antigenic because they cannot generate a strong immune response if they are not antigenic. In order to prevent any detrimental, poisonous, or allergic reactions in the body, the epitopes must once again be non-allergenic and non-toxic. Also, to be recognized as foreign antigenic sequences or particles, the epitopes must be non-identical to the human proteome.

To understand a clear picture, we divided our **7 epitopes** into **3 groups**, **Group 1** was of **MHC-II epitope**, only **1 result** we had obtained and we observed amino acids such as **THR, TYR, ASN, PHE, VAL, GLY** Again, for **Group 2** was of **MHC-I epitope**, we had obtained **4 results** and common amino acid residues we observed were **SER, GLY, ASP, PRO, ARG, TYR, GLU**. Lastly, **Group 3** was of **B-cell epitopes**, **2 results** we had obtained and we observed common amino acids such as **SER, GLY, ASP, PRO, ARG, TYR, GLU, LEU, HIS**. Among the **4 cluster**-based results we had obtained for **MHC-I** alleles, we observed the binding affinity-based result has the lowest binding energy of **-11.2**, for the MHC-I epitope spanning residues 122 to 130. Other three clusters also provided satisfactory results. Again, for **MHC-II** we had obtained only **1 result**, which was for MHC-II epitope spanning residues 89 to 99. The **binding energy** for MHC-II was **-10.2**, which was also in acceptable parameters. As

we know the lower the binding affinity, the stronger the binding energy would be. Thus we were able to determine which epitopes would best bind with MHC-I and MHC-II molecules respectively. Among the 7 filtered epitopes (**Table 3.13**), 4 MHC I and 1 MHC II epitopes were predicted. These 5 MHC epitopes were then used in NetMHC-4.0 and NetMHCII-2.3 for predicting the common HLA epitopes that bind with the identified MHC epitopes. A set of weak and strong MHC binders were identified (**Table 3.15**) which were used in population coverage analysis (**Table 3.16**). The population coverage analysis of the MHC epitopes and HLA alleles revealed that substantial populations throughout the world had the corresponding alleles and epitopes inside their genome. Among the African region, the prevalence of the selected alleles and the MHC epitopes were higher in North Africa and lowest in the East African population (**Table 3.16**).

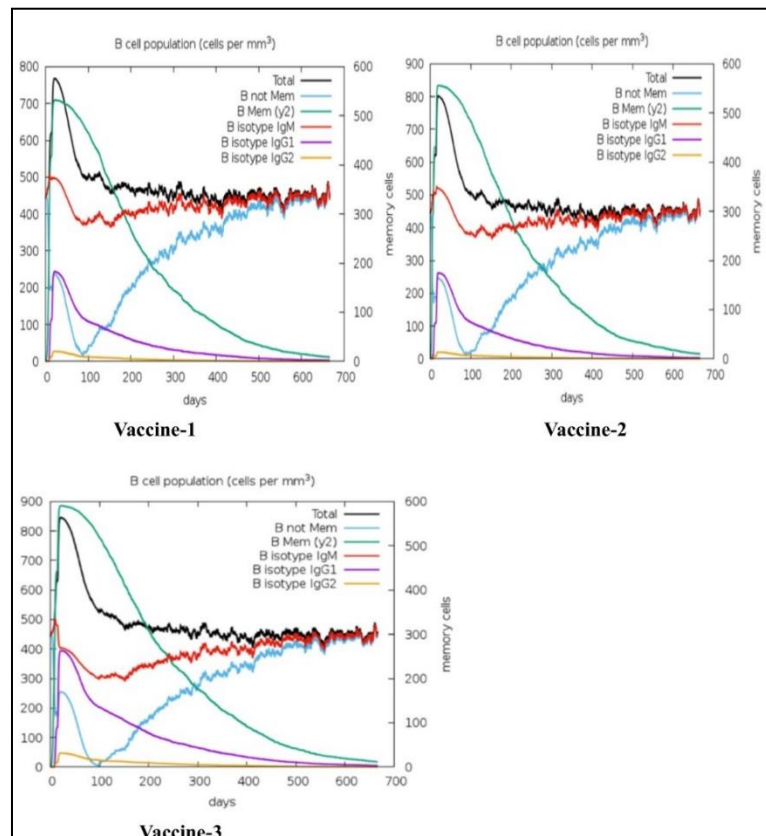
We had constructed 3 vaccines, the main and only structural difference among these vaccines were **3 different adjuvants**, **Vaccine 1** was made of **L7/L12 ribosomal protein**, **Vaccine 2** was made of **HBHA protein** and lastly, **Vaccine 3** was made of **beta-defensin-3** adjuvant. Furthermore, all other components were kept similar, such as the **junctional molecules** or **linkers** were kept similar and we had used **EAAAK**, **GGGGS**, **GPGPG**, and **KK** linkers. **EAAAK** was used as the **head linker molecule** for all three-vaccine construction. Again, **PADRE sequence** was also used for 3 vaccines. Finally, we used **4 CTL** (as **MHC-1**) molecules, such as **CTL-1**, **CTL-2**, **CTL-3**, **CTL-4** molecules, **1 HTL** (as **MHC-2**) molecule and **2 BCL** (as **B cell epitopes**) molecules such as **BCL-1** and **BCL-2**. A multi-epitope based vaccine is constructed based on these components because at the time of immunogenic based reaction, the vaccine would stimulate the Helper T cell, Cytotoxic-T cell and B-cell response as these cells play a role in our immune system[210]. B cell initiate antibody production and also assist in the function of memory cells[211]. Again, in the vaccine construction, a combination of T-cell and B-cell epitopes of the proteins were predicted, so that the vaccines would be able to provoke potential immune responses. In the next step (**Table 3.24**), the antigenicity, allergenicity and physicochemical properties i. e., number of amino acids, Molecular weight, Theoretical pI, Total number of negatively charged residues, Total number of positively charged residues, Formula, Total number of atoms, Extinction coefficients, estimated half-life, Instability index, Aliphatic index and Grand average of hydropathicity (GRAVY) of the three constructed vaccines were determined. All the proteins were found to be antigenic and non-allergen. In

physicochemical properties analysis, the pH at which a protein should not have a net charge is described by the theoretical pI. The amount of light absorbed by a substance at a specific wavelength is shown by its extinction coefficient [212, 213]. A compound's instability index indicates the likelihood that it will be stable, and a compound with an instability index of 40 or above is thought to be unstable [214]. The percentage volume of the amino acids in a protein's side chains occupied by the aliphatic amino acids, such as alanine, valine, etc., is known as the protein's aliphatic index [215]. A protein's GRAVY value is calculated by adding the hydrophathy values of each of its amino acids, then dividing the result by the total number of residues in the protein sequence. The hydrophilic and hydrophobic characteristics of a substance are represented by the negative and positive GRAVY values, respectively [216, 217]. Vaccine 3 has been found as basic with the theoretical PI of more than 9 which is 9.5. And all of the three vaccines have a good instability index which is less than 40. Additionally, the aliphatic index is a measure of a protein's thermal stability; the greater the aliphatic index, the more thermostable the protein. All of the vaccine designs were thought to be quite thermostable because all of them were predicted to have high aliphatic indexes. Furthermore, the vaccine designs' negative GRAVY values indicated that they may all have hydrophilic properties. Our constructed Vaccine 1 consisted of 308 amino acids, Vaccine 2 was of 337 amino acids and Vaccine 3 consisted of 241 amino acid length. In order to observe our vaccine in 2D format we used two tools. At first all 3 vaccines were observed by the PSIPRED tool, we studied the data, and observed the **Strand**, **Helix** and **Coil** region for all 3 vaccines. Again, another tool named JPRED4 was also used to predict the **Strand**, **Helix** and **Coil** region for our vaccine candidates. Then we conducted a compare and contrast analysis based on the results we had obtained from both of the tool, our aim was to observe that, whether the prediction from both PSIPRED and JPRED4 would match or not, we used another tool called 2dSS for our task, observation revealed that, in **most cases** and for all 3 of the vaccines, the PSIPRED and JPRED4 data provided us with similar results. We also generated 3d structures for our vaccines, the results were generated from I-TASSER tool, for Vaccine 1 we had observed multiple peptide residues such as GLU, ALA, LYS, MET, LEU, SER, THR, ASP, PHE, VAL, PRO, ILE, ARG, TYR, HIS, GLY, ASN, PHE, GLN(Glutamine), CYS, TRP. Again, for Vaccine 2 we observed, GLU, ALA, LYS, MET, ASN, ILE, PRO, ASP, LEU, GLY, THR, VAL, ARG, THR, PHE, GLN, SER, TRP, HIS. Finally, for Vaccine 3, we saw GLU, ALA, LYS, GLY, ILE, ASN,

