

***In-silico* Approach to Design a Novel Multi-epitope Vaccine Against
Zika Virus**

By

Radiya Tum Mardiya

20136003

Umaiya Binte Mahabub

20136009

A thesis submitted to the Department of Mathematical and Natural Sciences in partial fulfillment
of the requirements for the degree of Bachelor of Science in Biotechnology

Department of Mathematical and Natural Sciences

BRAC University

April 2024

© 2024. BRAC University

All rights reserved.

Declaration

It is hereby declared that.

1. The thesis submitted is our own original work while completing a degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate reference.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Umaiya Binte Mahabub

20136009

Radiya Tum Mardiya

20136003

Approval

The thesis” *In-silico* approach to Design a Novel Multi-epitope Vaccine Against Zika Virus” was submitted by.

1. Radiya Tum Mardiya (20136006)
2. Umaiya Binte Mahabub (20136009)

2020 has been accepted as satisfactory in partial fulfillment of the requirement for the Bachelor of Science in Biotechnology degree on 25th April 2024.

Examining Committee:

Supervisor:

Tushar Ahmed Shishir
Lecturer, Biotechnology Program
Department of Mathematics and Natural Sciences
BRAC University, Dhaka, Bangladesh

Program Coordinator:

Dr. Munima Haque
Associate Professor, Department of Mathematics and
Natural Sciences
BRAC University, Dhaka, Bangladesh

Departmental Head:

DR. A F M Yusuf Haider
Professor and Chairperson, Department of
Mathematics and Natural Sciences
BRAC University, Dhaka, Bangladesh

Ethics Statement

We, Radiya tum Mardiya and Umaiya Binte Mahabub, hereby certify that the following criteria are fulfilled in the manuscript “*In-silico* approach to Design a Novel Multi-epitope Vaccine Against Zika Virus.”

- 1) This content is our original work and has not been previously published.
- 2) All sources used are appropriately acknowledged through accurate citations and justified references.

Abstract

Zika virus is a global concern and a public health emergency proclaimed by WHO due to its dangerous repercussions after the infection. Since it poses a significant threat to emerging again, competent and adequate action must be taken to develop an efficient vaccine. The object of this study was to construct a multi-peptide vaccine against the deadly ZIKA virus by using a computational approach of immunoinformatic and bioinformatics. Three structure proteins (capsid, envelope, membrane) were selected due to their precious target in the host cell. The helper and cytotoxic T cells (CTL, HTL) and B cells were predicted as epitopes and fused with the propitious linkers and adjuvants molecules as promising vaccine constructs. These ZIKV vaccine constructs were checked thoroughly to inspect the highest immunogenicity and physiological properties. All the primary constructs were non-allergenic, non-toxic, and highly antigenic. The validation and selection criteria helped to choose the stronger candidates with considerable features to induce a high immunogenic response. The docking and structural simulation of the vaccine constructs with TLR4 receptor confirmed the binding stability and affinities to a great extent. MD simulations of the vaccine-TLR4 complex were performed to appraise the efficacy of the structural stability and integrity to find the only suitable candidate among all the primary and secondary selections. However, the in-vivo and in-vitro testing and clinical trial are required to justify the aim of the study.

Keywords: ZIKA, immunoinformatic, multi-epitope, genome, vaccine, MD simulation.

Dedication

The thesis is dedicated to the amazing individuals who helped us grow and learn throughout our academic journey. We came this far because of their good wishes and prayers from them.

To our beloved family, parents, and siblings who supported us, without you, we would never have come this far. We appreciate every moment and everything that you have done for us.

To our amazing supervisor and mentor, his guidance, dedication, and patience made us who we are today. Our faculties and confidantes made our four-year journey a wonderful experience and helped us shape our path.

Acknowledgment

We would like to express profound gratitude to our honorable thesis supervisor, Tushar Ahmed Shishir, Lecturer, Biotechnology program, Department of Mathematics and Natural Sciences, BRAC University, for allowing us to collaborate with him and believing in us. His dedication, tireless effort, patience, and unwavering support allowed us to continue the thesis journey successfully. His priceless encouragement and assistance pushed us forward to work on a bioinformatics thesis, overcoming all the obstacles. Throughout the journey, he inspired us to explore and learn the latest ideas, which helped us complete our thesis successfully. We are always grateful to him for his unswerving support and consolation.

We want to show our deepest gratitude to our co-supervisor, Afia Nowshin, a former lecturer in the Biotechnology program at the Department of Mathematics and Natural Sciences, BRAC University. She inspired us to continue the research, and her continuous encouragement profoundly helped us start an in-silico project.

We want to show our utmost gratitude to our program director, Dr. Munima Haque, Associate Professor and Course Coordinator of the Biotechnology Program, Department of Mathematics and Natural Science, BRAC University, and our honorable Professor and Chairperson, DR. A F M Yusuf Haidar, Department of Mathematics and Natural Science, BRAC University for permitting us to continue our thesis which helped us to assess our scientific aptitude towards research.

We are hopeful and grateful for all the guidance and assistance. We aspire to meet your expectations and goals in the future with our utmost endeavor. This thesis is our first research work and one of the most important parts of our career and undergraduate life. We will try to accomplish and continue to work on the valuable experience we gained.

Radiya Tum Mardiya, Umaiya Binte Mahabub.

March 2024

Table	of	Contents
Declaration.....		ii
Approval		iii
Abstract.....		v
Dedication		vi
Acknowledgment.....		vii
Table of Contents		viii
List of Tables		xi
List of Figures.....		xii
List of Acronyms		xvi
Chapter – 1		17
INTRODUCTION.....		18
1.1) Introduction of ZIKV.....		18
1.2) Zika Virus Throughout History		20
1.3) Zika virus, it is Classification and Replication Cycle Based on Genetic Evidence		21
1.4) Viral Structure of Zika Virus.....		23
1.5) Vectors and Host Range of Zika Virus:		26
1.6) Mode of Transmission.....		27
1.7) Symptoms and Complications by Zika Virus:.....		28
1.8) Complications of Infection.....		29
1.9) Aims and Objectives:		31
Chapter – 2		32
2. Materials and Method		32
2.1) Retrieval of Genome sequence and analysis		34
2.2) Physicochemical properties of protein sequences.....		34
2.3) Prediction of CTL, HTL, and B cell epitope.....		35
2.4) Population coverage analysis of the epitopes		35
2.5) Multiepitope Vaccine Construction		36
2.6) Structure prediction, and validation		37
2.7) Molecular Docking		37

2.8) Immune Simulation.....	38
2.9) Structural Simulation	38
2.10) Molecular Dynamic Simulation	39
Chapter – 3	40
3) RESULTS	40
3.1) Genome Sequence Retrieval.....	41
3.1.1) Phylogenetic tree	41
3.1.2) Analysis of the retrieved Complete Genome sequences of Zika Virus.....	42
3.1.3) Extraction of protein sequences	43
3.2) Epitope prediction	45
3.2.1) Cytotoxic T-Lymphocyte and Helper T-Lymphocyte Epitopes prediction	45
3.2.2) B cell Epitopes prediction	53
3.2.3) Selected Epitopes	55
3.2.4) Three-dimensional structure of selected epitopes.	57
3.3) Result Population Coverage	62
3.3.1) The MHC-I tables' statistics.....	63
3.3.2) MHC-II tables' statistics.....	67
3.4) Vaccine Formulation.....	70
3.4.1) Primary construction of 19 vaccines.....	70
3.4.2) physiological properties of 19 selected vaccines.	71
3.4.3) Secondary Structure Prediction of Vaccine Constructs	72
3.5) Selected vaccine analysis.....	92
3.5.1) Validation, ERRAT, and Ramachandran favored and disallowed region	92
3.5.2) Ramachandran Plot analysis.....	92
3.6) Molecular Docking	95
3.6.1) Target receptor and vaccine binding energy	97
3.6.2) Surface, Hydrogen, and non-bound Interaction analysis during docking.....	99
3.7) Structural Simulation	102
3.7.1) Fluctuation plot analysis.....	102
3.7.2) Contact Map analysis.....	106
3.8) Immune Simulation.....	109

3.9) Molecular Dynamic Simulation	121
3.9.1) RMS Fluctuation	127
3.9.2) RMSD	127
3.9.3) Hydrogen Bond.....	127
3.9.4) Rg (Radius of Gyration)	128
3.9.5) SASA.....	128
Chapter – 4	130
4) Discussion	130
Chapter – 5	142
5) Conclusion	142
Reference	144

List of Tables

Table 1.1: Functions and Locations of Zika Virus Proteins (Rawal et al., 2016)	25
Table 2. 1: Selected adjuvants and linkers.	36
Table 3.1: Accession number of the retrieved genome sequence, type, and origin.	42
Table 3.2: The start and end positions of the capsid, membrane, and envelope proteins.	43
Table 3.3: Physical and chemical properties of the annotated Protein sequences of capsid, envelope, and membrane	44
Table 3.4: MHC-I of envelope protein.	45
Table 3. 5: MHC-II for envelope protein.	48
Table 3. 6: MHC-I for capsid protein.....	50
Table 3. 7: MHC-II capsid protein.	50
Table 3. 8: MHC-I membrane protein.....	51
Table 3. 9: MHC-II of membrane protein.	52
Table 3. 10: B cell epitope prediction for envelope protein.	53
Table 3. 11: B cell epitope prediction for capsid.	54
Table 3. 12: B cell epitope prediction for membrane.....	55
Table 3. 13: Selected CTL, HTL and B-cell epitopes.....	56
Table 3. 14: 3D structure of the selected epitopes.	57
Table 3. 16: MHC-II population coverage.....	67
Table 3. 17: Physiological properties of 19 selected vaccines.	71
Table 3. 18: Ramachandran plot and ERRAT value of 10 selected vaccines.	92
Table 3. 19: Docking and Binding energy.	98
Table 3. 20: Interactive regions for chain A and B	108
Table 3. 21: Concentration of B memory cell, Helper T memory cell, Ag count, IgM+IgG and lastly, IFN-g count for ten selected vaccines.	121

List of Figures

Figure 1.1: Transmission of ZIKV around the world and major events according to WHO and Lancaster University	20
Figure 1.2: The Baltimore classification system is based on genome type and replication strategy.	22
Figure 1.3: Structural and Functional components of ZIKV and their locations (12) (Zhou et al., 2017)	24
Figure 1.4: Modes of ZIKV transmission between infected and uninfected persons through mosquito bite and mode of TBVs in preventing and controlling ZIKV infection.	27
Figure 1.5: How ZIKV enters the human population (Garcez et al., 2017)	27
Figure 1.6: Mode of transmission of Zika virus infection (Sager et al., 2018).....	28
Figure 1. 7: Symptoms of Zika Virus.	29
Figure 2.1: Methodology	33
Figure 3.1: Phylogenetic tree of the Zika variants	41
Figure 3.2: MHC-I population coverage-1	64
Figure 3. 3: MHC-I population coverage- 2.....	65
Figure 3. 4: MHC-II population coverage-1	68
Figure 3. 5: MHC-II population coverage-2	69
Figure 3. 6: Structure of vaccine 1. (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	73
Figure 3. 7: Structure of vaccine 2 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	74
Figure 3. 8: Structure of vaccine 3 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	75
Figure 3. 9: Structure of vaccine 4 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	76
Figure 3.10: Structure of vaccine 5 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	77
Figure 3. 11: Structure of vaccine 6 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	78

Figure 3. 12: Structure of vaccine 7 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	79
Figure 3. 13: Structure of vaccine 8 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	80
Figure 3. 14: Structure of vaccine 9 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	81
Figure 3. 15: Structure of vaccine 10 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	82
Figure 3. 16: Structure of vaccine 11 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	83
Figure 3. 17: Structure of vaccine 12 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	84
Figure 3. 18: Structure of vaccine 13 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	85
Figure 3. 19: Structure of vaccine 14 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	86
Figure 3. 20: Structure of vaccine 15 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	87
Figure 3. 21: Structure of vaccine 16 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	88
Figure 3. 22: Structure of vaccine 17 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	89
Figure 3. 23: Structure of vaccine 18 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	90
Figure 3. 24: Structure of vaccine 19 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.....	91
Figure 3. 25: Ramachandran Plot.....	93
Figure 3. 26: Ramachandran Plot.....	94
Figure 3. 27: Molecular Docking of Vaccine 16,18,12,11,9 with Receptor TLR4	96
Figure 3. 28: Molecular Docking of Vaccine 8,6,5,4, 2 with Receptor TLR4	97

Figure 3. 29: Vaccine5, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction	99
Figure 3. 30: Vaccine 12, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction	100
Figure 3. 31: Vaccine 16, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction	100
Figure 3. 32: Vaccine 18, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction	101
Figure 3. 33: Vaccine 11, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction	101
Figure 3. 34: Fluctuation Plot of vaccine-receptor complex.....	103
Figure 3. 35: Fluctuation Plot of vaccine-receptor complex.....	104
Figure 3. 36: Residue Contact Map of vaccine 16,18,12,11,9.	106
Figure 3. 37: Contact Map of vaccine 8,6,5,4,2.....	107
Figure 3. 38: Vaccine 2 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	110
Figure 3. 39: Vaccine 4 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	111
Figure 3. 40: Vaccine 5 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	112
Figure 3. 41: Vaccine 6 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	113
Figure 3. 42: Vaccine 8 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	114

Figure 3. 43: Vaccine 10 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	115
Figure 3. 44: Vaccine 11 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	116
Figure 3. 45: Vaccine 12 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	117
Figure 3. 46: Vaccine 16 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	119
Figure 3. 47: Vaccine 18 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins	120
Figure 3. 48: MD Simulation vaccine-11 A) RMSF B) RMSD C) H-Bond D) RG E) SAS.....	122
Figure 3. 49: MD Simulation vaccine-5 A) RMSF B) RMSD C) H-Bond D) RG E) SAS	123
Figure 3. 50: MD Simulation vaccine-12 A) RMSF B) RMSD C) H-Bond D) RG E) SAS.....	124
Figure 3. 51: MD Simulation vaccine-16 A) RMSF B) RMSD C) H-Bond D) RG E) SAS.....	125
Figure 3. 52: MD Simulation vaccine-18 A) RMSF B) RMSD C) H-Bond D) RG E) SAS.....	126

List of Acronyms

Sl	Short form	Abbreviation
1	ZIKV	Zika Virus
2	CHIKV	Chikungunya Virus
3	WNV	West Nile Virus
4	DENV	Dengue Virus
5	YFV	Yellow Fever Virus
6	HIV	Human Immunodeficiency Virus
7	CTL	Cytotoxic T lymphocyte
8	HTL	Helper T lymphocyte
9	MHC	Major Histocompatibility Complex
10	WHO	World Health Organization
11	RMSD	Root Mean Square Deviation
12	RMSF	Root Mean Square Fluctuation
13	RG	Radius of Gyration
14	SASA	Solvent Accessible Surface Area
15	FDA	Food and Drug Administration
16	NCBI	National Centre for Biotechnology Information
17	IEDB	Immune Epitope Database and Analysis Resources
18	TLR	Toll-like receptors

Chapter – 1

INTRODUCTION

1.1) Introduction of ZIKV

A virus is somewhat of an infectious particle without fundamental biological activity. A virus can infect many people despite its small dimensions and lack of complexity compared to bacteria and cells. The capsids of these infectious particles are little more than a protein encasing their genetic instructions. Upon encountering a live host cell, they change their latent state remarkably. The tiny intruders take control of the cell's machinery and make it replicate themselves. Disruption of the host by this viral takeover frequently results in illness. While some viruses are simple, others hide behind lipids that they borrowed. But their goal is one and only: to proliferate and remain alive by preying on the very life they jeopardize. The point is that even the tiniest intruders may cause a major stir, and these little nightmares serve as a reminder that life can take many shapes and sizes (Louten, 2016).