THR, LEU, GLN, TYR, CYS, ARG, VAL, SER, PRO, ILE, HIS, PHE, TRP. In order to obtain a clear visual for our vaccine with a **repacking of their side chain**, we used **Galaxy Refine** tool, this tool predicted **5 models** for each of the 3 vaccines that we had constructed thus a total of 15 models were obtained, we used PDBsum to study the best model for each of the vaccine, we observed Ramachandran plot model for 15 models, while studying the Ramachandran plot, we considered parameters such as, **Most Favored region(%)** and **Disallowed region(%)** value and compared it . Again, we also studied G-Factors for 15 models as well, parameters such as, **Phi-psi distribution score, Parameter Average Score, Main-Chain bond angle, Main-Chain covalent forces Average Score** and **OVERALL AVERAGE** were our main target for comparing areas among the vaccines. Another tool named SAVES was also used to generate our best vaccine candidates. Overall analysis from SAVES and PDBsum combined data revealed that, for **Vaccine 1**, best candidate was **Model 4**, for **Vaccine 2**, best candidate was **Model 2**, lastly for **Vaccine 3**, best candidate was **Model 3**. The 3D structure of protein files of the three vaccines were used for vaccine protein disulphide engineering. Introducing disulphide bonds into proteins involves improving the proteins stability, modification of functional characteristics, protein dynamics analysis etc. The disulphide modified vaccines then undergone into antigenicity, allergenicity and physicochemical properties analysis like before.

All of the modified vaccines were found to antigen and non-allergen in VaxiJen 2.0 and AllerTOP v.2.0 server (**Table 3.30**). In addition, from physicochemical proteins analysis, Vaccine 3 has been found as basic as it has 9.42 theoretical PI value. Therefore, these vaccines were found to be stable and they had similar estimated half-life. We had obtained 3D structures for our vaccine and we used them to conduct the docking procedure, the docking procedure itself has its own significance, the binding affinity value would provide us with a clear picture for using best vaccine candidate from three of the vaccine models that we had designed, furthermore, the docking would give us a clear picture regarding how the vaccines would interact with different MHC allele [217]. As MHC allele candidates, we had chosen 1QEW: Chain A - HLA-A 0201 and 2G9H: Chain B - HLA-DRA1.1QEW for MHC-I and **2G9H** for MHC-2. After analysing the lowest binding affinity scores for our vaccines, we had merged them into a dataset and we had observed the outcome. Based on the scores provided by PRODIGY, for **MHC-I(1QEW)**, we saw **Vaccine 3** held the lowest binding energy affinity which was **-16.5**, at the same time scores for Vaccine 1 and Vaccine 2 had good

outcomes as well for MHC-I. Similarly, in case of **MHC-2(2G9H)**, we saw **Vaccine 1** held the lowest binding energy affinity which was **-15.8**, at the same time Vaccine 2 and Vaccine 3 had good outcomes as well. For providing better humoral immunity the three vaccine constructs should have effective conformational B cell epitopes. ElliPro server was used by keeping the default parameter to predict the B lymphocytic epitopes. And this tool predicted continuous and discontinuous epitopes from the three vaccines (**Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, Table 3.32, Table 3.33, Table 3.34, Table 3.35, Table 3.36, Table 3.37**).



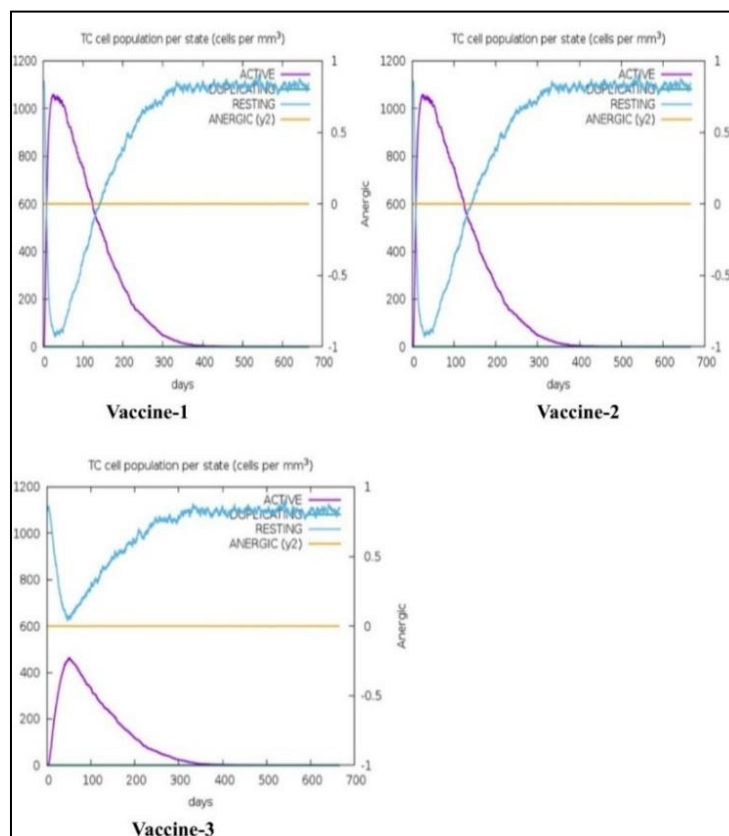
**Figure 4.1: Vaccine 1, Vaccine 2, Vaccine 3 B-cell population comparison (cells per mm<sup>3</sup>)**

Vaccine 1 (**Figure 4.1**), Based on **B cell population** graph, we observed that, **B Mem(y2)**, numerical range was a **little over 700 cells per mm<sup>3</sup>**, within day 0 to 100 and gradually decreasing over time (day range 600-700). Then, **B isotype IgM**, numerical range was between **400-500 cells per mm<sup>3</sup>**, while reaching the peak value a little over **500 cells per mm<sup>3</sup>** from **0-100 days** range. Furthermore, **B isotype IgG1** our graph showed us, from day 0 to 100 range the peak value was **between 200-300 cells per mm<sup>3</sup>**. Finally, **B isotype, IgG2**, showed a little numerical value between 0-100 cells per mm<sup>3</sup> from 0-100 days range then gradually decreasing over the 600-700 days range.



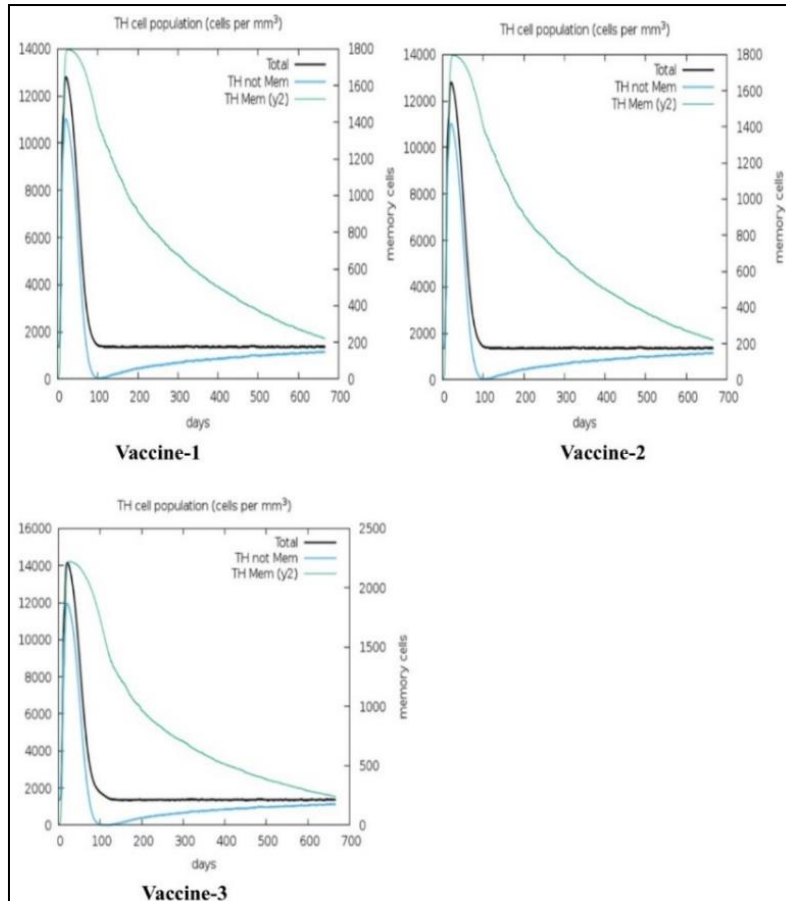
**Vaccine 2(Figure 4.1).** Based on **B cell population** graph, we observed that, **B Mem(y2)**, numerical range was a **little over 800 cells per mm<sup>3</sup>**, within day 0 to 100 and gradually decreasing over time (day range 600-700). Then, **B isotype IgM**, numerical range was **between 400-500 cells per mm<sup>3</sup>**, while reaching the peak value a **little over 500 cells per mm<sup>3</sup>** from **0-100 days** range. Furthermore, **B isotype IgG1**, our graph showed us, from day 0 to 100 range, the **peak value** was **between 200-300 cells per mm<sup>3</sup>**. Finally, **B isotype IgG2**, showed a **little numerical value** between 0-100 **cells per mm<sup>3</sup>** from 0-100 days range then gradually decreasing over the 600-700 days range.