The genetic material of most viruses is a little fragment of single- or double-stranded DNA or RNA. A capsid is a protein encasement that encases nucleic acid. A virion is a completely infectious viral particle. An envelope encases the whole viral particle in certain viruses (Y. Wang et al., 2022).

The Zika virus, which belongs to the Flaviviridae family and, more specifically, the Flavivirus genus, is referred to by the acronym ZIKV. After being taken from a monkey in 1947, it was discovered for the first time in 1948 that mosquitoes spread across Africa (Sharma & Lal, 2017). *A. Aegyptie* CHIKV, WNV, DENV, YFV, HIV, Ebola, and other members of the same family are near relatives of ZIKV. There are two distinct strains of the Zika virus: one from Asia and one from Africa. Eighty percent of people infected with the Zika virus do not show any symptoms at all (Bullard-Feibelman et al., 2017). Concerns have been reignited because of recent rare outbreaks and growing occurrences in particular locations. This comes after the virus gained importance in the Americas and Pacific despite sporadic human infections over the course of fifty years. In 2015, an epidemic of Zika virus (ZIKV) caused severe microcephaly in infants in Brazil, drawing global attention to the disease. Rapid global dissemination allowed this pandemic to engulf the Americas and numerous Pacific and Southeast Asian islands. Reports of a 19% increase in Guillain Barre syndrome (GBS) and a 20-fold increase in parental microcephaly in 2015 compared to the previous year caused a situation of alarm (Wiwanitkit, 2016). The Zika virus was proclaimed a public health emergency by the World Health Organization (WHO) in February 2016 because of its link to

congenital deformities. There will be a sustained effort to combat the Zika virus, even if the "Emergency" designation was withdrawn in November 2016 (Ruivinho & Gama-Carvalho, 2023). So far, vaccines against YFV, TBEV, JEV, and DNEV have been produced and are effective. There is currently no ZIKA virus vaccine that has reached the clinical trial stage. There is currently no vaccination available to protect against the ZIKA virus. Sofosbuvir is the sole medicine the FDA has approved for inhibiting Zika virus infection in a non-human model. However, the challenge posed by ZIKV today is too great for this medicine to manage. Comparing the vaccine against the Zika virus with those of related species may be helpful in creating a vaccine against the Zika virus.

Acetaminophen and ibuprofen are two types of medications that can help manage symptoms like fever and pain. Unfortunately, there is no specific cure for the Zika virus infection. It is important to focus on prevention to prevent the spread of the epidemic. The best way to do this is by taking measures to avoid getting bitten by mosquitoes. To protect ourselves from mosquitoes, it is important to use insect repellents that are registered with the EPA. We should also wear long sleeves and pants to cover our skin, and avoid areas where mosquitoes are likely to breed, such as places with standing water near houses (Dar et al., 2016). However, these approaches do have a few disadvantages. Eliminating all potential breeding places can be challenging, especially in densely populated areas. Additionally, repellents may be uncomfortable as they must be applied frequently. It is not a promising idea to wear certain types of clothing when it is really hot outside (Bullard-Feibelman et al., 2017).

Taking preventative measures is extremely important because pregnant women face a significantly higher risk. You should consider doing more monitoring because there is a risk of microcephaly and other birth problems. The applicable procedures are still the same. However, it is important to note that while condoms can help lower the risk of sexual transmission, they are not perfect and have their limitations. It would be ideal if pregnant women could avoid going to places where there is currently an outbreak of Zika. This brings attention to a basic limitation: complete avoidance may not always be possible, especially for people who live close to sensitive areas. It is important to keep researching Zika because we do not have a vaccine or specific treatment for it (Dikhit et al., 2016).

1.2) Zika Virus Throughout History

In 1947, the Zika virus was first found in monkeys in the Zika forest in Uganda. Since then, not much study has been done on this virus because it was thought to be safe. The first major spread happened almost sixty years after it was first found.

- The first case was found on Yap Island in Micronesia in 2007; 70% of the people who lived there were affected (Nicolson et al., 1981)
- In 2013, it spread to French Polynesia.
- New Caledonia got Zika in 2014.
- The Zika virus was identified in Brazil in May 2015, and it was linked to a sharp rise in microcephaly and GBS.
- In January 2016, the Zika virus had spread to many countries in the Americas, such as the Republic of Cabo Verde, Colombia, Ecuador, El Salvador, French Guinea, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Martinique, Mexico, Panama, Paraguay, Saint Martin, Samoa, Suriname, and Venezuela.

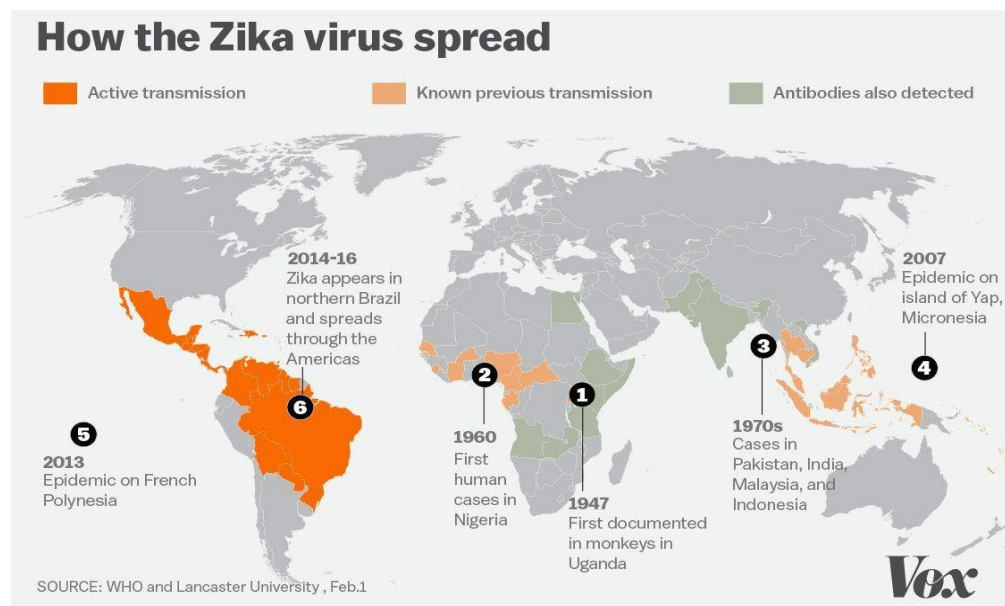


Figure 1.1: Transmission of ZIKV around the world and major events according to WHO and Lancaster University.

The WHO and the CDC say that Zika could spread everywhere. They put countries in various parts of the world into different risk groups. Based on a study reviewed by the WHO, the CDC, and the European CDC as of March 2018, these are the countries:

Asia: Bangladesh, Burma (Myanmar), Cambodia, India, Indonesia, Laos, Malaysia, Maldives, Pakistan, Philippines, Singapore, Thailand, Timor-Leste (the Democratic Republic of Timor-Leste), Vietnam.

The Pacific Islands: Fiji, Marshall Islands, Papua New Guinea, Samoa, Solomon Islands, Tonga.

The Caribbean: Anguilla; Antigua and Barbuda; Aruba; Barbados; Bonaire; British Virgin Islands; Cuba; Curaçao; Dominica; Dominican Republic; Grenada; Haiti; Jamaica; Montserrat; the Commonwealth of Puerto Rico, a US territory; Saba; Saint Kitts and Nevis; Saint Lucia; Saint Martin; Saint Vincent and the Grenadines; Sint Eustatius; Sint Maarten; Trinidad and Tobago; Turks and Caicos Islands; US Virgin Islands.

North America: Mexico.

Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama.

South America: Argentina, Bolivia, Brazil, Colombia, Ecuador, French Guiana, Guyana, Paraguay, Peru, Suriname, Venezuela.

Africa: Angola, Benin, Burkina-Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Congo, Côte d'Ivoire, Democratic Republic of the Congo (Congo-Kinshasa), Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Mali, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South Sudan, Sudan, Tanzania, Togo, Uganda (Nicolson et al., 1981).

Using distance matrix analysis on the viral E protein, NS1, NS2a, NS2b, and NS3, it was found that the Brazilian sample (BR_ZIKA_AB_ES) was at least 12.2% different from several strains from Africa. Once more, the highest rate of difference between the Brazilian isolate and some Asian strains was found to be 1.9%. This makes it noticeably clear that the Asian lineage has turned harmful and is causing major health issues all over the world.

1.3) Zika Virus, It's Classification and Replication Cycle Based on Genetic Evidence

Based on the process in which their mRNA is produced, there are seven primary categories of viruses. The Baltimore Classification is the name of this order. Since the process known as "Transcription" or mRNA synthesis plays a significant role in the way that various viruses live, it is of utmost importance to determine how the virus managed to survive (Borucki et al., 2019). The Baltimore Classification includes all the following categories:

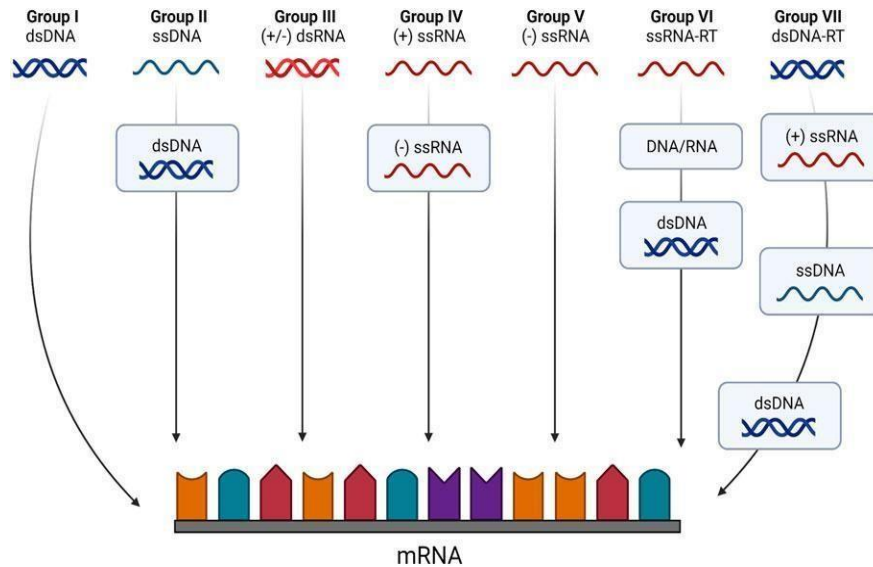


Figure 1.2: The Baltimore classification system is based on genome type and replication strategy.

Despite the wide variety of viruses, their life cycles can be broadly defined using the following terms in sequential order: entry, latency, transcription, replication, and exit (Lazear & Diamond, 2016). Since it is a positive strand RNA virus, Zika follows the regulations of the fourth class. The plus-sense RNA genome is thus identical to the messenger RNA genome. Thus, it can be translated into distinct proteins by means of the cellular machinery once it is uncoated in the cell, which results in the production of a single polyprotein. One such protein is the viral polymerase protein, which is an RNA-dependent RNA polymerase. It produces complementary minus-sense RNA by copying the plus-sense genomic RNA. The process by which viral translation transitions to transcription utilizing the same genomic RNA as a template remains a mystery. Synthesis of new plus-sense strands begins with new minus-sense strands (Nicolson et al., 1981). There are separate ways that viruses can act even when they are in the same stage of their life cycle. There are eight ways to get across the host barrier, eleven ways to copy their DNA, and more than four ways to get out of the cell.

If the exact steps of the Zika virus life cycle are given, they can be summed up as:

- Attachment
- Apoptotic Mimicry
- Viral endocytosis
- Fusion with host endosomal membrane

- Viral factory
- Ds RNA templated transcription and replication
- Cytoplasmic Capsid Assembly
- Viral budding by host ESCRT complexes
- Viral budding by exocytosis

ss (+) RNA virus replication usually occurs in membrane pockets in the host cell's cytoplasm. It has been found that Zika antigens are present in the nuclei of host cells, which suggests that Zika virus replication may be different from that of other flaviviruses.

1.4) Viral Structure of Zika Virus

Its diameter is approximately 60 nanometers, and its shape is spherical. Its positive-strand RNA molecule contains approximately 10,800 nucleotides (Ramacciotti et al., 2019). Three structural proteins (C protein, M protein, E protein) and seven non-structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5) are generated from the RNA polyprotein that is transcribed into a single polyprotein (Shragai et al., 2017). The length of the Zika virus polyprotein is approximately 3400 amino acids.

The image presented herein illustrates the structural configuration of the genetic material and structural proteins of the ZIKA virus. Additionally, the sequential arrangement of the genetic codes is illustrated. The Zika virus generates ten structural and non-structural proteins via transcription of this mRNA.

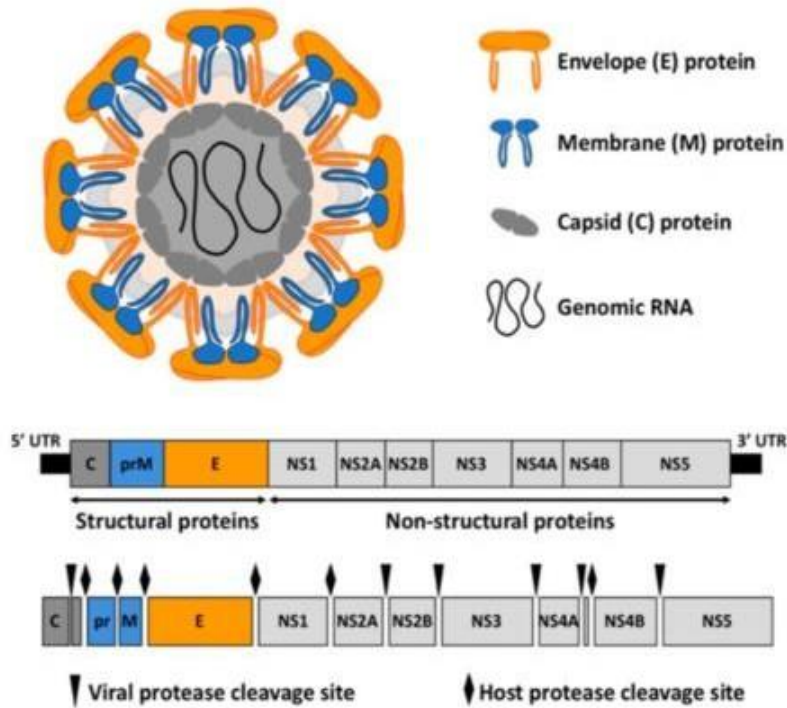


Figure 1.3: Structural and Functional components of ZIKV and their locations (Zhou et al., 2017).