**Vaccine 3(Figure 4.1),** Based on **B cell population** graph, we observed that, **B Mem(y2)**, numerical range was **between 800-900 cells per mm<sup>3</sup>**, within day 0 to 100 and gradually decreasing over time (day range 600-700). Then, **B isotype IgM**, numerical range was **between 400-500 cells per mm<sup>3</sup>**, while reaching the peak value a **little over 500 cells per mm<sup>3</sup>** from **0-100 days** range. Furthermore, **B isotype IgG1**, our graph showed us, from day 0 to 100 range, the **peak value** was **almost 400 cells per mm<sup>3</sup>**. Finally, **B isotype IgG2**, showed a **little numerical value** between 0-100 **cells per mm<sup>3</sup>** from 0-100 days range then gradually decreasing over the 600-700 days range.



**Figure 4.2: Vaccine 1, Vaccine 2, Vaccine 3 TC population per state comparison**

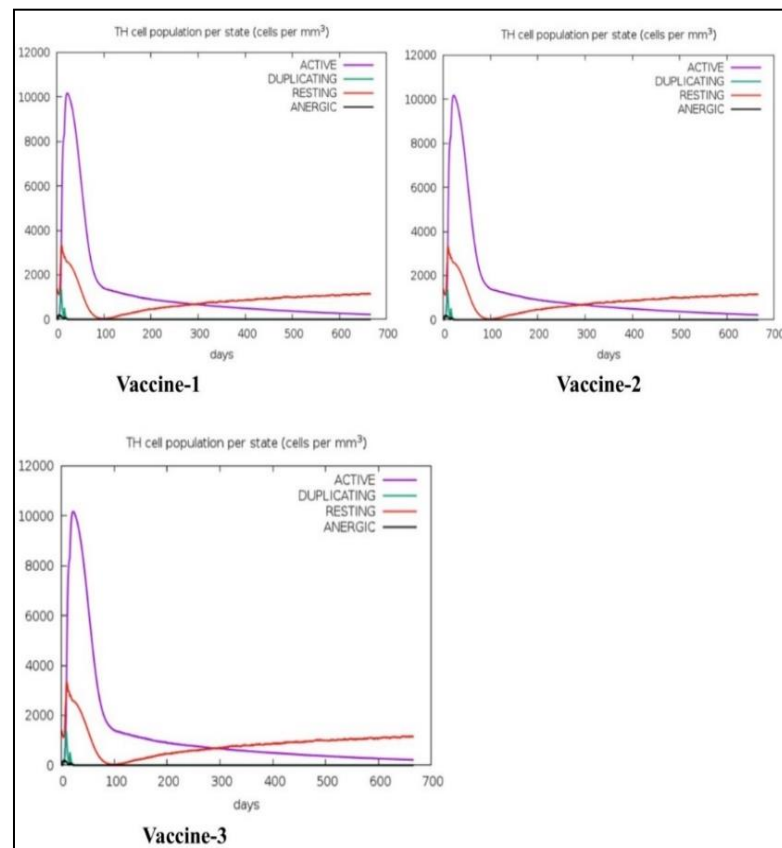
Based on the **TC cell population per state (Figure 4.2)** graph, in case of **Vaccine 1(Figure 4.2)**, we observed that there had been a **peak of ACTIVE Cyto-toxic T cell** number a **little over 1000 cells per mm<sup>3</sup> between 0-100 day** range, then the numerical ratio had **decreased** over the time, again the **RESTING Cyto-toxic T cell** number were **between 1000-1200 cells per mm<sup>3</sup>** at first, then **drastically the ratio started to drop**, the **lowest numerical ratio** was **between 0-200 cells per mm<sup>3</sup> between 0-100 day** range, later gradually increasing over time, Finally we again observed an **Equilibrium state** of the graph where the **Active state Cytotoxic T cell** had been **equal to Resting state of Cytotoxic T cell**, from the graph, the **intercept point** was, **between day 100-200** and **Active and Resting T cell ratio** was a **little below 600 cells per mm<sup>3</sup>**. Based on the **TC cell population per state** graph, in case of **Vaccine 2(Figure 4.2)**, we observed that there had been a **peak of ACTIVE Cyto-toxic T cell** number was **almost equal to 1000 cells per mm<sup>3</sup> between 0-100 day** range, then the numerical ratio had **decreased** over the time, again the **RESTING Cyto-toxic T cell** number were **between 1000-1200 cells per mm<sup>3</sup>** at first, then **drastically the ratio started to drop**, the **lowest numerical ratio** was **between 0-200 cells per mm<sup>3</sup> between 0-100 day** range, later gradually increasing over time, Finally, we again observed an **Equilibrium state** of the graph where the **Active state Cytotoxic T cell** had been **equal to Resting state of Cytotoxic T cell**, from the graph, the **intercept point** was, **between day 100-200** and **Active and Resting T cell ratio** was a **little below 600 cells per mm<sup>3</sup>**. In the case of vaccine 3(Figure 4.2), we had observed a very **unique result**, which is, **ACTIVE Cyto-toxic T cell** number, reached the highest peak at a **little over 400 cells per mm<sup>3</sup>, between day 0-100 range**. Again, for **Resting Cyto-toxic T cells**, we observed from a value of **1000-1200 cells per mm<sup>3</sup>**, the **ratio dropped** and then **gradually increased** over time. Most importantly, another noticeable thing was, like Vaccine 1 and Vaccine 2 model, we had **obtained an equilibrium state** from their graphs, from Vaccine 3, we **did not see any equilibrium state (Figure 4.2)**, thus we analyzed that, in case of vaccine 3, **no Resting Cytotoxic T cell would become active**.



**Figure 4.3: Vaccine 1, Vaccine 2, Vaccine 3 TH (Helper T) cell population comparison**

Based on the **TH cell population graph** we observed (**Figure 4.3**), in case of **Vaccine 1 (Figure 4.3)**, numerical value of **Total TH cell population** was held at a **peak between 12000-14000 cells per mm<sup>3</sup>**, between **day 0-100**, then decreased gradually and maintained a stable value, from **Day 100 to between 600-700 days**. Again, **TH Not Mem**, held a peak of numerical value **between 10000-12000 cells per mm<sup>3</sup>**. Finally, **TH Mem(y2)** had a peak at **14000 cells per mm<sup>3</sup>** based on the graph then had a steady decreased ratio over time (day 600-700). In the case of **Vaccine 2(Figure 4.3)**, the numerical value of **Total TH cell population** was a **little over 12000 cells per mm<sup>3</sup>**, between **day 0-100**, then decreased gradually and maintained a stable value, from **Day 100 to between 600-700 days**. Again, **TH Not Mem**, held a peak of numerical value **of a little over 10000 cells per mm<sup>3</sup>**. Finally, **TH Mem(y2)** had a peak **between 12000-14000 cells per mm<sup>3</sup>** based on the graph then had a steady decreased ratio over time (day 600-700) Lastly, for **Vaccine 3(Figure 4.3)**, the numerical value of **Total TH cell population** was a **little over 14000 cells per mm<sup>3</sup>**, between **day 0-100**, then decreased gradually and maintained a **stable value**, from **Day 100 to between 600-700 days**.

Again, **TH Not Mem**, held a peak of numerical value of **12000 cells per mm<sup>3</sup>**. Finally, **TH Mem(y2)** had a peak at **14000 cells per mm<sup>3</sup>** based on the graph then had a steady decreased ratio over time (day 600-700).



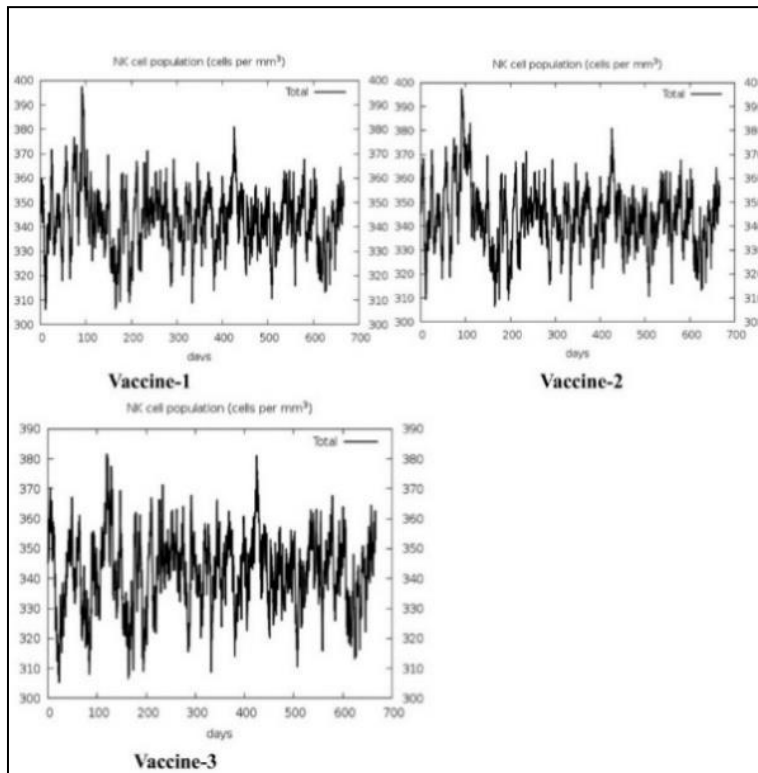
**Figure 4.4: Vaccine 1, Vaccine 2, Vaccine 3 TH cell population per state comparison**

Based on the **TH cell population per state** graph (Figure 4.4), in case of **Vaccine 1** (Figure 4.4), we observed that there had been a **peak** of **ACTIVE** Helper T cell number a **little over 10000 cells per mm<sup>3</sup>**, between **0-100 day** range, then the numeral ratio had **decreased** over the time, again the **RESTING(not active)** Helper T cell number were at **between 0-2000 cells per mm<sup>3</sup>** at first then, reached a **peak of 2000-4000 cells per mm<sup>3</sup>**, between **0-100 day** range, Finally we again observed an **Equilibrium state** of the graph, where the **Active state** Helper T cell had been **equal** to **Resting state(not active)** of Helper T cell, from the graph, the **intercept point** was, **between day 200-300** and **Active and Resting T cell ratio** was **between 0-2000 cells per mm<sup>3</sup>**.

Based on the **TH cell population per state** graph, in case of **Vaccine 2** (Figure 4.4), we observed that there had been a **peak** of **ACTIVE** Helper T cell number **reaching 10000 cells per mm<sup>3</sup>**, between **0–100-day** range, then the numeral ratio had **decreased**

over the time, again the **RESTING (not active) Helper T cell** number obtained a **peak a little over 3000 cells per mm<sup>3</sup>, between 0-100 day range**, Finally we again observed an **Equilibrium state** of the graph, where the **Active state Helper T cell** had been equal to **Resting state(not active) of Helper T cell**, from the graph, the **intercept point** was, **between day 200-300** and **Active and Resting Helper T cell ratio** was **between 0-1000 cells per mm<sup>3</sup>**.

Lastly, in case of **Vaccine 3 (Figure 4.4)**, we observed that there had been a **peak of ACTIVE Helper T cell** number were **between 10000-30000 cells per mm<sup>3</sup>, between 0-100 day range**, then the numeral ratio had **decreased** over the time, again the **RESTING(not active) Helper T cell** number obtained a **peak between 2000-4000 cells per mm<sup>3</sup>, between 0-100 day range**, Finally we again observed an **Equilibrium state** of the graph, where the **Active state Helper T cell** had been equal to **Resting state(not active) of Helper T cell**, from the graph, the **intercept point** was, **between a little over day 300** and **Active and Resting Helper T cell ratio** was **between 0-2000 cells per mm<sup>3</sup>**.



**Figure 4.5: Vaccine 1, Vaccine 2, Vaccine 3 NK-cell population comparison**

We also observed the **NK cell population** graph (**Figure 4.5**), where in case of **Vaccine 1 (Figure 4.5)**, we observed **2 peak points** of NK cell population. In case of **peak point 1**, the NK cells were **between 390-400 cells per mm<sup>3</sup>, between 0-100 days**. Again, for

**peak point 2**, observed NK cells were, a little over 380 **cells per mm<sup>3</sup>**, **between 400-500** days. In the case of **Vaccine 2 (Figure 4.5)**, we also observed **2 peak** points of NK cell population from our graph. In case of **peak point 1**, the NK cells numerical value was between **390-400 cells per mm<sup>3</sup>**, **between 0-100** days. Again, for **peak point 2**, observed NK cells were, **a little over 380 cells per mm<sup>3</sup>**, **between 400-500** days. Lastly, **Vaccine 3 (Figure 4.5)**, we observed **2 peak** points of the NK cell population. In case of **peak point 1**, the NK cells were **a little over 380 cells per mm<sup>3</sup>**, **between 100-200** days. Again, for **peak point 2**, observed NK cells were, a little over 380 **cells per mm<sup>3</sup>**, **between 400-500** days.

## Chapter 5. Conclusion

A bioinformatics approach has employed in our study to develop epitope based vaccines against the phylogroup II of bat *Lyssavirus*, targeting the surface protein Glycoprotein (G). Previously many lab-based methodology and computational technology has been used in constructing the vaccines of various *Lyssavirus* strains. In our study, we used different bioinformatics tools for constructing our desired vaccines against *Lyssavirus* phylogroup II i.e., Lagos bat virus (LBV), Mokola virus (MOKV) and Shimoni bat virus (SHBV). All available Glycoprotein sequences from each species were compared and conserved regions of these sequences were analysed for determining their properties. Therefore, by targeting these Glycoprotein sequences we concluded three desired vaccine models against Phylogroup II *Lyssavirus* strains.

Here, in our study the epitope-based vaccines depicted high antigenicity and non-allergenicity. Besides, these vaccines are expected to be highly stable and generate both humoral and cell mediated immune responses in body as they contained multiple B cell and MHC epitopes against Phylogroup II strains. Population coverage analysis indicated that it can cover a huge portion of African population and it is an important analysis since *Lyssavirus* phylogroup II has more influence in the African region.

By analysing all of the result of bioinformatics tools, we can come to a conclusion that all these three vaccines should be effective and immunogenic. However, since all the analysis and prediction were done by computational procedures, it is not sufficient enough to confirm the final outcome of our study. Hence, we conclude that our constructed vaccines require further *in vivo* and *in vitro* assessment as well as clinical trials to make it effectively usable for mankind. We hope that our study will be a great contribution to the bioinformatics research purpose.