The Zika virus is a 10794 base long positive sense single stranded RNA molecule. It has two non-coding regions on either side, which are called the 5' NCR and the 3' NCR. The Zika virus's open reading frame looks like this: A polyprotein is coded for by 5'-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'. This polyprotein is then cut into capsid (C), precursor membrane (prM), envelope (E), and non-structural proteins (NS). The E protein makes up most of the surface of the virion and helps with replication by sticking to host cells and fusing membranes. NS1, NS3, and NS5 are big proteins that are remarkably like each other. NS2A, NS2B, NS4A, and NS4B are smaller proteins that do not like water. There are 428 nucleotides in the 3' NCR that may help with translation, RNA packaging, cyclization, genome stability and recognition. The 3' NCR makes a loop, and the 5' NCR lets a methylated nucleotide cap, or a genome-linked protein do the translation. It has two non-coding regions on either side, which are called the 5' NCR and the 3' NCR. This is what the Zika virus's open reading frame looks like. A polyprotein is coded for by 5'-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'. This polyprotein is then cut into capsid (C),

precursor membrane (prM), envelope (E), and non-structural proteins (NS). The E protein makes up most of the surface of the virion and helps with replication by sticking to host cells and fusing membranes. NS1, NS3, and NS5 are big proteins that are like each other. NS2A, NS2B, NS4A, and NS4B are smaller proteins that do not like water. 428 nucleotides can be found in the 3' NCR. They may help with translation, RNA packaging, cyclization, genome stability and recognition. The 3' NCR makes a loop, and the 5' NCR lets a methylated nucleotide cap, or a genome-linked protein do the translation.

Structural and non-structural proteins play an important part in controlling the Zika virus's life cycle and how it spreads. Here is a list of what they are and what they do:

Table 1.1: Functions and Locations of Zika Virus Proteins (Rawal et al., 2016)

Protein	Location	Function
Structural Protein		
Capsid	Structural Core protein	Nucleocapsid formation by binding to viral RNA
Precursor Membrane	Structural Surface Protein	E protein stabilization, Host Cell fusion
Envelope	Structural Surface Protein	Host Receptor binding, Host cell fusion, Viral entry
Non-structural Proteins Mediate viral transcription and mitigate host antiviral responses		
NS1		Virus replication
NS2a		Virus Transcription, Virus assembly
NS2b		NS3 cofactor for appropriate serine protease function
NS3	Encoded by viral RNA but transcribed via host cellular machinery.	Serine protease activity, Helicase activity, Triphosphatase activity
NS4a		Viral replication
NS4b		Viral replication
NS5		Viral RNA dependent, Polymerase activity, RNA capping, Methyl Transferase Activity

1.5) Vectors and Host Range of Zika Virus:

The most common way for ZIKV to spread is through mosquito bites. The most important mosquito that spreads ZIKV is *Aedes (Stegomyia)*, but the virus can also be spread by *Anopheles*, *Culex*, *Mansonia*, and *Hermaphrodites*. ZIKV was first found in *Aedes africanus*, which is an important part of the ZIKV sylvatic transmission cycle. The *Aedes Aegypti* and *Aedes Albopictus* mosquitoes are more important in the urban spread cycle because they live in more places. Figure 1.4 shows the pattern of how the ZIKV virus is spread from infected to healthy people by mosquito bites. The reservoir host-mosquito-reservoir host transfer cycle takes between 2 and 5 days for the reservoir host and between 5 and 7 days for the mosquito transmission without a vector, which can happen through sexual contact, blood transfer, or the placenta. Other ways that viruses can be passed are through breast milk, saliva, or urine but there is no solid proof that ZIKV can be transmitted this way. More in-depth research is needed to determine how likely it is that ZIKV can be transmitted uniquely. At this point, these routes that are not spread by vectors are not getting as much study attention as those that are spread by mosquitoes. The methods used to get rid of *Ae. aegypti* are still not working because there are a lot of mosquitoes around. It looks like these plans were made as a public health reaction to a ZIKV epidemic. To use these methods, the vector's life cycle must be stopped in its native environment, susceptible people must be exposed to the vector less, the vector's source must be limited to urban populations, the vector must be eliminated, and infected mosquitoes must not bite people. Because of these things, this mosquito is an ideal vector for spreading dangerous viruses like Zika, Dengue, West Nile, and Yellow Fever.

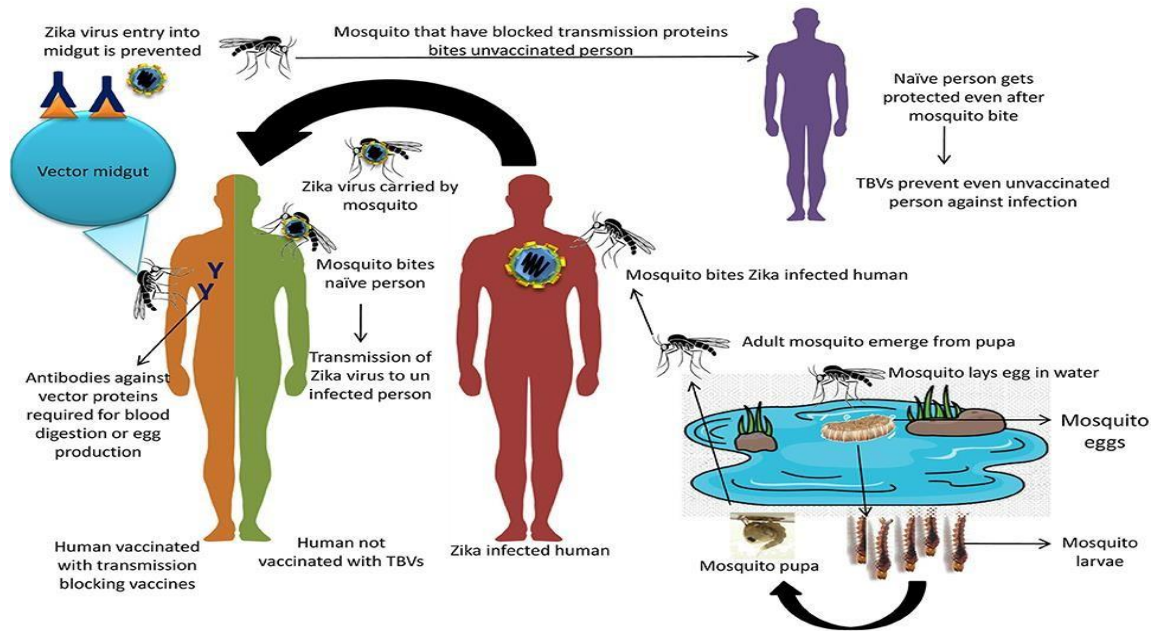


Figure 1.4: Modes of ZIKV transmission between infected and uninfected persons through mosquito bite and mode of TBVs in preventing and controlling ZIKV infection.

1.6) Mode of Transmission:

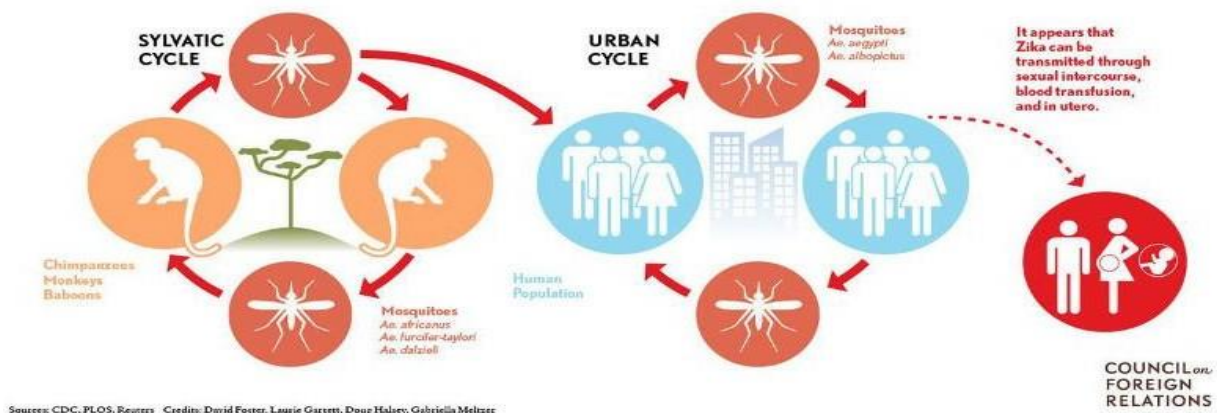


Figure 1.5: How ZIKV enters the human population (Garcez et al., 2017).

Aedes Aegypti and *Aedes Albopictus* mosquitoes infected with the Zika virus are the main ways the disease gets spread. In warm and subtropical areas, these mosquitoes do very well. When a mosquito bites a person who has Zika, it gets the virus. The most biting times of the day for *Aedes* mosquitoes are early in the morning and late afternoon or evening. The Zika virus can also be passed from a pregnant woman to her unborn child through sexual interaction, blood transfusions, and organ transplants.

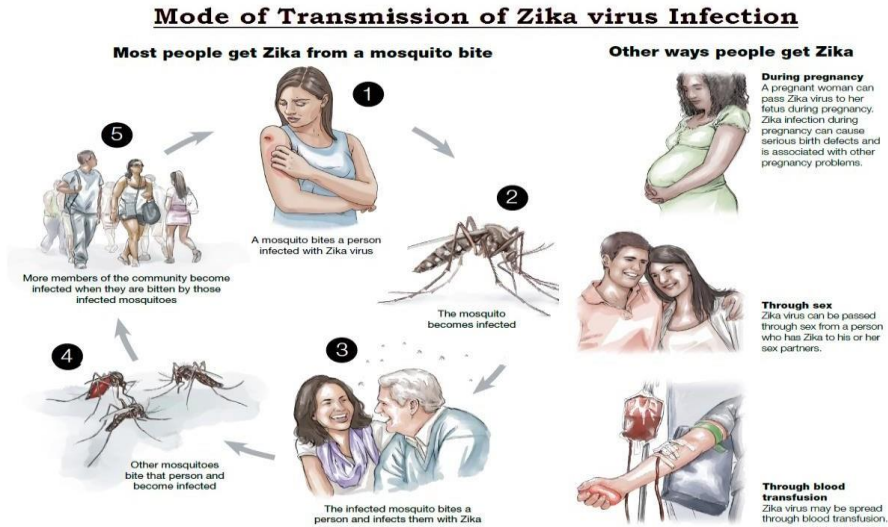


Figure 1.6: Mode of transmission of Zika virus infection (Sager et al., 2018).

1.7) Symptoms and Complications by Zika Virus:

The first signs of Zika can show up 3 to 12 days after an infected mosquito bites a person, but it can be shorter or longer for some. Most people who have the Zika virus do not have any signs. If someone gets the Zika virus, they usually only get sick for two to seven days and feel fine. Some of the following signs and symptoms may point to an infection with the Zika virus:

- Rash
- Itching or pruritus
- Fever
- Headache
- Arthralgia or arthritis
- Myalgia
- Conjunctivitis
- Lower back pain
- Retro-orbital pain

Zika symptoms are like those of dengue (which is caused by a related flavivirus) or chikungunya (which is caused by an alphavirus). These viruses often spread together in places where Zika virus is present. To make the right determination, lab tests are needed.

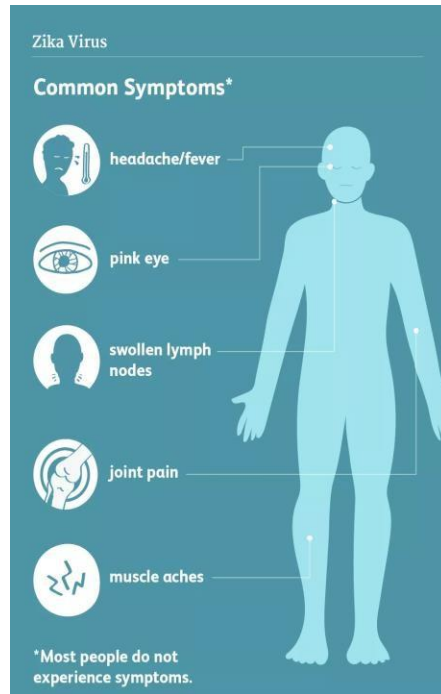


Figure 1. 7: Symptoms of Zika Virus.

1.8) Complications of Infection:

Guillain-Barré syndrome (GBS) is an extremely rare but potentially fatal consequence of Zika infection. It occurs when the immune system mistakenly targets its own nerve cells. Rare as it is, the disorder can weaken the muscles that govern respiration and, in the worst instances, the muscles that control the arms and legs. Gonorrhea symptoms often last for weeks or months. Some patients may experience long-term nerve damage, but the majority will make a full recovery. In rare cases, GBS can be fatal. At this time, there is no vaccine that can protect against the ZIKA virus. This is the only medicine that the FDA has allowed to stop the Zika virus from infecting animals other than humans. ZIKV, on the other hand, is too much for this medicine to oversee right now. To make a vaccine against the Zika virus, it might help to look at how it differs from viruses that are linked to it.

Acetaminophen and ibuprofen are two types of painkillers that can help you deal with signs like fever and pain. Unfortunately, there is no known way to get rid of the Zika virus. To stop the disease from spreading, it is important to focus on stopping it. Taking steps to avoid getting bitten by bugs is the best way to do this. If we want to stay safe from mosquitoes, we should only use bug sprays that have been approved by the EPA. Also, we should cover our skin with long sleeves

and pants and stay away from places where mosquitoes like to breed, like still water near homes (Dar et al., 2016). Nevertheless, there are some problems with these methods as well. It can be hard to get rid of all places for breeding, especially in areas with many people. Also, repellents may be uncomfortable to use because they need to be put on so often. Some kinds of clothes should not be worn when it is hot outside.

Because pregnant women are at a much higher risk, it is particularly important to take precautions. There is a chance of microcephaly and other birth problems, so you might want to do more tracking. The steps that need to be taken are still the same. However, it is important to remember that condoms are not perfect and have their limits, even though they can help lower the risk of SST. For the sake of their health, pregnant women should stay away from places where Zika is spreading. Another crucial point that is brought up is that people may not always be able to avoid everything, especially if they live near sensitive places. Zika needs more study because there is not a vaccine or a specific way to treat it (Dikhit et al., 2016).