## References

1. Zandi, F., et al., *Rabies Infection: An Overview of Lyssavirus-Host Protein Interactions*. Iran Biomed J, 2021.**25**(4): p.226-42.
2. Walker, P.J., et al., *ICTV Virus Taxonomy Profile: Rhabdoviridae 2022*. J Gen Virol, 2022.**103**(6).
3. Lafon, M., *Rabies virus receptors*. J Neurovirol, 2005.**11**(1): p.82-7.
4. Charlton, K.M., et al., *The long incubation period in rabies: delayed progression of infection in muscle at the site of exposure*. Acta Neuropathol, 1997.**94**(1): p.73-7.
5. Shankar, V., B. Dietzschold, and H. Koprowski, *Direct entry of rabies virus into the central nervous system without prior local replication*. J Virol, 1991.**65**(5): p.2736-8.
6. Lycke, E. and H. Tsiang, *Rabies virus infection of cultured rat sensory neurons*. J Virol, 1987.**61**(9): p.2733-41.
7. Tsiang, H., P.E. Ceccaldi, and E. Lycke, *Rabies virus infection and transport in human sensory dorsal root ganglia neurons*. J Gen Virol, 1991.**72** ( Pt 5): p.1191-4.
8. Ugolini, G., *Rabies virus as a transneuronal tracer of neuronal connections*. Adv Virus Res, 2011.**79**: p.165-202.
9. Gluska, S., et al., *Rabies Virus Hijacks and accelerates the p75NTR retrograde axonal transport machinery*. PLoS Pathog, 2014.**10**(8): p. e1004348.
10. Bauer, A., et al., *Anterograde glycoprotein-dependent transport of newly generated rabies virus in dorsal root ganglion neurons*. J Virol, 2014.**88**(24): p.14172-83.
11. Bauer, A., et al., *A Dynein Light Chain 1 Binding Motif in Rabies Virus Polymerase L Protein Plays a Role in Microtubule Reorganization and Viral Primary Transcription*. J Virol, 2015.**89**(18): p.9591-600.
12. Wang, J., et al., *Metabotropic glutamate receptor subtype 2 is a cellular receptor for rabies virus*. PLoS Pathog, 2018.**14**(7): p. e1007189.
13. Charlton, K.M., G.A. Casey, and J.B. Campbell, *Experimental rabies in skunks: mechanisms of infection of the salivary glands*. Can J Comp Med, 1983.**47**(3): p.363-9.
14. Haage, V., et al., *Comprehensive gene expression meta-analysis identifies signature genes that distinguish microglia from peripheral monocytes/macrophages in health and glioma*. Acta Neuropathol Commun, 2019.**7**(1): p.20.
15. Udow, S.J., R.A. Marrie, and A.C. Jackson, *Clinical features of dog- and bat-acquired rabies in humans*. Clin Infect Dis, 2013.**57**(5): p.689-96.
16. Jackson, A.C., *Rabies and Other Rhabdovirus Infections*, in *Harrison's Principles of Internal Medicine, 19e*, D. Kasper, et al., Editors.2014, McGraw-Hill Education: New York, NY.
17. Jackson, A.C., *Human Rabies: a 2016 Update*. Curr Infect Dis Rep, 2016.**18**(11): p.38.
18. Li, X.Q., L. Sarmiento, and Z.F. Fu, *Degeneration of neuronal processes after infection with pathogenic, but not attenuated, rabies viruses*. J Virol, 2005.**79**(15): p.10063-8.
19. Fernandes, E.R., et al., *In situ apoptosis of adaptive immune cells and the cellular escape of rabies virus in CNS from patients with human rabies transmitted by *Desmodus rotundus**. Virus Res, 2011.**156**(1-2): p.121-6.



20. Fu, Z.F. and A.C. Jackson, *Neuronal dysfunction and death in rabies virus infection*. Journal of NeuroVirology, 2005.11(1): p.101-106.
21. Scott, C.A., et al., *Structural abnormalities in neurons are sufficient to explain the clinical disease and fatal outcome of experimental rabies in yellow fluorescent protein-expressing transgenic mice*. J Virol, 2008.82(1): p.513-21.
22. Guigoni, C. and P. Coulon, *Rabies virus is not cytolytic for rat spinal motoneurons in vitro*. Journal of NeuroVirology, 2002.8(4): p.306-317.
23. Gholami, A., et al., *Mitochondrial dysfunction in Lyssavirus-induced apoptosis*. J Virol, 2008.82(10): p.4774-84.
24. Hooper, D.C., et al., *Local nitric oxide production in viral and autoimmune diseases of the central nervous system*.1995.92(12): p.5312-5316.
25. Koprowski, H., et al., *In vivo expression of inducible nitric oxide synthase in experimentally induced neurologic diseases*. Proc Natl Acad Sci U S A, 1993.90(7): p.3024-7.
26. Hemachudha, T., J. Laothamatas, and C.E. Rupprecht, *Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges*. Lancet Neurol, 2002.1(2): p.101-9.
27. World Health, O., *WHO expert consultation on rabies: third report*. WHO technical report series;1012.2018, Geneva: World Health Organization.
28. Boland, T.A., et al., *Phylogenetic and epidemiologic evidence of multiyear incubation in human rabies*. Ann Neurol, 2014.75(1): p.155-60.
29. Horton, D.L., et al., *Antigenic and genetic characterization of a divergent African virus, Ikoma Lyssavirus*. J Gen Virol, 2014.95(Pt 5): p.1025-1032.
30. Aréchiga Ceballos, N., et al., *Novel Lyssavirus in bat, Spain*. Emerg Infect Dis, 2013.19(5): p.793-5.
31. Hu, S.C., et al., *Lyssavirus in Japanese Pipistrelle, Taiwan*. Emerg Infect Dis, 2018.24(4): p.782-785.
32. Gunawardena, P., et al., *Lyssavirus in Indian Flying Foxes, Sri Lanka*. Emerging Infectious Disease journal, 2016.22(8): p.1456.
33. Gnanadurai, C.W., et al., *Differential Host Immune Responses after Infection with Wild-Type or Lab-Attenuated Rabies Viruses in Dogs*. PLoS Negl Trop Dis, 2015.9(8): p. e0004023.
34. Faul, E.J., et al., *Rabies virus infection induces type I interferon production in an IPS-1 dependent manner while dendritic cell activation relies on IFNAR signaling*. PLoS Pathog, 2010.6(7): p. e1001016.
35. Walker, P.J., et al., *Evolution of genome size and complexity in the rhabdoviridae*. PLoS Pathog, 2015.11(2): p. e1004664.
36. Albertini, A.A., R.W. Ruigrok, and D. Blondel, *Rabies virus transcription and replication*. Adv Virus Res, 2011.79: p.1-22.
37. Nadin-Davis, S.A. and C. Fehlner-Gardiner, *Lyssaviruses: current trends*. Adv Virus Res, 2008.71: p.207-50.
38. Jackson, A.C., *Chapter 1 - History of Rabies Research*, in *Rabies (Third Edition)*, A.C. Jackson, Editor.2013, Academic Press: Boston. p.1-15.
39. Wunner, W.H. and K.-K. Conzelmann, *Chapter 2 - Rabies Virus*, in *Rabies (Third Edition)*, A.C. Jackson, Editor.2013, Academic Press: Boston. p.17-60.
40. Schnell, M.J., et al., *The cell biology of rabies virus: using stealth to reach the brain*. Nat Rev Microbiol, 2010.8(1): p.51-61.
41. Lentz, T.L., et al., *Is the acetylcholine receptor a rabies virus receptor?* Science, 1982.215(4529): p.182-4.

42. Thoulouze, M.I., et al., *The neural cell adhesion molecule is a receptor for rabies virus*. J Virol, 1998.72(9): p.7181-90.
43. Tuffereau, C., et al., *Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus*. Embo j, 1998.17(24): p.7250-9.
44. Tuffereau, C., et al., *Interaction of Lyssaviruses with the low-affinity nerve-growth factor receptor p75NTR*. J Gen Virol, 2001.82(Pt 12): p.2861-2867.
45. Ito, N., et al., *Rescue of rabies virus from cloned cDNA and identification of the pathogenicity-related gene: glycoprotein gene is associated with virulence for adult mice*. J Virol, 2001.75(19): p.9121-8.
46. Kucera, P., et al., *Pathways of the early propagation of virulent and avirulent rabies strains from the eye to the brain*. 1985.55(1): p.158-162.
47. Yan, X., et al., *The rabies virus glycoprotein determines the distribution of different rabies virus strains in the brain*. J Neurovirol, 2002.8(4): p.345-52.
48. Dougherty, K.D. and T.A. Milner, *p75NTR immunoreactivity in the rat dentate gyrus is mostly within presynaptic profiles but is also found in some astrocytic and postsynaptic profiles*. J Comp Neurol, 1999.407(1): p.77-91.
49. Caillet-Saguy, C., et al., *Strategies to interfere with PDZ-mediated interactions in neurons: What we can learn from the rabies virus*. Prog Biophys Mol Biol, 2015.119(1): p.53-9.
50. Azimzadeh Jamalkandi, S., et al., *Systems Biomedicine of Rabies Delineates the Affected Signaling Pathways*. Front Microbiol, 2016.7: p.1688.
51. Yin, K., et al., *SNAP25 regulates the release of the Rabies virus in nerve cells via SNARE complex-mediated membrane fusion*. Veterinary Microbiology, 2020.245: p.108699.
52. Masatani, T., et al., *Rabies virus nucleoprotein functions to evade activation of the RIG-I-mediated antiviral response*. J Virol, 2010.84(8): p.4002-12.
53. Lahaye, X., et al., *Hsp70 protein positively regulates rabies virus infection*. J Virol, 2012.86(9): p.4743-51.
54. Zhang, J., et al., *Cellular chaperonin CCT $\gamma$  contributes to rabies virus replication during infection*. J Virol, 2013.87(13): p.7608-21.
55. Fooks, A.R. and A.C. Jackson, *Rabies: scientific basis of the disease and its management*. 2020: Academic Press.
56. Morin, B., et al., *An In Vitro RNA Synthesis Assay for Rabies Virus Defines Ribonucleoprotein Interactions Critical for Polymerase Activity*. J Virol, 2017.91(1).
57. Lieu, K.G., et al., *The rabies virus interferon antagonist P protein interacts with activated STAT3 and inhibits Gp130 receptor signaling*. J Virol, 2013.87(14): p.8261-5.
58. Li, Y., et al., *Rabies virus phosphoprotein interacts with ribosomal protein L9 and affects rabies virus replication*. 2016.488: p.216-224.
59. Liu, J., et al., *Rabies virus phosphoprotein P5 binding to BECN1 regulates self-replication by BECN1-mediated autophagy signaling pathway*. Cell Commun Signal, 2020.18(1): p.153.
60. Liu, J., et al., *BECN1-dependent CASP2 incomplete autophagy induction by binding to rabies virus phosphoprotein*. Autophagy, 2017.13(4): p.739-753.
61. Komarova, A.V., et al., *Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis*. Nucleic Acids Res, 2007.35(5): p.1522-32.
62. Sonthonnax, F., et al., *Lyssavirus matrix protein cooperates with phosphoprotein to modulate the Jak-Stat pathway*. Sci Rep, 2019.9(1): p.12171.

63. Luco, S., et al., *RelAp43, a member of the NF- $\kappa$ B family involved in innate immune response against Lyssavirus infection*. PLoS Pathog, 2012.8(12): p. e1003060.
64. Badrane, H., et al., *Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity*. J Virol, 2001.75(7): p.3268-76.
65. Voloch, C.M.A.U.C.R.T.A.U.M.B.S.C.G.T.I.A.o.A.E.i.L.G.R.P.D. Viruses, 2014.6, 4465-4478 DOI: 10.3390/v6114465.
66. Rabies-Bulletin-Europe. *Classification*. Available from: <https://www.who-rabies-bulletin.org/site-page/classification>.
67. Fooks, A.R., et al., *Current status of rabies and prospects for elimination*. Lancet, 2014.384(9951): p.1389-99.
68. Kuzmin, I.V., et al., *Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions*. Virus Research, 2008.136(1): p.81-90.
69. Weir, D.L.A.U.C.S.A.A.A.U.V.B.K.A.U.C.J.Y.L.S.I.L.A.U., V. Characterization of Cross-Reactive Human Monoclonal Antibodies That Potently Neutralize Australian Bat *Lyssavirus*, and L. Other Phylogroup Viruses, 2021.13, DOI: 10.3390/v13030391.
70. Servat, A.W.M.C.F.C.-P.o.I.R.V.f.V.U.a.B.L.O.i.E. Viruses, 2019.11, DOI: 10.3390/v11100936.
71. Gluska, S., et al., *Rabies Virus Hijacks and Accelerates the p75NTR Retrograde Axonal Transport Machinery*. PLOS Pathogens, 2014.10(8): p. e1004348.
72. Schnell, M.J., et al., *The cell biology of rabies virus: using stealth to reach the brain*. Nature Reviews Microbiology, 2010.8(1): p.51-61.
73. *Host Defense Evasion Mechanisms of Rabies Virus*. Available from: [https://microbewiki.kenyon.edu/index.php/Host\\_Defense\\_Evasion\\_Mechanisms\\_of\\_Rabies\\_Virus](https://microbewiki.kenyon.edu/index.php/Host_Defense_Evasion_Mechanisms_of_Rabies_Virus).
74. Fooks, A.R., A.C. Banyard, and H.C.J. Ertl, *New human rabies vaccines in the pipeline*. Vaccine, 2019.37 Suppl 1(Suppl 1): p. A140-a145.
75. Hicks, D.J., A.R. Fooks, and N. Johnson, *Developments in rabies vaccines*. Clin Exp Immunol, 2012.169(3): p.199-204.
76. Zhu, S. and C. Guo, *Rabies Control and Treatment: From Prophylaxis to Strategies with Curative Potential*. Viruses, 2016.8(11).
77. Ertl, H.C.J., *New Rabies Vaccines for Use in Humans*. Vaccines (Basel), 2019.7(2).
78. Wiktor, T.J., et al., *Protection from rabies by a vaccinia virus recombinant containing the rabies virus glycoprotein gene*. Proc Natl Acad Sci U S A, 1984.81(22): p.7194-8.
79. Cadoz, M., et al., *Immunisation with canarypox virus expressing rabies glycoprotein*. Lancet, 1992.339(8807): p.1429-32.
80. Jallet, C., et al., *Chimeric Lyssavirus glycoproteins with increased immunological potential*. J Virol, 1999.73(1): p.225-33.
81. Lodmell, D.L. and L.C. Ewalt, *Post-exposure DNA vaccination protects mice against rabies virus*. Vaccine, 2001.19(17-19): p.2468-73.
82. Conzelmann, K.K. and M. Schnell, *Rescue of synthetic genomic RNA analogs of rabies virus by plasmid-encoded proteins*. J Virol, 1994.68(2): p.713-9.
83. Chen, Z., et al., *A novel rabies vaccine based on a recombinant parainfluenza virus 5 expressing rabies virus glycoprotein*. J Virol, 2013.87(6): p.2986-93.

84. Ge, J., et al., *Newcastle disease virus-vectored rabies vaccine is safe, highly immunogenic, and provides long-lasting protection in dogs and cats.* J Virol, 2011.85(16): p.8241-52.
85. Ge, J., et al., *Newcastle disease virus-based live attenuated vaccine completely protects chickens and mice from lethal challenge of homologous and heterologous H5N1 avian influenza viruses.* J Virol, 2007.81(1): p.150-8.
86. Martinez-Sobrido, L., et al., *Protection against respiratory syncytial virus by a recombinant Newcastle disease virus vector.* J Virol, 2006.80(3): p.1130-9.
87. Nakaya, T., et al., *Recombinant Newcastle disease virus as a vaccine vector.* J Virol, 2001.75(23): p.11868-73.
88. Amann, R., et al., *A new rabies vaccine based on a recombinant ORF virus (parapoxvirus) expressing the rabies virus glycoprotein.* J Virol, 2013.87(3): p.1618-30.
89. Weyer, J., C.E. Rupprecht, and L.H. Nel, *Poxvirus-vectored vaccines for rabies-a review.* Vaccine, 2009.27(51): p.7198-201.
90. Blancou, J., et al., *Oral vaccination of the fox against rabies using a live recombinant vaccinia virus.* Nature, 1986.322(6077): p.373-5.
91. Rupprecht, C.E., et al., *Oral immunization and protection of raccoons (Procyon lotor) with a vaccinia-rabies glycoprotein recombinant virus vaccine.* Proc Natl Acad Sci U S A, 1986.83(20): p.7947-50.
92. Rupprecht, C.E., et al., *Human infection due to recombinant vaccinia-rabies glycoprotein virus.* N Engl J Med, 2001.345(8): p.582-6.
93. Wu, Q., et al., *Rabies-virus-glycoprotein-pseudotyped recombinant baculovirus vaccine confers complete protection against lethal rabies virus challenge in a mouse model.* Vet Microbiol, 2014.171(1-2): p.93-101.
94. Houimel, M. and K. Dellagi, *Peptide mimotopes of rabies virus glycoprotein with immunogenic activity.* Vaccine, 2009.27(34): p.4648-55.
95. Castel, G., et al., *Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity.* J Virol, 2009.83(20): p.10808-20.
96. Liu, M.A., *DNA vaccines: an historical perspective and view to the future.* Immunol Rev, 2011.239(1): p.62-84.
97. Ferraro, B., et al., *Clinical applications of DNA vaccines: current progress.* Clin Infect Dis, 2011.53(3): p.296-302.
98. Li, L., F. Saade, and N. Petrovsky, *The future of human DNA vaccines.* J Biotechnol, 2012.162(2-3): p.171-82.
99. Ulmer, J.B., et al., *RNA-based vaccines.* Vaccine, 2012.30(30): p.4414-8.
100. Ertl, H.C., *Novel vaccines to human rabies.* PLoS Negl Trop Dis, 2009.3(9): p. e515.
101. Xiang, Z.Q., et al., *Vaccination with a plasmid vector carrying the rabies virus glycoprotein gene induces protective immunity against rabies virus.* Virology, 1994.199(1): p.132-40.
102. Perrin, P., et al., *Immunization of dogs with a DNA vaccine induces protection against rabies virus.* Vaccine, 1999.18(5-6): p.479-86.
103. Saxena, S., et al., *A sindbis virus replicon-based DNA vaccine encoding the rabies virus glycoprotein elicits immune responses and complete protection in mice from lethal challenge.* Vaccine, 2008.26(51): p.6592-601.
104. Lodmell, D.L., et al., *DNA vaccination of mice against rabies virus: effects of the route of vaccination and the adjuvant monophosphoryl lipid A (MPL).* Vaccine, 2000.18(11-12): p.1059-66.