In the past few years, bioinformatics works on the ZIKA virus have become immensely popular. Some of those studies gave us ideas for this one. The study found B Cell epitopes for the Zika virus vaccine. A small number of B cell studies on the Zika virus were found. However, many studies have tried to guess what T cell epitopes are in ZIKV polyprotein. They found 23 epitopes of MHC I molecules and 48 epitopes of MHC II molecules in one study (Byler et al., 2016). The forecast and analysis of possible antigenic CTL epitopes in the Zika virus were done in a different study. The immune-informatics method was used to look for possible major histocompatibility complex classes. I restricted epitopes that might be able to make a human cell-mediated immune reaction. After searching for a long time, 63 epitopes were found that bind strongly to all 42 types of human leukocyte antigens (A. K. Gupta et al., 2016). Molecular docking was used in another study to find plants that fight ZIKA. The virus protease, methyltransferase, and RNA-dependent RNA polymerase were modeled and matched with a collection of phytochemicals that we had on hand. Finally, 43 chemicals worked on the targets like drugs (Z. Wang & Ma'ayan, 2016).

1.9) Aims and Objectives:

- To learn more about the Zika virus and its effects right now.
- The goals are to find out what the Zika virus can do and to compare it to other viruses that are similar.
- To find the conserved regions of all related viruses that stay the same and describe what they do.
- To construct a full vaccine with suitable epitopes that can fight the Zika virus.

Chapter – 2

Materials and Methods

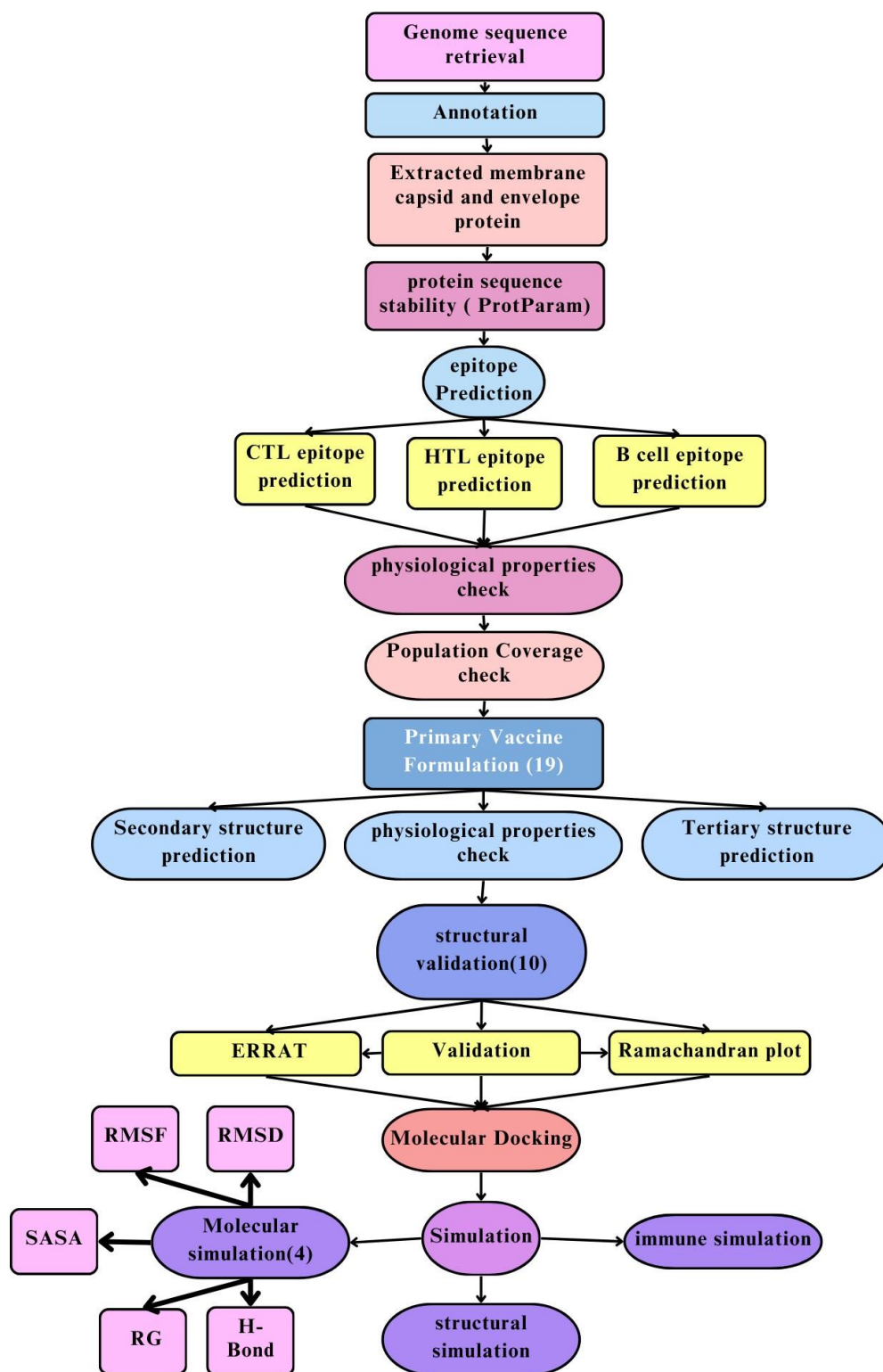


Figure 2.1: Overview of the Methodology

2.1) Retrieval of Genome sequence and analysis

We took the complete genome of Zika virus sequences from the BV-BRC server (<https://www.bv-brc.org/>) (Olson et al., 2023) including Asian and African variants. The variants were checked against the Genome Detective Virus tool (<https://www.genomedetective.com/app/typingtool/virus/>) (Vilsker et al., 2019) to confirm the origin of the virus. A phylogenetic tree was generated by the Iqtree2 tool and visualized with the iTOL (<https://itol.embl.de/tree/103133140163172171706901924>) (Letunic & Bork, 2021) server. The genome sequences were annotated using VAPiD to polyprotein and later from the polyprotein, capsid, membrane, and envelope protein were extracted using Samtools. Multiple Sequence alignments were done using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Madeira et al., 2022) to find the common and conserved region of the sequence.

2.2) Physicochemical properties of protein sequences

The stability index was then checked using the Protparam tool (<https://web.expasy.org/protparam/>) (Duvaud et al., 2021). signifying the score for capsid protein as stable along with the aliphatic index and Grand Average of hydropathicity (GRAVY). The Glycoprotein or membrane protein is also classified as a stable protein with standard antigenic value. Envelope protein instability index computed and classified as table protein with moderate antigenic value. The weight of the Half-life of the respective protein was checked.

Antigenicity and allergenicity are some of the most important factors to check hence the constructed vaccines were checked by VaxiJen 2.0 (<https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova & Flower, 2007) since the Antigenic properties of vaccines assure an immune response and Allerprot (<https://allercatpro.bii.a-star.edu.sg/>) (Nguyen et al., 2022) were used to check the antigenicity.

Solubility analysis was also checked via the Protein-Sol online server <https://protein-sol.manchester.ac.uk> (Hebditch et al., 2017); it checks the solubility from the sequence scaled solubility value (QuerySol) as the predicted solubility. The average for the experimental dataset (PopAvrSol) is 0.45, and therefore any scaled solubility value greater than 0.45 is predicted to have a higher solubility, and any protein with a lower scaled solubility value is predicted to be less soluble, so we only took the value more than 0.45 for the selected vaccines and analyzed further because of their stability. To Check the protein stability, hydrophilic or hydrophobicity of the

constructed vaccines, Protparam was again used for the accurate predictions. The physical and chemical parameters of the given sequence of the protein are calculated. The tool can measure the molecular weight, theoretical Pi, amino acid compositions, estimated half-life, instability index, aliphatic index, and GRAVY, which stands for the Grand Average of Hydropathicity. All these indicate a better candidate for vaccines.

2.3) Prediction of CTL, HTL, and B cell epitope.

The cytotoxic T cells potentially kill the virus and eliminate germs, regenerate immune response and decrease the proportion of the pathogens. Hence, a prominent vaccine must contain the CTL to perform a better immune response against the vaccine. (The Immune Epitope Database) IEDB (<https://www.iedb.org/>) (Vita et al., 2019) server was used to predict the CTL epitopes as MHC-I selecting the HLA total set. Tools of IEDB are delineated to host the epitope prediction and analyze it in the context of human and other animal species with reference to infectious disease. Helper T cells are crucial for adaptive immunity and are key in generating antibodies and antigen-presenting cells. Both prominently required vaccines were also predicted using the IEDB online server. The best functional epitopes were selected using the threshold value of 0.8. A few criteria were chosen to identify the best epitopes. ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/multi_submit.php) (S. Gupta et al., 2013) was used to determine the toxicity. The epitopes must not show allergenicity. As a result, AllerProt was used to predict the epitopes' allergenicity. Again, VaxiJen is used to calculate the antigenicity for antigen classification, which depends on physicochemical characteristics. Additionally, The B cell plays a key role in adaptive immune response, which is in charge of humoral immunity. Bepipred 2.0 (<http://tools.iedb.org/bcell/>), (Chou & Fasman, 2006) The sequential B cell epitope Prediction tool predicted the specific B cell epitope from a protein sequence. Later, the selected epitopes were used to generate a three-dimensional structure using PyMOL. (The PyMOL MolecularGraphics System, 2010) tool.

2.4) Population coverage analysis of the epitopes

The IEDB server was employed to calculate the population coverage and determine the individual and multi-epitope impact on the population. The server allows the analysis and prediction of the overall coverage of the infected patient population earmarked by the selected uni-peptide or multi-

peptide-based vaccines by determining peptides with multiple and different HLA binding specificities. The targeted MHC-I and MHC-II were selected, and the allelic frequency was used to predict correlated epitopes. Class I and Class II have been done both individually and combined. MCH population studies employing the population coverage tool by IEDB (<http://tools.iedb.org/population/>) (Bui et al., 2006) showed descriptive statistics on population coverage. The world population of the class covers average hits, and Pc90 is the analysis done all over the world, selecting both African, Asian, and African separately. The highest coverage was seen on many continents at the same time.

2.5) Multiepitope Vaccine Construction

The objective of the study was to construct the best candidate for a vaccine against the Zika virus and assess them based on their efficiency. The adjuvants for an increased immune response have been used in the vaccine constructs. Selected adjuvants and linkers are.

Table 2.1: Selected adjuvants and linkers.

NAME	SEQUENCE
Beta Defensin-3	GIINTLQKYYCRVRRGRCVLSCLPKKEEQIGKCSTRGRKCCRRKK
50s ribosomal protein	MSITKDQIIIEAVAAMSVM DVVELISAMEEKFGVSA AAAVAVAAG PVEAAEEKTEFDVILKAAGANKVAVIKAVRGATGLGLKEAKDLV ESAPAALKEGVSKDDAEALKKALEEAGAEVEVK
Linker	EAAK
PADRE	AKFVAAWTLKAAA
MHC- I	AA Y and GGGGS
MHC- II	GPGPG
B CELL	KK
Histidine	HHHHHH

2.6) Structure prediction, and validation

The secondary structure was predicted using NetSurfP -3.0 (<https://services.healthtech.dtu.dk/services/NetSurfP-3.0/>) (Høie et al., 2022) server to understand the protein surface accessibilities, secondary structure, and dihedral angles. Moreover, tertiary structure prediction widens the ability to understand and analyze the protein structural data analysis. Therefore, the three-dimensional vaccines were predicted from the online server Rosetta Fold (<https://rosetta.bakerlab.org/myqueue.php>). Then, PyMOL was used to generate a three-dimensional image. It analyzes the protein sequence and investigates how the protein's amino acids work and interact based on the three-track network; it accurately images the submitted protein's three-dimensional structure. Statistics and principles from ab initio progress evaluate the helix, coil, side chain, and other structural properties.

The validation process helps to enhance the 3d model quality of the selected protein sequence. Then, after that, validation was checked via saves v6.0 (<https://saves.mbi.ucla.edu/>) (Baek et al., 2021). The ERRAT, verify 3D, PROcheck, and whatcheck algorithms were used for each vaccine construct.

2.7) Molecular docking

The molecular docking was performed by using the Cluspro server (<https://cluspro.org>) (Desta et al., 2020) that uses high-throughput proteomics to calculate the interaction between two proteins or ligands. It uses the three computational steps, which include sampling many confirmations that are close to billions, generating the best models of the complex by root-mean-square (RMSD) based clustering structure that consists of the lowest energy, and refining the structure with energy minimization. We submitted the Protein Bank Data (PDB) of the vaccine as a ligand and receptor TLR4 PDB file.

Binding affinities for the constructed vaccines against the receptor TLR4 were done by a prodigy server (<https://wenmr.science.uu.nl/prodigy/>) (Honorato et al., 2021) which calculated the Gibbs free energy for each selected vaccine. The server allows us to predict the binding affinities for the submitted vaccine complex and determine the molecular interfaces from crystallographic ones. Gibbs free energy plays a significant role in understanding the thermodynamics of protein structure.

The residue Interaction Network Generator (RING) (<https://ring.biocomputingup.it>) (Clementel et al., 2022) can indicate the protein's atomic non-covalent interactions by processing the multi-state structure, including molecular dynamic and structural ensembles. It calculates linearly with the input size and generates the statistically probabilistic network and conformation-dependent contact maps for each interaction. It enables the defining the hydrogen atom donors and acceptors, the addition or deletion of hydrogen atoms with accurate labeling and arrangements, the identification of generalized aromatic rings, novel interactions that cover Pi-hydrogen bonds, Vaccine constructions with more than three hydrogen bonds in terms of intermolecular interactions demonstrate a stable protein.

2.8) Immune simulation

The vaccine candidates were set to perform the immune stimulation by C-immsim (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>) (Rapin et al., 2010). It administered immunoglobulin and immune complexes. By submitting the sequence, the increase of the b-cell population, T-cell population, and coverage throughout the vaccination, count of active dendritic cell and macrophage population after a single shot of vaccination, it also includes the changes of concentration of several other types of cytokines throughout the three injections.

2.9) Structural simulation

For the structural simulation of the targeted vaccine, we used the online server CABS-FLEX 2.0 (<https://biocomp.chem.uw.edu.pl/CABSflex2/submit>) (Kuriata et al., 2018). It stabilized the constructed structure to be constant and overlooked the functionalities of the vaccine construct.

The expression level of any protein, depending on its location, structure, and interactions, provides important insights into the function and efficiency of the vaccine. The simulation is done automatically by analyzing and processing the trajectory and regenerating all the atom representations to make the structure as accurate as possible with a spatial resolution that includes a C-alpha chain format. Also, the K-medoids approach was used for trajectory clustering, which enables the understanding of multi-step reconstruction with higher optimization. Selected four vaccines were submitted to generate the individual and clustered models, contact maps, fluctuation plots, and other project information.

2.10) Molecular Dynamic simulation

To run the molecular dynamics simulation GROMAC (H. Bekker et al., 1993). Open-source software was used to calculate and analyze various parameters of the vaccines. RMSF (root mean square fluctuation), RMSD (root mean square deviation), RG (radius of gyration), SASA (solvent accessible surface area), and Hydrogen bond were generated for the simulation by GROMAC. A force field was added, and solvation was done by adding SPCE216 as a water molecule. The charm27 captured the vaccine in a dodecahedron box that was added to centralize the vaccine protein. Na⁺ and Cl⁻ ions were added to the neutralization of charges of the protein, and energy minimization was conducted with the input algorithm. The thermodynamic analysis was inclined and adjusted to the temperature, pressure, and volume. The system used NVT and NPT in a default steepest descent algorithm with 500000 steps. All the non-bonds, surface, and hydrogen bond interactions were conducted under the algorithm, and finally, 50000 picosecond simulation was done with the TLR4-vaccine complex.

Chapter – 3

RESULTS

3.1) Genome Sequence Retrieval

Zika Virus genome sequence was retrieved from the BV BRC server by selecting genome type as complete genome sequence. The collection year was selected as the most recent to discover neoteric variants. No backdated viruses were chosen due to the possibility of getting lost due to the fast evolution capabilities of RNA viruses. Fifteen viruses of recent time were checked and selected. Using the genome detective virus tool, all selected Zika variants were checked again; 15 of them were originally from Asia, and 3 of them were from Africa. It can be understood that later variants spread firmly around the world. Later, the genomes sequences were annotated and converted to protein sequences for later analysis. Lastly, the capsid, envelope, and membrane proteins were extracted to generate epitopes against them.

3.1.1) Phylogenetic tree

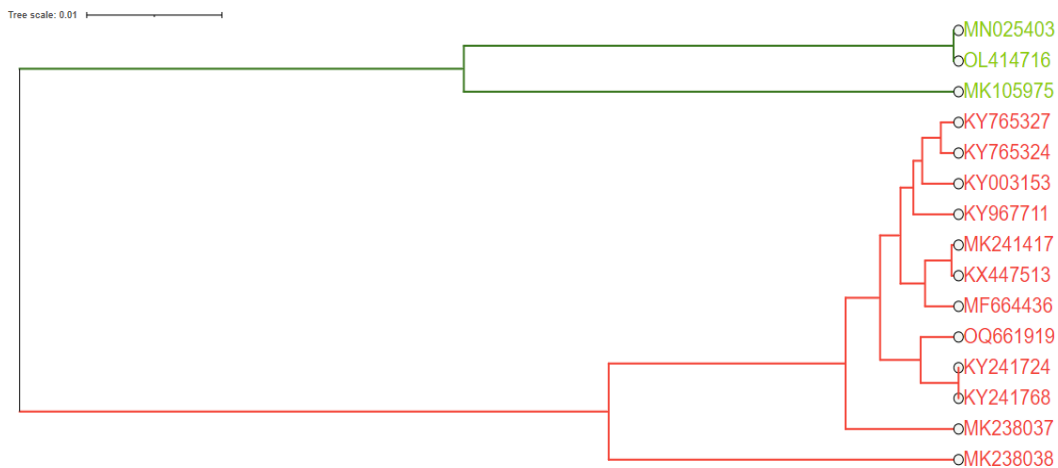


Figure 3.1: Phylogenetic tree of the Zika variants

Phylogeny is the study of how distinct groups of creatures, like species or populations of a species, have changed over time. A phylogenetic tree is a graphical representation of phylogeny that shows how intricately linked organisms are to each other and how far apart they are in terms of evolution. There are six main parts of a phylogenetic tree: the root, the branch, the nodes, the leaves, and the clades.

In an evolutionary tree, branch length is an important measure that shows how many changes have happened in the branch.

You can use either distance-based or character-based ways to build a phylogenetic tree. The distance-based methods, on the other hand, are more common and work better.

The Zika virus transmitted to humans through mosquitoes and the most recent variants we choose are from Asia which caused the most suburban-urban transmission. In the multifurcating tree, 3 recent Zika viruses MN025403, OLA414716, MK105975 colored in green. from Africa and the other 12 Asian variants colored in red, descendants arrived from each of the interior nodes, all connected to one root.

3.1.2) Analysis of the retrieved Complete Genome sequences of Zika Virus

After retrieving the genome sequences from BV-BRC the, it is again checked by Genome Detective Virus Tool Version 2.72 for origin of the variants of the Zika virus. After annotating the sequence to the protein sequence, the instability, GRAVY, and aliphatic index were checked for further analysis. The table below shows the accession number of the retrieved genome sequence, type, and the origin of them.

Table 3.1: Accession number of the retrieved genome sequence, type, and the origin.

Accession no	Organism	Sequence type	Origin
KY765324	Zika virus strain	Complete genome.	Asian
KY765327	Zika virus strain	polyprotein gene, complete cds.	Asian
MK238037	Zika virus strain	polyprotein gene, complete cds.	Asian
MK238038	Zika virus strain	polyprotein gene, complete cds.	Asian
KY967711	Zika virus strain	polyprotein gene, complete cds.	Asian
OQ661919	Zika virus strain	complete genome.	Asian
KY241724	Zika virus strain	polyprotein gene, complete cds.	Asian
KY241768	Zika virus strain	polyprotein gene, complete cds.	Asian
MF664436	Zika virus strain	polyprotein gene, complete cds.	Asian
KY003153	Zika virus strain	polyprotein gene, complete cds.	Asian
KX447513	Zika virus strain	polyprotein gene, complete cds.	Asian

KX806557	Zika virus strain	complete genome.	Asian
MK105975	Zika virus strain	complete genome.	African
MK241416	Zika virus strain	complete genome.	African
MN025403	Zika virus strain	complete genome.	African

3.1.3) Extraction of protein sequences

The table contains the start and end positions of the three specific proteins which are envelope, capsid and membranes that are used to predict the epitopes for the multipartite vaccine designing.

Table 3.2: The start and end positions of the capsid, membrane, and envelope proteins.

Protein Accession	Organism	Capsid		Envelop		Membrane	
		Start	End	Start	End	Start	End
MK105975.1	Zika	99	464	969	2468	465	967
MN025403.1	Zika	93	458	963	2474	459	962
MT505349.1	Zika	107	177	1691	3202	Not Found	
MT505350.1	Zika	107	172	1691	3202	Not Found	
OK054351.1	Zika	94	459	964	2463	462	962
OL414716.1	Zika	104	469	974	2485	470	973
OM522327.1	Zika	107	472	977	2488	473	976
OM666891.1	Zika	99	464	969	1423	465	968
OM666892.1	Zika	95	460	965	1419	461	964
OM666893.1	Zika	98	463	968	1422	464	967
OM964565.1	Zika	31	396	901	2412	397	900
OM964566.1	Zika	31	396	901	2412	397	900
OM964567.1	Zika	31	396	901	2412	397	900
OM964568.1	Zika	31	396	901	2412	397	900
ON209935.1	Zika	108	473	978	2489	474	977

Later, each sequence was evaluated via ExPasy-ProtParam tool to computationally calculate the diverse types of physical and chemical properties like the Grand Average of Hydropathy value (GRAVY) of the protein. Besides, the instability index was checked to acknowledge the measurements and determination of given protein. Also, an aliphatic index was calculated to predict stability.

Table 3.3: Physical and chemical properties of the annotated Protein sequences of capsid, envelope, and membrane.

Accession no	Capsid			Envelope			Membrane		
	GRAVY	Aliphatic index	Stability	GRAVY	Aliphatic Index	Stability	GRAVY	Aliphatic index	Stability
MK105975.1	0.103	114.43	stable	-0.066	80.90	stable	-0.186	82.13	stable
MN025403.1	0.103	114.43	stable	-0.060	80.83	stable	-0.163	82.44	stable
MT505349.1	0.452	106.97	stable	-0.107	84.31	stable	-0.637	60.95	stable
MT505350.1	0.452	106.97	stable	-0.107	84.31	stable	-0.637	60.95	stable
OK054351.1	0.134	110.41	stable	-0.077	82.10	stable	-0.228	78.86	stable
OL414716.1	0.074	115.50	stable	-0.060	80.83	stable	-0.163	82.44	stable
OM522327.1	0.134	110.41	stable	-0.071	82.02	stable	-0.202	80.12	stable
OM666891.1	0.134	110.41	stable	-0.100	81.79	stable	-0.202	80.12	stable
OM666892.1	0.134	110.41	stable	-0.077	82.10	stable	-0.202	80.12	stable
OM666893.1	0.134	110.41	stable	-0.077	82.10	stable	-0.202	80.12	stable
OM964565.1	0.134	110.41	stable	-0.071	82.02	stable	-0.202	80.12	stable
OM964566.1	0.134	110.41	stable	-0.072	81.97	stable	-0.202	80.12	stable
OM964567.1	0.134	110.41	stable	-0.071	82.02	stable	-0.202	80.12	stable
OM964568.1	0.134	110.41	stable	-0.071	82.02	stable	-0.202	80.12	stable
ON209935.1	0.134	110.41	stable	-0.071	82.02	stable	-0.202	80.12	stable

After finding these 15 sequences, desired regions like membrane, capsid, or envelope proteins were extracted. The conserved regions were checked using the multiple sequence alignment tool Clustal Omega. Then, the epitopes from these regions were predicted.

3.2) Epitope prediction

3.2.1) Cytotoxic T-Lymphocyte and Helper T-Lymphocyte Epitopes Prediction

3.2.1) (a) MHC-I and MHC-II for Envelope proteins

Epitopes are a diverse group of amino acids that are exposed on the exterior part of a molecule and cause to generate antigen reactions, and later bind to an antibody. Using the IEDB server, a capping threshold of 0.8 for better affinity to the MHC-I or CTL epitopes and MHC-II or HTL was predicted for the envelope. 38 MHC-I epitopes for envelope protein were selected primarily. 22 MHC-II epitopes were selected. Then, it was sorted according to different parameters in the table. Later, the peptide sequence went for antigenic, toxicity, and allergenicity tests, respectively.

Table 3.4: MHC-I of the envelope protein.

No	Allele	Length	Peptide Sequence	Score	Antigenicity and score	Toxicity	Allergenicity
1	HLA-B*57:01	9	KSLFGGMSW	0.993924	ANTIGEN 0.7431	Non-Toxin	Nonallergic
2	HLA-B*58:01	9	KSLFGGMSW	0.992316	ANTIGEN 0.7431	Non-Toxin	Nonallergic
3	HLA-A*68:02	9	ETLHGTVTV	0.979822	ANTIGEN 0.5422	Non-Toxin	Nonallergic
4	HLA-B*15:01	9	SQHSGMIGY	0.976799	ANTIGEN 1.2234	Non-Toxin	Nonallergic
5	HLA-A*01:01	10	YLDKQSDTQY	0.975213	NON-ANTIGEN 0.3207	Non-Toxin	Nonallergic
6	HLA-B*40:01	9	QEGAVHTAL	0.963484	NON-ANTIGEN 0.0330	Non-Toxin	Nonallergic

7	HLA-B*44:03	10	AEMDGAKGKL	0.937879	NON-ANTIGEN 0.1803	Non-Toxin	Nonallergic
8	HLA-A*01:01	9	GLDFSDLYY	0.937595	ANTIGEN 1.7878	Non-Toxin	Nonallergic
9	HLA-B*44:02	10	AEMDGAKGKL	0.930873	NON-ANTIGEN 0.1803	Non-Toxin	Nonallergic
10	HLA-A*01:01	10	ELDPPFGDSY	0.929162	ANTIGEN 0.8833	Non-Toxin	Nonallergic
11	HLA-B*15:01	10	YLDKQSDTQY	0.926636	NON-ANTIGEN 0.3207	Non-Toxin	Nonallergic
12	HLA-A*02:03	9	GLFGKGSV	0.919098	ANTIGEN 0.8202	Non-Toxin	Nonallergic
13	HLA-A*01:01	10	TGLDFSDLYY	0.91336	ANTIGEN 1.8632	Non-Toxin	Nonallergic
14	HLA-A*68:02	9	TTVSNMAEV	0.913108	ANTIGEN0.6972	Non-Toxin	Nonallergic
15	HLA-B*40:01	10	AEMDGAKGKL	0.912437	NON-ANTIGEN 0.1803	Non-Toxin	Nonallergic
16	HLA-A*03:01	10	RLKMDKLRLK	0.905263	NON-ANTIGEN -0.4819	Non-Toxin	Nonallergic
17	HLA-B*58:01	9	GADTGTPHW	0.899948	NON-ANTIGEN 0.3712	Non-Toxin	Nonallergic
18	HLA-B*40:01	9	KEWFHDIPL	0.891315	NON-ANTIGEN 0.0523	Non-Toxin	Nonallergic
19	HLA-B*57:01	10	FKSLFGGMSW	0.884249	ANTIGEN 0.5143	Non-Toxin	Nonallergic
20	HLA-B*58:01	10	KSIQPENLEY	0.882521	ANTIGEN 1.6902	Non-Toxin	Nonallergic
21	HLA-B*57:01	10	KSIQPENLEY	0.881357	ANTIGEN 1.6902	Non-Toxin	Nonallergic
22	HLA-A*02:03	9	SLGKGIHQI	0.878415	ANTIGEN 0.4021	Non-Toxin	Nonallergic

23	HLA-A*68:01	9	TVSNMAEVR	0.877489	ANTIGEN 0.7175	Non-Toxin	Nonallergic
24	HLA-A*68:01	10	TTVSNMAEVR	0.862622	ANTIGEN 0.7688	Non-Toxin	Nonallergic
25	HLA-B*07:02	9	TPNSPRAEA	0.856156	NON-ANTIGEN -0.1529	Non-Toxin	Nonallergic
26	HLA-A*02:01	9	SLGKGIHQI	0.855117	ANTIGEN 0.4021	Non-Toxin	Nonallergic
27	HLA-B*08:01	9	DAHAKRQTV	0.853751	ANTIGEN 0.4481	Non-Toxin	Nonallergic
28	HLA-B*57:01	10	VGDKKITHHW	0.851551	ANTIGEN 0.7202	Non-Toxin	Nonallergic
29	HLA-A*30:02	9	SQHSGMIGY	0.849265	ANTIGEN 1.2234	Non-Toxin	Nonallergic
30	HLA-A*32:01	9	KSLFGGMSW	0.829265	ANTIGEN 0.7431	Non-Toxin	Nonallergic
31	HLA-A*68:02	9	DTAWDFGSV	0.81275	ANTIGEN 1.9031	Non-Toxin	Nonallergic
32	HLA-B*58:01	10	AGADTGTPHW	0.810739	NON-ANTIGEN 0.3400	Non-Toxin	Nonallergic
33	HLA-A*68:02	9	TAAFTFTKV	0.80816	ANTIGEN 0.8764	Non-Toxin	Nonallergic
34	HLA-A*03:01	10	KLFSGHLKCR	0.807871	ANTIGEN 0.7229	Non-Toxin	Non-allergic
35	HLA-B*15:01	10	GSQHSGMIGY	0.807045	ANTIGEN 0.8890	Non-Toxin	Nonallergic
36	HLA-B*57:01	10	KSLFGGMSWF	0.805493	NON-ANTIGEN 0.3525	Non-Toxin	Nonallergic
37	HLA-A*01:01	9	SIQPENLEY	0.803348	ANTIGEN 2.1083	Non-Toxin	Nonallergic
38	HLA-B*58:01	10	FKSLFGGMSW	0.801729	ANTIGEN 0.5143	Non-Toxin	Non-allergic

MHC-II envelope proteins were predicted and represented with length, sequence, score, and physiological properties.