105. Margalith, M. and A. Vilalta, *Sustained protective rabies neutralizing antibody titers after administration of cationic lipid-formulated pDNA vaccine*. *Genet Vaccines Ther*, 2006.4: p.2.
106. Deering, R.P., et al., *Nucleic acid vaccines: prospects for non-viral delivery of mRNA vaccines*. *Expert Opin Drug Deliv*, 2014.11(6): p.885-99.
107. Zamore, P.D., et al., *RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals*. *Cell*, 2000.101(1): p.25-33.
108. Ketzinel-Gilad, M., Y. Shaul, and E. Galun, *RNA interference for antiviral therapy*. *J Gene Med*, 2006.8(8): p.933-50.
109. Kumar, P., et al., *Transvascular delivery of small interfering RNA to the central nervous system*. *Nature*, 2007.448(7149): p.39-43.
110. Kaur, M., et al., *Rabies vaccines: where do we stand, where are we heading?* *Expert Rev Vaccines*, 2015.14(3): p.369-81.
111. Huang, C.T., et al., *Enhancement of blood-brain barrier permeability is required for intravenously administered virus neutralizing antibodies to clear an established rabies virus infection from the brain and prevent the development of rabies in mice*. *Antiviral Res*, 2014.110: p.132-41.
112. Chen, S., et al., *Bispecific antibodies in cancer immunotherapy*. *Hum Vaccin Immunother*, 2016.12(10): p.2491-2500.
113. Pardridge, W.M., *Blood-brain barrier drug delivery of IgG fusion proteins with a transferrin receptor monoclonal antibody*. *Expert Opin Drug Deliv*, 2015.12(2): p.207-22.
114. Boado, R.J., et al., *Humanization of anti-human insulin receptor antibody for drug targeting across the human blood-brain barrier*. *Biotechnol Bioeng*, 2007.96(2): p.381-91.
115. Zuchero, Y.J., et al., *Discovery of Novel Blood-Brain Barrier Targets to Enhance Brain Uptake of Therapeutic Antibodies*. *Neuron*, 2016.89(1): p.70-82.
116. Malerczyk, C., et al., *Duration of immunity: an anamnestic response 14 years after rabies vaccination with purified chick embryo cell rabies vaccine*. *J Travel Med*, 2007.14(1): p.63-4.
117. Coffman, R.L., A. Sher, and R.A. Seder, *Vaccine adjuvants: putting innate immunity to work*. *Immunity*, 2010.33(4): p.492-503.
118. Dowling, J.K. and A. Mansell, *Toll-like receptors: the swiss army knife of immunity and vaccine development*. *Clin Transl Immunology*, 2016.5(5): p. e85.
119. Luo, T.R., et al., *Antigenic and functional analyses of glycoprotein of rabies virus using monoclonal antibodies*. *Microbiol Immunol*, 1998.42(3): p.187-93.
120. Schnell, M.J., T. Mebatsion, and K.K. Conzelmann, *Infectious rabies viruses from cloned cDNA*. *Embo j*, 1994.13(18): p.4195-203.
121. Ito, N., et al., *Characterization of M gene-deficient rabies virus with advantages of effective immunization and safety as a vaccine strain*. *Microbiol Immunol*, 2005.49(11): p.971-9.
122. Alberer, M., et al., *Safety and immunogenicity of a mRNA rabies vaccine in healthy adults: an open-label, non-randomised, prospective, first-in-human phase I clinical trial*. *Lancet*, 2017.390(10101): p.1511-1520.
123. Hensley, S.E., et al., *Dendritic cell maturation, but not CD8+ T cell induction, is dependent on type I IFN signaling during vaccination with adenovirus vectors*. *J Immunol*, 2005.175(9): p.6032-41.

124. Maki, J., et al., *Oral vaccination of wildlife using a vaccinia-rabies-glycoprotein recombinant virus vaccine (RABORAL V-RG®): a global review*. *Vet Res*, 2017.48(1): p.57.
125. *Human vaccinia infection after contact with a raccoon rabies vaccine bait - Pennsylvania, 2009*. *MMWR Morb Mortal Wkly Rep*, 2009.58(43): p.1204-7.
126. Rosatte, R.C., et al., *Aerial distribution of ONRAB baits as a tactic to control rabies in raccoons and striped skunks in Ontario, Canada*. *J Wildl Dis*, 2009.45(2): p.363-74.
127. Xiang, Z., et al., *Novel, chimpanzee serotype 68-based adenoviral vaccine carrier for induction of antibodies to a transgene product*. *J Virol*, 2002.76(6): p.2667-75.
128. Wang, C., et al., *A simian-adenovirus-vectored rabies vaccine suitable for thermostabilisation and clinical development for low-cost single-dose pre-exposure prophylaxis*. *PLoS Negl Trop Dis*, 2018.12(10): p. e0006870.
129. Fontana, D., et al., *Rabies virus-like particles expressed in HEK293 cells*. *Vaccine*, 2014.32(24): p.2799-804.
130. Fernández-Núñez, E.G., et al., *Transient expression of rabies virus G-glycoprotein using BHK-21 cells cultured in suspension*. *Biotechnol Lett*, 2015.37(6): p.1153-63.
131. Wojczyk, B.S., et al., *Purification of a secreted form of recombinant rabies virus glycoprotein: comparison of two affinity tags*. *Protein Expr Purif*, 1996.7(2): p.183-93.
132. Brooks, S.A., *Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system*. *Mol Biotechnol*, 2004.28(3): p.241-55.
133. Yendo, A.C., et al., *A rabies vaccine adjuvanted with saponins from leaves of the soap tree (Quillaja brasiliensis) induces specific immune responses and protects against lethal challenge*. *Vaccine*, 2016.34(20): p.2305-11.
134. Wang, X., et al., *A CpG oligodeoxynucleotide acts as a potent adjuvant for inactivated rabies virus vaccine*. *Vaccine*, 2008.26(15): p.1893-901.
135. Kalimuddin, S., et al., *A phase II randomized study to determine the safety and immunogenicity of the novel PIKA rabies vaccine containing the PIKA adjuvant using an accelerated regimen*. *Vaccine*, 2017.35(51): p.7127-7132.
136. Mebatsion, T., et al., *Highly stable expression of a foreign gene from rabies virus vectors*. *Proc Natl Acad Sci U S A*, 1996.93(14): p.7310-4.
137. Morimoto, K., Y. Shoji, and S. Inoue, *Characterization of P gene-deficient rabies virus: propagation, pathogenicity and antigenicity*. *Virus Res*, 2005.111(1): p.61-7.
138. Cenna, J., et al., *Replication-deficient rabies virus-based vaccines are safe and immunogenic in mice and nonhuman primates*. *J Infect Dis*, 2009.200(8): p.1251-60.
139. Papp, Z., L.A. Babiuk, and M.E. Baca-Estrada, *The effect of pre-existing adenovirus-specific immunity on immune responses induced by recombinant adenovirus expressing glycoprotein D of bovine herpesvirus type 1*. *Vaccine*, 1999.17(7-8): p.933-43.
140. Galletti, R., P. Beauverger, and T.F. Wild, *Passively administered antibody suppresses the induction of measles virus antibodies by vaccinia-measles recombinant viruses*. *Vaccine*, 1995.13(2): p.197-201.
141. Hammer, S.M., et al., *Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine*. *N Engl J Med*, 2013.369(22): p.2083-92.