Table 3. 5: MHC-II for envelope protein.

No	Allele	Length	Peptide Sequence	Score	Antigenicity	Toxicity	Allergenicity
1	HLA-DRB1*07:01	15	AFTFTKVPAETLHGT	0.9028	NON-ANTIGEN 0.2817	Non-Toxin	Non-allergic
2	HLA-DRB1*01:01	15	NLEYRIMLSVHGSQH	0.9534	ANTIGEN 0.5466	Non-Toxin	Non-allergic
3	HLA-DRB1*04:05	15	TAAFTFTKVPAETLH	0.9190	NON-ANTIGEN 0.2749	Non-Toxin	Non-allergic
4	HLA-DRB1*01:01	15	ENLEYRIMLSVHGSQ	0.9302	NON-ANTIGEN 0.3983	Non-Toxin	Non-allergic
5	HLA-DRB1*01:01	15	AFTFTKVPAETLHGT	0.9255	NON-ANTIGEN 0.2817	Non-Toxin	Non-allergic
6	HLA-DRB1*11:01	15	VGGVFNSLGKGIHQI	0.9045	NON-ANTIGEN 0.2486	Non-Toxin	Non-allergic
7	HLA-DRB1*01:01	15	AAFTFTKVPAETLHG	0.9011	NON-ANTIGEN 0.1991	Non-Toxin	Non-allergic
8	HLA-DRB1*08:02	15	SDTQYVCKRTLVD RG	0.8598	NON-ANTIGEN -0.0058	Non-Toxin	Non-allergic
9	HLA-DRB1*07:01	15	AAFTFTKVPAETLHG	0.8966	NON-ANTIGEN 0.1991	Non-Toxin	Non-allergic
10	HLA-DRB1*08:02	15	DTQYVCKRTLVD RGW	0.8418	NON-ANTIGEN 0.1207	Non-Toxin	Non-allergic
11	HLA-DRB1*04:01	15	NLEYRIMLSVHGSQH	0.8696	ANTIGEN 0.5466	Non-Toxin	Non-allergic

12	HLA-DRB1*07:01	15	AETLHGTVTVEVQYA	0.8391	ANTIGEN 0.9004	Non-Toxin	Non-allergic
13	HLA-DRB1*07:01	15	TAAFTFTKVPAETLH	0.8436	NON-ANTIGEN 0.2749	Non-Toxin	Non-allergic
14	HLA-DRB1*04:05	15	CTAAFTFTKVPAETL	0.8921	NON-ANTIGEN 0.0511	Non-Toxin	Non-allergic
15	HLA-DRB1*04:05	15	LCTAAFTFTKVPAET	0.8770	NON-ANTIGEN 0.1171	Non-Toxin	Non-allergic
16	HLA-DRB1*11:01	15	GGVFNSLGKGIHQIF	0.8925	NON-ANTIGEN -0.1368	Non-Toxin	Non-allergic
17	HLA-DRB1*04:01	15	ENLEYRIMLSVHGSQ	0.8250	NON-ANTIGEN 0.3983	Non-Toxin	Non-allergic
18	HLA-DRB1*11:01	15	GKAFEATVRGAKRMA	0.8590	NON-ANTIGEN 0.3411	Non-Toxin	Non-allergic
19	HLA-DRB1*11:01	15	MQTLTPVGRLITANP	0.8570	NON-ANTIGEN 0.2260	Non-Toxin	Non-allergic
20	HLA-DRB1*11:01	15	SVGGVFNSLGKGIHQ	0.8361	NON-ANTIGEN 0.3172	Non-Toxin	Non-allergic
21	HLA-DRB1*11:01	15	IGKAFEATVRGAKRM	0.8050	NON-ANTIGEN 0.3664	Non-Toxin	Non-allergic
22	HLA-DRB1*01:01	15	PENLEYRIMLSVHGS	0.8082	ANTIGEN 0.4835	Non-Toxin	Non-allergic

3.2.1 (b) MHC-I and MHC-II for Capsid proteins

Capsid proteins are selected to generate epitopes against structural components called nucleocapsids, which help package viral genomes. Epitope scores with more than 0.5 were selected

for a higher quality affinity for MHC-II, and more than 0.8 were for MHC-I. 6 epitopes were selected fundamentally for both MHC-I and MHC-II. The table contains antigenicity, toxicity, and allergenicity tests.

Table 3. 6 : MHC-I for capsid protein

No	Allele	Length	Peptide	Sequence Score	Antigenicity and Score	Toxicity	Allergenicity
1	HLA-A*68:02	9	TTVSNMAEV	0.913108	Antigen 0.6611	Non-toxic	Non-allergic
2	HLA-A*68:01	9	TVSNMAEVR	0.877489	Antigen 0.7175	Non-toxic	Non-allergic
3	HLA-A*68:01	10	TTVSNMAEVR	0.862622	Antigen 0.7688	Non-toxic	Non-allergic
4	HLA-A*01:01	9	MAEVRSYCY	0.719988	Antigen 1.2831	Non-toxic	Non-allergic
5	HLA-B*15:01	9	SQKVIYLV	0.707585	Antigen 0.6601	Non-toxic	Non-allergic
6	HLA-A*03:01	9	LLGSSTSQK	0.719988	Antigen 0.7202	Non-toxic	Non-allergic

MHC-II capsid proteins were predicted and represented with length, sequence, score, and physiological properties.

Table 3. 7: MHC-II capsid protein.

No	Allele	Length	Sequence	Score	Antigenicity	Toxicity	Allergenicity
1	HLA-DRB1*12:01	15	TKHLIKVEN WIFRNP	0.6482	NON-ANTIGEN -0.0851	Non-Toxin	Non-allergic
2	HLA-DRB1*04:05	15	VRSYCYEAS ISDMAS	0.6101	ANTIGEN 0.7147	Non-Toxin	Non-allergic

3	HLA-DRB1*15:01	15	TKHLLIKVEN WIFRNP	0.5799	NON-ANTIGEN -0.0851	Non-Toxin	Non-allergic
4	HLA-DRB1*04:05	15	EVRSYCYE ASISDMA	0.5465	ANTIGEN 1.0152	Non-Toxin	Non-allergic
5	HLA-DQA105:01/ DQB103:01	15	IFRNPGFAL VAVAIA	0.5311	ANTIGEN 0.6476	Non-Toxin	Non-allergic
6	HLA-DRB1*12:01	15	YTKHLLIKVE NWIFRN	0.5256	NON-ANTIGEN 0.0807	Non-Toxin	Non-allergic

3.2.1) (c) MHC-I and MHC-II for Membrane proteins

Eleven epitopes were selected against the membrane proteins that scored more than 0.8. Membrane proteins are especially involved in cellular procedures. Seven MHC-II epitopes were generated based on the same peptide score, which is more than 0.8. Both were then evaluated for later analysis for antigenicity, toxicity, and allergenicity.

Table 3. 8: MHC-I membrane protein.

No	Allele	Length	Peptide Sequence	Score	Antigenicity and score	Toxicity	Allergenicity
1	HLA-B*15:01	9	SQHSGMIGY	0.976799	ANTIGEN 1.2234	Non-Toxin	Nonallergic
2	HLA-A*01:01	10	YLDKQSDTQY	0.975213	NON-ANTIGEN 0.3207	Non-Toxin	Nonallergic
3	HLA-A*01:01	9	GLDFSDLYY	0.937595	ANTIGEN 1.7878	Non-Toxin	Nonallergic
4	HLA-B*15:01	10	YLDKQSDTQY	0.926636	NON-ANTIGEN 0.3207	Non-Toxin	Nonallergic
5	HLA-A*02:03	9	GLFGKGSV	0.919098	ANTIGEN 0.8202	Non-Toxin	Nonallergic

6	HLA-B*58:01	9	GADTGTPHW	0.899948	NON-ANTIGEN 0.3712	Non-Toxin	Nonallergic
7	HLA-B*40:01	9	KEWFHDIPL	0.891315	NON-ANTIGEN 0.0523	Non-Toxin	Nonallergic
8	HLA-B*07:02	9	TPNSPRAEA	0.856156	NON-ANTIGEN -0.1529	Non-Toxin	Nonallergic
9	HLA-A*30:02	9	SQHSGMIGY	0.849265	ANTIGEN 1.2234	Non-Toxin	Nonallergic
10	HLA-B*58:01	10	AGADTGTPHW	0.810739	NON-ANTIGEN 0.3400	Non-Toxin	Nonallergic
11	HLA-B*15:01	10	GSQHSGMIGY	0.807045	ANTIGEN 0.7666	Non-Toxin	Nonallergic

MHC-II membrane proteins were predicted and represented with length, sequence, score, and physiological properties.

Table 3. 9: MHC-II of membrane protein.

No	Allele	Length	Peptide Sequence	Score	Antigenicity and score	Toxicity	Allergenicity
1	HLA-DRB1*01:01	15	NLEYRIMLS VHGSQH	0.9534	ANTIGEN 0.5466	Non-Toxin	Non-allergic
2	HLA-DRB1*01:01	15	ENLEYRIM LSVHGSQ	0.9302	NON-ANTIGEN 0.3983	Non-Toxin	Non-allergic
3	HLA-DRB1*04:01	15	NLEYRIMLS VHGSQH	0.8696	ANTIGEN 0.5466	Non-Toxin	Non-allergic
4	HLA-DRB1*08:02	15	SDTQYVCK RTLVD RG	0.8598	NON-ANTIGEN -0.0058	Non-Toxin	Non-allergic
5	HLA-DRB1*08:02	15	DTQYVCKR TLVDRGW	0.8418	ANTIGEN 0.8202	Non-Toxin	Non-allergic

6	HLA-DRB1*04:01	15	ENLEYRIM LSVHGSQ	0.8250	NON-ANTIGEN 0.1207	Non-Toxin	Non-allergic
7	HLA-DRB1*01:01	15	PENLEYRIM LSVHGS	0.8082	ANTIGEN 0.4104	Non-Toxin	Non-allergic

3.2.2) B cell Epitopes prediction

3.2.2) (a) B cell epitope prediction for Envelope

The envelope protein sequence was given as input, and the B cell epitope was predicted using the BepiPred2.0 server of IEDB. Eighteen epitopes were predicted, but lengths less than ten amino acid sequences were automatically disdained. The predicted score and length are shown in the table and are later assessed for antigenicity, toxicity, and allergenicity.

Table 3. 10: B cell epitope prediction for envelope protein.

No	Peptide sequence	Length	Antigenicity	Toxicity	Allergenicity
1	GVSNRDFV	8	ANTIGEN 1.3853	Non-Toxin	Nonallergic
2	S	1	No result	No result	
3	E	1	No result	No result	
4	SDMASDSRCPTQGEAYLDKQS DTQYVCKRTLVD RGWGN	38	NON-ANTIGEN 0.2545	Non-Toxin	Nonallergic
5	G	1	No result	No result	
6	SQHSGMIGYETDEN	14	ANTIGEN 0.8727	Non-Toxin	Nonallergic
7	P	1	No Result	No result	
8	TGLD	4	No Result	No result	
9	FHDIPLPWHAGADTGT PHWN NKE	23	ANTIGEN 0.9099	Non-Toxin	Nonallergic
10	AHA	3	No result	No result	

11	AKGK	4	No result	No result	Nonallergic
12	TFTKVPAETL	10	NON-ANTIGEN -0.4575	Non-Toxin	Nonallergic
13	TDG	3	No Result	No result	
14	MQTL	4	No Result	No result	
15	STEN	4	No result	No result	Nonallergic
16	KITHHWHRSGSTIGK	15	NON-ANTIGEN 0.3983	Non-Toxin	Nonallergic
17	EA	2	No result	No result	
18	WDFGSVGG	8	ANTIGEN 2.2400	Non-Toxin	Nonallergic

3.2.2 (b) B cell epitope prediction for Capsid

The capsid protein sequence was given as input and predicted the B cell epitope using the BepiPred2.0 server of IEDB. Four epitopes were generated based on the capsid protein sequence, and less than ten peptide sequences were cut off. The predicted score and length are shown in the table and are later assessed for antigenicity, toxicity, and allergenicity.

Table 3. 11: B cell epitope prediction for capsid.

No	Peptide	Length	Antigenicity and score	Toxicity	Allergenicity
1	SSTSQ	5	No Result	No Result	No Result
2	RCIGVSNRDFVEGMSG	16	ANTIGEN 1.2515	Non-Toxin	Non-allergic
3	MAE	3	No Result	No Result	No Result
4	IS	2	No Result	No Result	No Result

3.2.2) (C) B cell epitope prediction for Membrane

Using the BepiPred2.0 service of IEDB, the membrane protein sequence was used as input to predict the B cell epitope. Four epitopes were produced based on the capsid protein sequence, and fewer than ten peptide sequences were removed. The table displays the anticipated score and duration, which is subsequently assessed for allergenicity, toxicity, and antigenicity.

Table 3. 12: B cell epitope prediction for membrane.

No.	Peptide	Length	Antigenicity	Toxicity	Allergenicity
1	EWFHDIPLPWHAGADTGTPHWNNK	24	ANTIGEN 0.6397	Non-Toxin	Nonallergic
2	G	1	No result	No result	No result
3	GSQHSGMIGYETDEN	15	ANTIGEN 0.7170	Non-Toxin	Nonallergic
4	RTGLD	5	No Result	No Result	No result
5	TQGEAYLDKQSDTQYVCKRTLVD RGWGN	28	NON-ANTIGEN 0.1243	Non-Toxin	Nonallergic
6	W	1	No result	No result	No result

3.2.3) Selected Epitopes

The best candidates for the epitopes were selected for further analysis based on the highest antigenic values, non-allergenicity, and non-toxicity. The table below contains the types of proteins: Capsid, envelope, and membrane. Name epitopes are MHC-I, MHC-II, and B cell, with length and 3D figure to better understand the inspection.

The selected epitopes are shown in the table below.

Table 3. 13: Selected CTL, HTL, and B-cell epitopes.

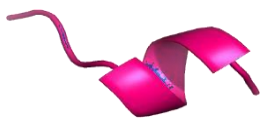
Capsid			
CTL (MHC-I)		HTL(MHC-II)	B CELL
1	TVSNMAEVR	VRSYCYEASISDMAS	RCIGVSNRDFVEGMSG
2	LLGSSTSQK	EVRSYCYEASISDMA	
3	MAEVRSYCY	IFRNPGFALVAVAIA	
Envelope			
CTL (MHC-I)		HTL(MHC-II)	B CELL
1	SIQPENLEY	HAKRQTVVVLGSQEG	GVSNRDFV
2	KSLFGGMSW	AETLHGTVTVEVQYA	SQHSGMIGYETDEN
3	SQHSGMIGY	AHAKRQTVVVLGSQE	FHDIPLPWHAGADTGTPHWNNKE
4	ELDPPFGDSY		WDFGSVGG
5	GLFGKGSLV		
6	TGLDFSDLYY		
7	GLDFSDLYY		
8	KSIQPENLEY		
9	SQHSGMIGY		
10	DTAWDFGSV		
11	TAAFTFTKV		

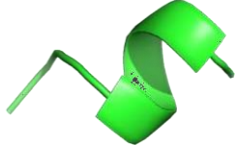
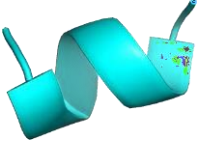
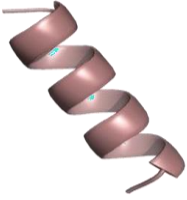

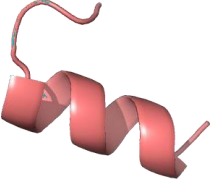
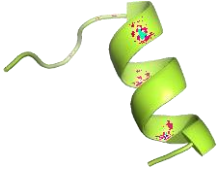
12	GSQHSGMIGY		
Membrane			
CTL (MHC-I)		HTL(MHC-II)	B CELL
1	SQHSGMIGY	NLEYRIMLSVHGSQH	GSQHSGMIGYETDEN
2	GLDFSDLYY		
3	GLFGKGSLV		
4	TGLDFSDLYY		
5	KSIQPENLEY		
6	GSQHSGMIGY		
7	SIQPENLEY		

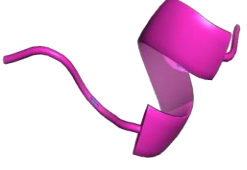




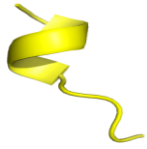

3.2.4) Three-dimensional structure of selected epitopes.



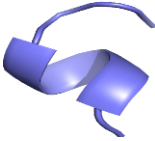

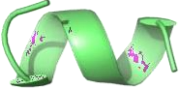
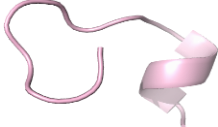


Three-dimensional structures were generated using the PyMOL tool to visualize the system and shape of the selected peptide structure.

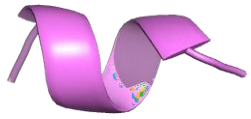

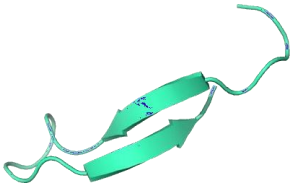
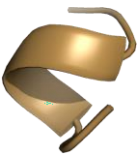
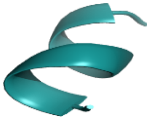


Table 3.14: 3D structure of the selected epitopes.



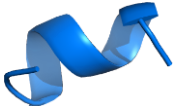


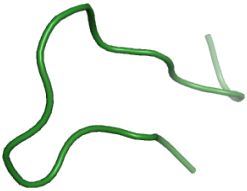
Type of Proteins	Name of epitopes	Sequence	Length	Figure
Capsid	MHC-I	TVSNMAEVR	9	

		LLGSSTSQK	9	
		MAEVRSYCY	9	
	MHC-II	VRSYCYEASISDMAS	15	
		EVRSYCYEASISDMA	15	
		IFRNPGFALVAVAIA	15	
	B cell	RCIGVSNRDFVEGMSG	16	

Envelope	MHC-I	SIQPENLEY	9	
		KSLFGGMSW	9	
		SQHSGMIGY	9	
		ELDPPFGDSY	10	
		GLFGKGS LV	9	
		TGLDFSDLYY	10	
		GLDFSDLYY	9	

		KSIQPENLEY	10	
		SQHSGMIGY	9	
		DTAWDFGSV	9	
		TAAFTFTKV	9	
		GSQHSGMIGY	10	
	MHC-II	HAKRQTVVVLGSQEG	15	
		AETLHGTVTVEVQYA	15	
		AHAKRQTVVVLGSQE	15	

	Bcell	GVSNRDFV	8	
		SQHSGMIGYETDEN	14	
		FHDIPLPWHAGADTGTP HWNNKE	23	
		WDFGSVGG	8	
		SQHSGMIGY	9	
		GLDFSDLYY	9	
		GLFGKGSLV	9	

Membrane	MHC-I	TGLDFSDLYY	10	
		KSIQPENLEY	10	
		GSQHSGMIGY	10	
		SIQPENLEY	9	
	MHC-II	NLEYRIMLSVHGSQH	15	
B cell	GSQHSGMIGYETDEN	15		

3.3) Result Population Coverage

Population coverage was done for the MHC-I and MHC-II epitopes to visualize the effect of the consensus epitopes. It is the most important analysis to construct a multi-epitope vaccine. Population coverage is a crucial part of the analysis that uses the HLA genotype frequencies for specific populations that predict the highest and lowest coverage.

3.3.1) The MHC-I tables' statistics

Table 3.: MHC-I population coverage

population /location	Coverage	Average hit	pc 90
World	88.09%	5.07	0.84
South America	63.13%	2.4	0.27
North America	89.18%	4.73	0.92
Central America	5.69%	0.18	0.11
East Africa	78.94%	4.4	0.47
North Africa	81.44%	4.15	0.54
Central Africa	71.11%	2.78	0.35
South Africa	75.27%	4.36	0.4
West Africa	79.41%	3.65	0.49
Southeast Asia	60.72%	3.35	0.25
Northeast Asia	61.18%	3.28	0.26
East Asia	63.13%	2.67	0.27
South Asia	68.03%	4.05	0.31
Europe	95.65%	6.57	1.44
West indies	91.28%	5.26	1.08
Oceania	53.21%	2.68	0.21

MHC-I population coverage

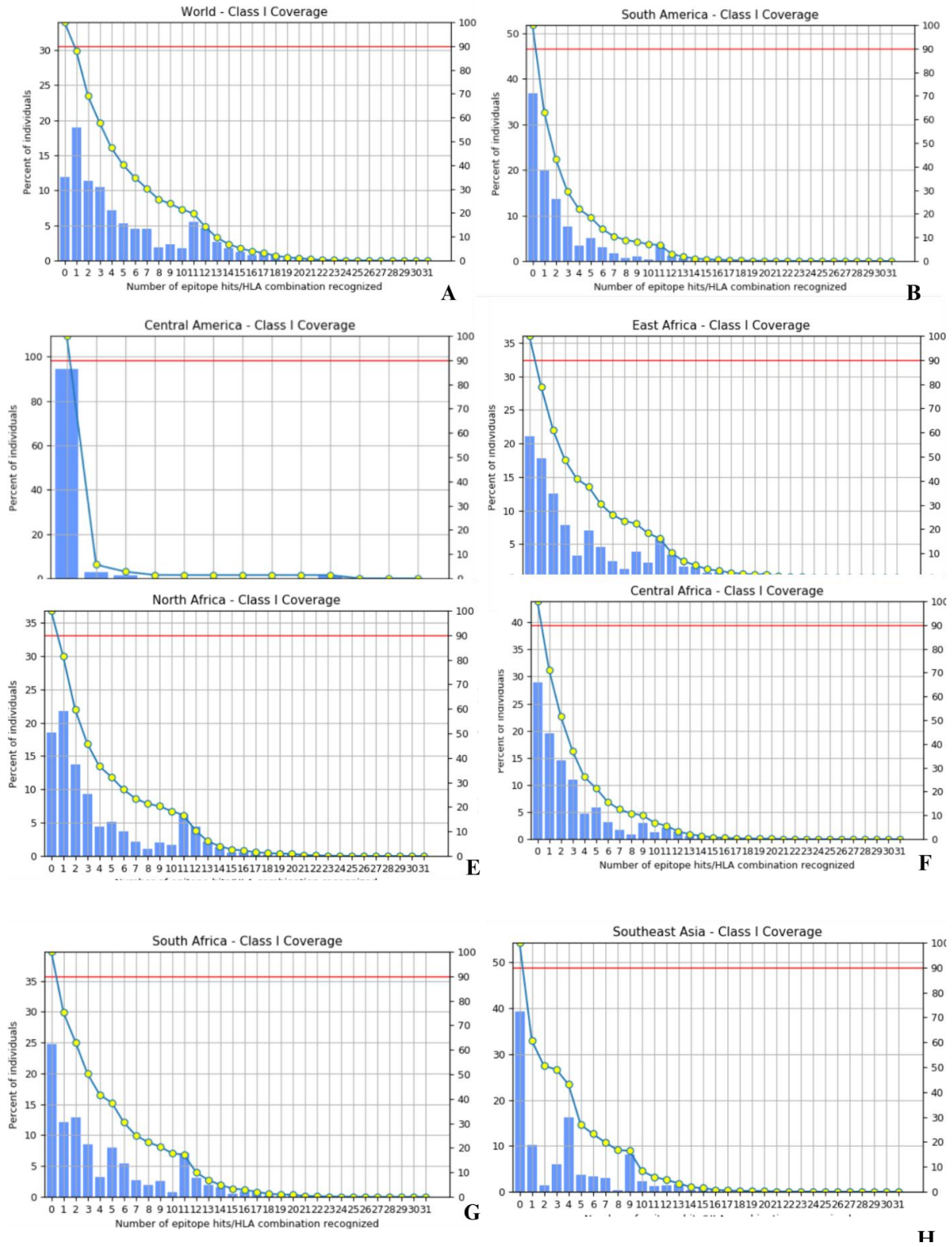


Figure 3.2: MHC-I population coverage-1

MHC-I population coverage

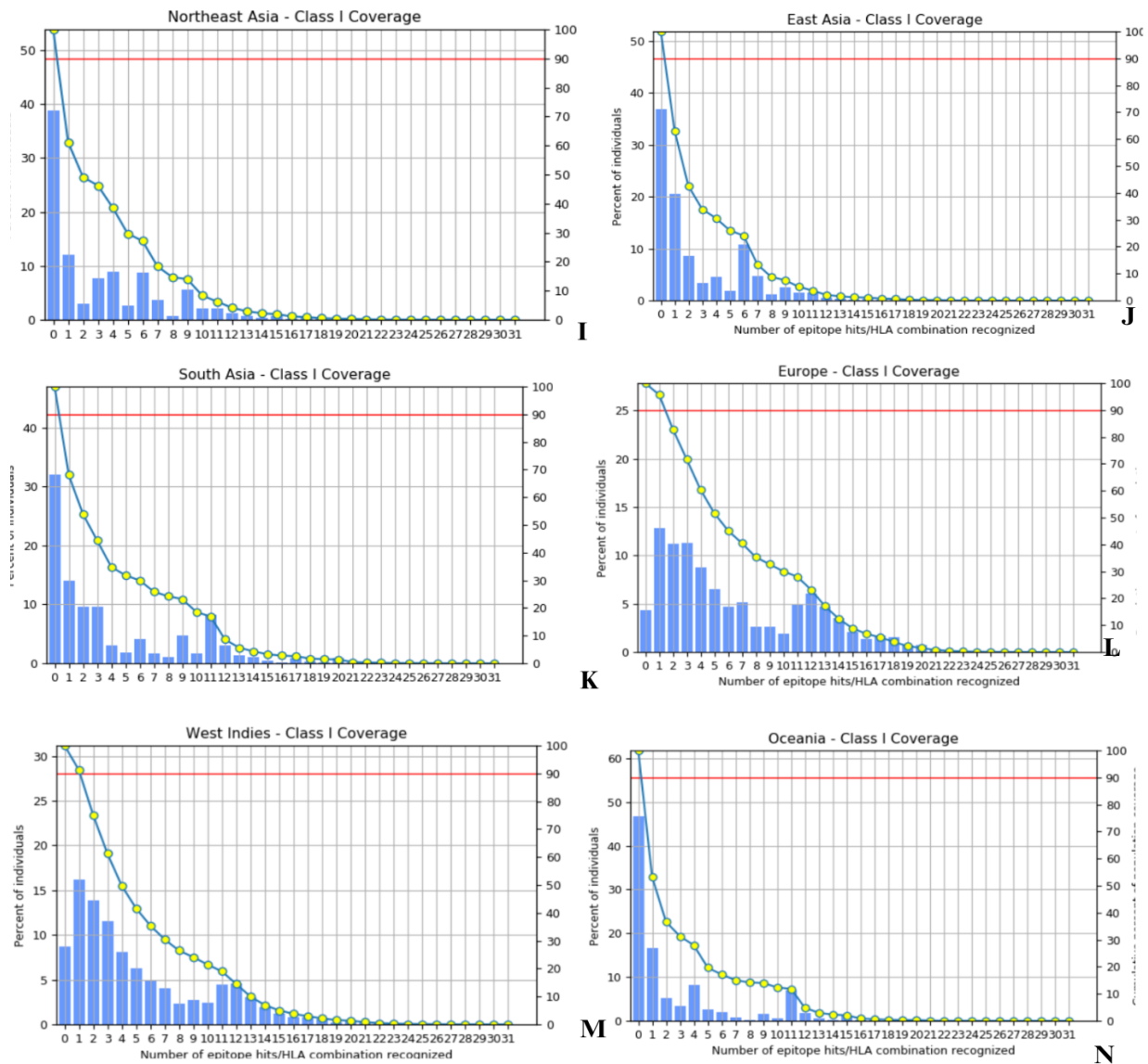


Figure 3.3: MHC-I population coverage- 2

The numbers show that the population spread, average hit rates, and PC90 values differ in distinct parts of the world. Notably, the World has an 88.09% reach rate, with an average hit rate of 5.07 and a PC90 of 0.84; the next high coverage region is South America, with a spread of 63.13%, an average hit rate of 2.4 and a PC90 of 0.27. Central America, however, has the lowest coverage, at 5.69%, with an average hit of 0.18 and PC90 of 0.11. This shows that there are significant problems

with usability in that area. It covers 78.94% of the continent, with an average hit rate of 4.4 and a PC90 of 0.47. North Africa covers 81.44% of the continent, with an average hit rate of 4.15 and a PC90 of 0.54. Northeast Asia has a coverage rate of 61.18%, an average hit rate of 3.28, and a PC90 of 0.26, which means it is moderately accessible. East Asia and Southeast Asia have coverage rates of 63.13%, with average hits of 2.67 and 3.35 and PC90 values of 0.27 and 0.25, respectively.