142. Casimiro, D.R., et al., *Comparative immunogenicity in rhesus monkeys of DNA plasmid, recombinant vaccinia virus, and replication-defective adenovirus vectors expressing a human immunodeficiency virus type 1 gag gene*. J Virol, 2003.77(11): p.6305-13.
143. Tatsis, N., et al., *Adenoviral vectors persist in vivo and maintain activated CD8+ T cells: implications for their use as vaccines*. Blood, 2007.110(6): p.1916-23.
144. Xiang, Z.Q., et al., *Protection of non-human primates against rabies with an adenovirus recombinant vaccine*. Virology, 2014.450-451: p.243-9.
145. Fooks, A.R., *Development of oral vaccines for human use*. Curr Opin Mol Ther, 2000.2(1): p.80-6.
146. Bourhy, H., B. Kissi, and N. Tordo, *Molecular Diversity of the Lyssavirus Genus*. Virology, 1993.194(1): p.70-81.
147. Fooks, A.R.A.U.S.R.M.W.T.N.F.C.M.A.U.M.T.M.L.M.A.U.B.A.C. *Viruses*, 2021.13, DOI: 10.3390/v13091769.
148. UniProtKB.
149. M-Coffee.
150. MEGA X.
151. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data*. Nucleic Acids Res, 2020.48(W1): p. W449-w454.
152. NetMHCpan-4.1.
153. IDEB recommended 2.22.
154. Jespersen, M.C., et al., *BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes*. Nucleic Acids Res, 2017.45(W1): p. W24-w29.
155. Bepipred Linear Epitope Prediction 2.0.
156. VaxiJen 2.0
157. AllerTOP v.2.0.
158. ToxinPred
159. blastp
160. Lamiable, A., et al., *PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex*. Nucleic Acids Res, 2016.44(W1): p. W449-54.
161. PEP-FOLD3.
162. Sehnal, D., et al., *Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures*. Nucleic Acids Research, 2021.49(W1): p. W431-W437.
163. Web3dMol.
164. Honorato, R.V., et al., *Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem*. Front Mol Biosci, 2021.8: p.729513.
165. HADDOCK.
166. Vangone, A. and A.M. Bonvin, *Contacts-based prediction of binding affinity in protein-protein complexes*. Elife, 2015.4: p. e07454.
167. Xue, L.C., et al., *PRODIGY: a web server for predicting the binding affinity of protein-protein complexes*. Bioinformatics, 2016.32(23): p.3676-3678.
168. PRODIGY.
169. NetMHC - 4.0
170. NetMHCII 2.3

171. Berger, A., *HLA typing*. *Bmj*, 2001.322(7280): p.218.
172. *Population coverage*
173. Misra, N., et al., *Population coverage analysis of T-Cell epitopes of Neisseria meningitidis serogroup B from Iron acquisition proteins for vaccine design*. *Bioinformatics*, 2011.6(7): p.255-61.
174. Sarkar, B., et al., *Immunoinformatics-guided designing and in silico analysis of epitope-based polyvalent vaccines against multiple strains of human coronavirus (HCoV)*. *Expert Rev Vaccines*, 2022.21(12): p.1851-1871.
175. *ProtParam server*.
176. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server*. *Bioinformatics*, 2000.16(4): p.404-405.
177. *PSIPRED*.
178. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server*. *Nucleic Acids Res*, 2015.43(W1): p. W389-94.
179. Lotun, D.P., et al., *2dSS: a web server for protein secondary structure visualization*. 2019: p.649426.
180. *2dSS*.
181. Zheng, W., et al., *Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations*. *Cell Rep Methods*, 2021.1(3).
182. *I-TASSER*.
183. Heo, L., H. Park, and C. Seok, *GalaxyRefine: Protein structure refinement driven by side-chain repacking*. *Nucleic Acids Res*, 2013.41(Web Server issue): p. W384-8.
184. *Galaxy-Refine*.
185. Laskowski, R.A., et al., *PDBsum: Structural summaries of PDB entries*. *Protein Sci*, 2018.27(1): p.129-134.
186. *PDBsum*.
187. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*. *Protein Sci*, 1993.2(9): p.1511-9.
188. Bowie, J.U., R. Lüthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure*. *Science*, 1991.253(5016): p.164-70.
189. Lüthy, R., J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles*. *Nature*, 1992.356(6364): p.83-5.
190. Pontius, J., J. Richelle, and S.J. Wodak, *Deviations from standard atomic volumes as a quality measure for protein crystal structures*. *J Mol Biol*, 1996.264(1): p.121-36.
191. *UCLA*.
192. *Disulfide by Design server*.
193. Dombkowski, A.A., *Disulfide by Design: a computational method for the rational design of disulfide bonds in proteins*. *Bioinformatics*, 2003.19(14): p.1852-3.
194. Craig, D.B. and A.A. Dombkowski, *Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins*. *BMC Bioinformatics*, 2013.14: p.346.
195. Sarkar, B., et al., *Immunoinformatics-guided designing and in silico analysis of epitope-based polyvalent vaccines against multiple strains of human coronavirus (HCoV)*. *Expert Review of Vaccines*, 2022.21(12): p.1851-1871.
196. *UCSF chimera Software*.
197. *RCSB PDB - Mol\* 3D Viewer*



198. Yurina, V. and O.R. Adianingsih, *Predicting epitopes for vaccine development using bioinformatics tools*. Ther Adv Vaccines Immunother, 2022.10: p.25151355221100218.
199. *ElliPro*
200. Rapin, N., et al., *Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system*. PLoS One, 2010.5(4): p. e9862.
201. *C-ImmSim*
202. Hajighahramani, N., et al., *Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus*. Infect Genet Evol, 2017.48: p.83-94.
203. Pandey, R.K., S. Sundar, and V.K. Prajapati, *Differential Expression of miRNA Regulates T Cell Differentiation and Plasticity During Visceral Leishmaniasis Infection*. Front Microbiol, 2016.7: p.206.
204. Alexander, J., et al., *Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses*. J Immunol, 2000.164(3): p.1625-33.
205. Wu, C.-Y., et al., *Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation*. Journal of Biomedical Science, 2010.17(1): p.88.
206. Magnan, C.N., A. Randall, and P. Baldi, *SOLpro: accurate sequence-based prediction of protein solubility*. Bioinformatics, 2009.25(17): p.2200-7.
207. Magnan, C.N., et al., *High-throughput prediction of protein antigenicity using protein microarray data*. Bioinformatics, 2010.26(23): p.2936-43.
208. Saha, S. and G.P. Raghava, *AlgPred: prediction of allergenic proteins and mapping of IgE epitopes*. Nucleic Acids Res, 2006.34(Web Server issue): p. W202-9.
209. Nandy, A. and S.C. Basak, *Bioinformatics in Design of Antiviral Vaccines*. Encyclopedia of Biomedical Engineering.2019:280-90. doi: 10.1016/B978-0-12-801238-3.10878-5. Epub 2018 Sep 13.
210. Shey, R.A., et al., *In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases*. Scientific Reports, 2019.9(1): p.4409.
211. Chauhan, V., et al., *Excavating SARS-coronavirus 2 genome for epitope-based subunit vaccine synthesis using immunoinformatics approach*. J Cell Physiol, 2021.236(2): p.1131-1147.
212. Pei, H., et al., *Expression of SARS-coronavirus nucleocapsid protein in Escherichia coli and Lactococcus lactis for serodiagnosis and mucosal vaccination*. Appl Microbiol Biotechnol, 2005.68(2): p.220-7.
213. Morla, S., A. Makhija, and S. Kumar, *Synonymous codon usage pattern in glycoprotein gene of rabies virus*. Gene, 2016.584(1): p.1-6.
214. Hamasaki-Katagiri, N., et al., *The importance of mRNA structure in determining the pathogenicity of synonymous and non-synonymous mutations in haemophilia*. Haemophilia, 2017.23(1): p. e8-e17.
215. Yazdani, Z., et al., *Design an Efficient Multi-Epitope Peptide Vaccine Candidate Against SARS-CoV-2: An in silico Analysis*. Infect Drug Resist, 2020.13: p.3007-3022.
216. Joshi, A., et al., *Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach*. Informatics in Medicine Unlocked, 2020.19: p.100338.

217. Lizbeth, R.G., et al., *Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine*. J Biomol Struct Dyn, 2021.39(13): p.4878-4892.