The coverage rate in South Asia is 68.03%, the average hit rate is 4.05, and the PC90 is 0.31. The coverage rate in West Africa is 79.41%, the average hit rate is 3.65, and the PC90 is 0.49. With an average hit rate of 4.73 and a PC90 of 0.92, North America has one of the highest coverage rates at 89.18%. This shows that much work has been put into making things accessible. Europe has the best coverage, with an impressive 95.65%, an average hit rate of 6.57, and a PC90 of 1.44, showing that they take mobility very seriously. The coverage in Central Africa is 71.11%, the average hit is 2.78, and the PC90 is 0.35. The coverage in South Africa is 75.27%, the average hit is 4.36, and the PC90 is 0.4. West Indies has a coverage rate of 91.28%, an average hit rate of 5.26, and a PC90 of 1.08; Oceania has a coverage rate of 53.21%, an average hit rate of 2.68, and a PC90 of 0.21. When we look at the MHC-II sample, East Asia has 80.97% coverage, an average hit of 3.91, and a PC90 of 0.53. Northeast Asia has 94.69% coverage, an average hit of 4.05, and a PC90 of 2.02. With a coverage rate of 99.22%, an average hit rate of 5.57, and a PC90 of 2.57, Southeast Asia has a coverage rate of 70.09%, an average hit rate of 2.06, and a PC90 of 0.33, With a coverage rate of 99.9%, an average hit rate of 6.35, and a PC90 of 2.98, Europe has the highest coverage. Central Africa has a coverage rate of 97.21%, an average hit rate of 4.64, and a PC90 of 2.32. Southwest Asia has a coverage rate of 68.48%, an average hit rate of 1.92, and a PC90 of 0.32. The West Indies have an average hit rate of 3.37, a coverage rate of 84.04%, and a PC90 of 0.63. North Africa has an average hit rate of 81.9% on the year. With an average hit rate of 6.07 and a PC90 of 3.09, North America has one of the highest coverage rates at 99.99%. This shows that a lot of work has gone into making things accessible. The coverage in Central America is 84.31%, the average hit is 2.51, and the PC90 is 0.64. The coverage in South America is 96.79%, the average hit is 4.61, and the PC90 is 2.19. A PC90 of 2.08 and a coverage rate of 96.07% are seen in Oceania.

3.3.2) MHC-II tables' statistics

Table 3. 15: MHC-II population coverage.

Population/Location	Coverage	Average hit	PC90
World	84.11%	3.35	0.63
South America	96.79%	4.61	2.19
North America	99.99%	6.07	3.09
Central America	84.31%	2.51	0.64
East Africa	97.77%	4.94	2.41
North Africa	81.9%	3.02	0.55
Central Africa	97.21%	4.64	2.32
West Africa	99.19%	5.32	2.9
Southeast Asia	70.09%	2.06	0.33
East Asia	97.77%	4.94	2.41
Northeast Asia	94.69%	4.05	2.02
South Asia	99.22%	5.57	2.57
Southwest Asia	68.48%	1.92	0.32
Europe	99.9%	6.35	2.98
West Indies	84.04%	3.37	0.63
Oceania	96.07%	4.25	2.08

MHC-II population coverage

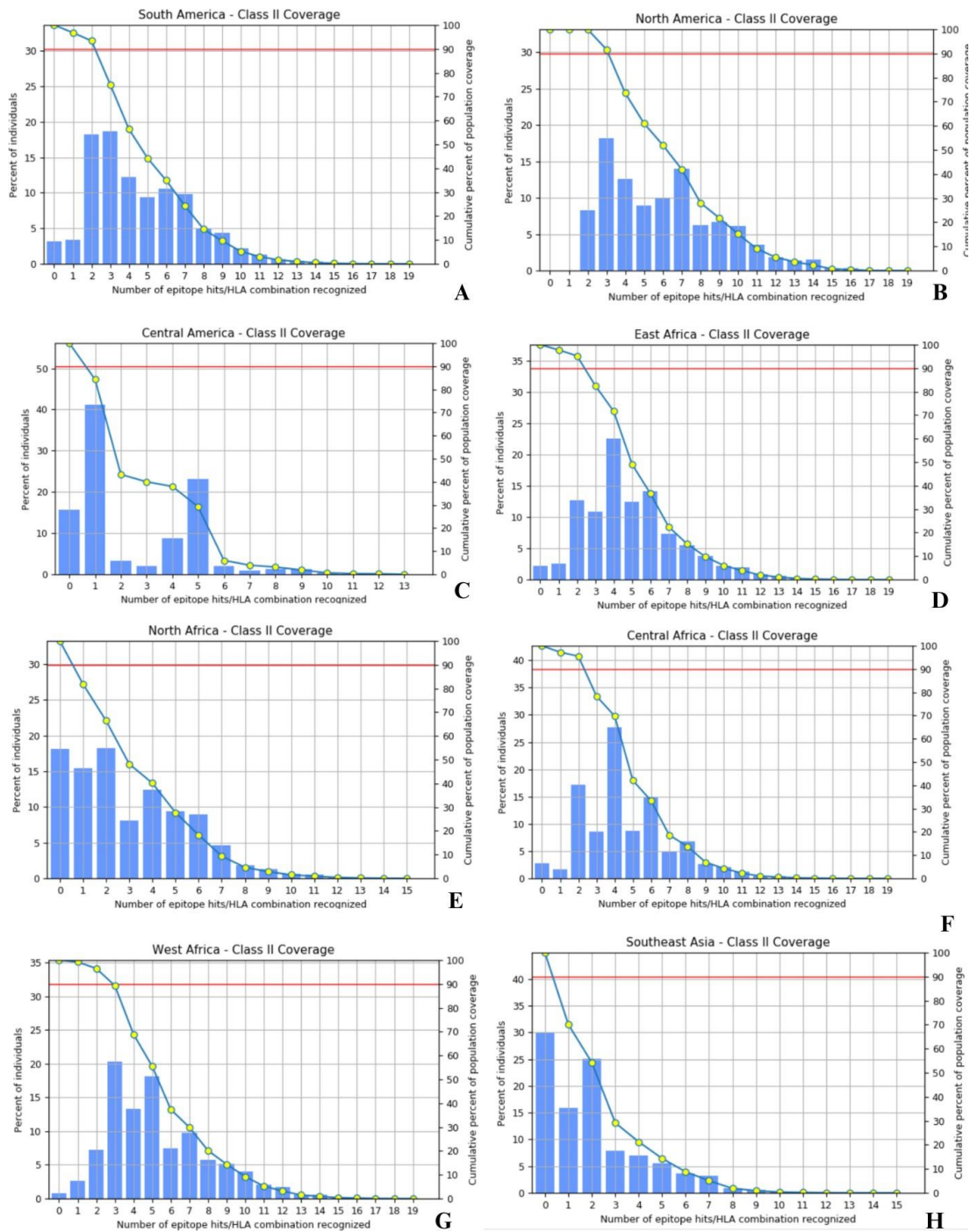


Figure 3. 4: MHC-II population coverage-1

MHC-II population coverage

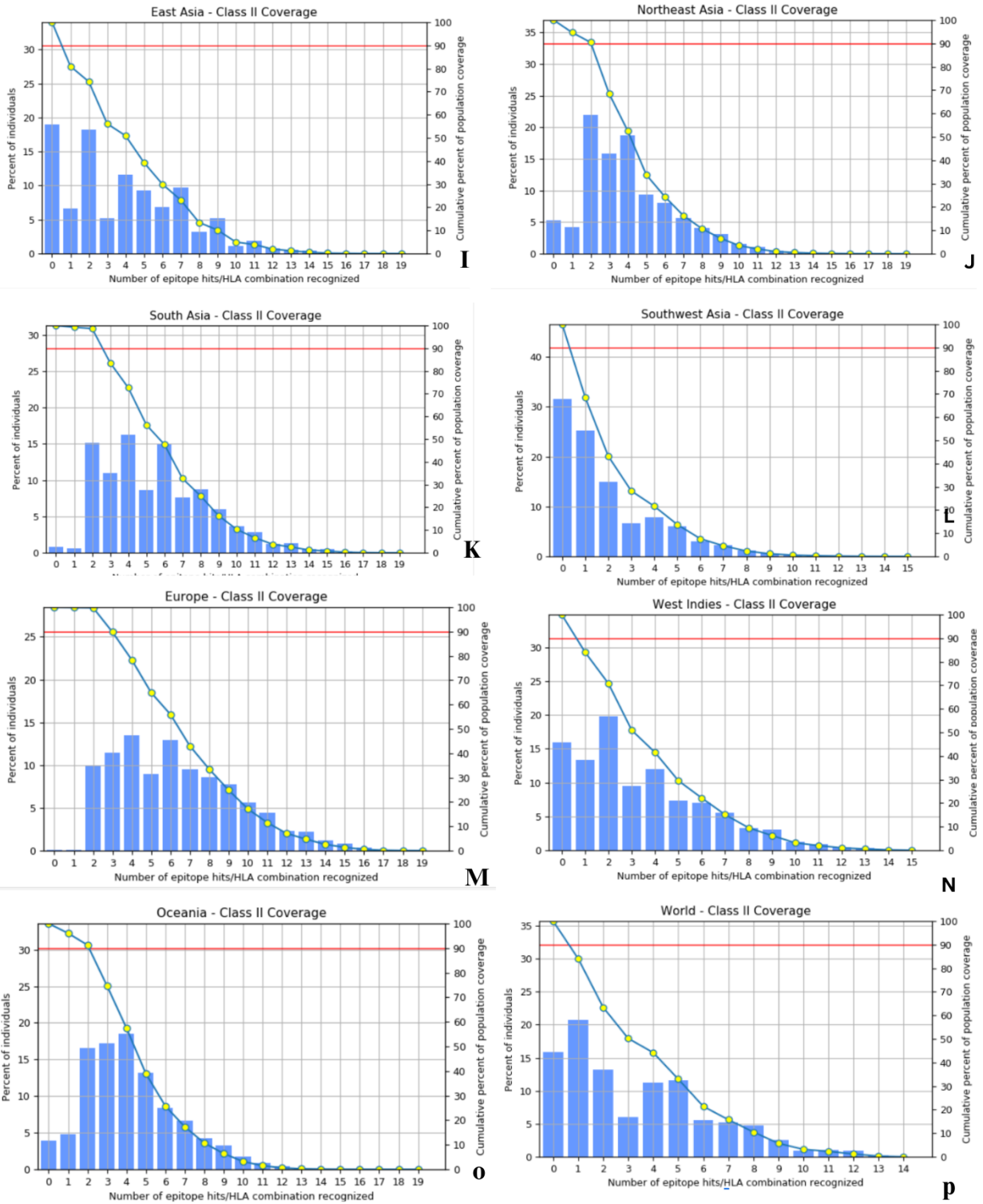


Figure 3. 5: MHC-II population coverage-2

The data table represented population coverage, average hit rates, and PC90 numbers in various parts of the world. Notably, South Asia has the best coverage, at 99.22%, with a high average hit rate of 5.57 and a PC90 value of 2.57. Southeast Asia, on the other hand, has a coverage rate of only 70.09%, an average hit rate of 2.06%, and a PC90 value of 0.33. Europe has excellent coverage at 99.9%, with a strong average hit rate of 6.35 and a PC90 score of 2.98. More research shows North America has almost complete coverage (99.99%), an average hit rate of 6.07, and a PC90 value of 3.09. Central America, on the other hand, only gets 84.31% covered, with an average hit rate of 2.51 and a PC90 value of 0.64.

3.4) Vaccine Formulation

The vaccines were constructed using the specific linkers, adjuvants, and epitope sequences from the selected epitopes. Nineteen vaccines were constructed with the selected epitope that has been analyzed before.

3.4.1) Primary construction of 19 vaccines

The secondary structure was predicted using the NetSurf-3.0 server tool to observe the amino acid sequences and the arrangements of the residues. It was done to predict the structure for the function, local conformation, and further analysis. Later, with the help of Pymol, tertiary structures were generated. All 19 different vaccines have primary differences like separate adjuvants, different linkers, and discrete protein sequences. Based on the other physicochemical properties, structure analysis, molecular docking, and in-depth simulation, a suitable candidate was chosen.

- To add linkers with the epitope PADRE sequences were used.
- EAAAK was used to link the adjuvants with the rest sequence.
- GGGGS and AYY both were used as CTL linkers.
- GPGPG was used as HTL epitopes linker.
- KK was used as a B cell epitope linker.
- Histidine (HHHHHH) was tagged.

3.4.2) physiological properties of 19 selected vaccines.

Constructed 19 vaccines were evaluated for the antigenic value by Vaxijen server and all showed to be antigenic. The values for each vaccine were respectively 0.8492, 0.7877, 0.8169, 0.7668, 0.7829, 0.8157, 0.7781, 0.7628, 0.7702, 0.6699, 0.7041, 0.7528, 0.7260, 0.6738, 0.7953, 0.8427, 0.7595, 0.9075, 0.9101 showed in the table with other values as the physiological properties. All the vaccine constructs showed stability along with that, molecular weight, solubility, GRAVY, and length are also shown. No significant hits were observed, and the threshold E-value was 0.001 which means all of them were non-allergenic. Moreover, they are highly soluble and have more than 0.45 threshold. These values are 0.537, 0.456, 0.566, 0.606, 0.629, 0.479, 0.581, 0.549, 0.531, 0.635, 0.505, 0.583, 0.571, 0.590, 0.584, 0.482, 0.509, 0.490, 0.479.

Table 3. 16: Physiological properties of 19 selected vaccines.

No	Antigenicity	Allergenicity	Solubility	Length	Grand average of hydrophobicity (GRAVY)	Molecular weight
1	Antigenic	non-allergenic	Soluble	274	-0.259	29649.54
2	Antigenic	non-allergenic	Soluble	448	-0.249	47407.36
3	Antigenic	non-allergenic	Soluble	468	-0.363	48097.51
4	Antigenic	non-allergenic	Soluble	350	-0.011	36783.54
5	Antigenic	non-allergenic	Soluble	368	-0.126	36873.11
6	Antigenic	non-allergenic	Soluble	452	-0.292	48017.89
7	Antigenic	non-allergenic	Soluble	454	-0.360	46821.16
8	Antigenic	non-allergenic	Soluble	356	-0.065	37606.39
9	Antigenic	non-allergenic	Soluble	351	-0.185	36844.66
10	Antigenic	non-allergenic	Soluble	316	-0.297	33378.15
11	Antigenic	non-allergenic	Soluble	326	-0.366	34806.62
12	Antigenic	non-allergenic	Soluble	310	-0.295	33110.77
13	Antigenic	non-allergenic	Soluble	322	-0.394	34185.85

14	Antigenic	non-allergenic	Soluble	311	-0.305	32986.53
15	Antigenic	non-allergenic	Soluble	311	-0.348	32886.39
16	Antigenic	non-allergenic	Soluble	316	-0.437	34065.62
17	Antigenic	non-allergenic	Soluble	318	-0.453	34461.11
18	Antigenic	non-allergenic	Soluble	310	-0.449	33501.88
19	Antigenic	non-allergenic	Soluble	310	-0.428	33616.00

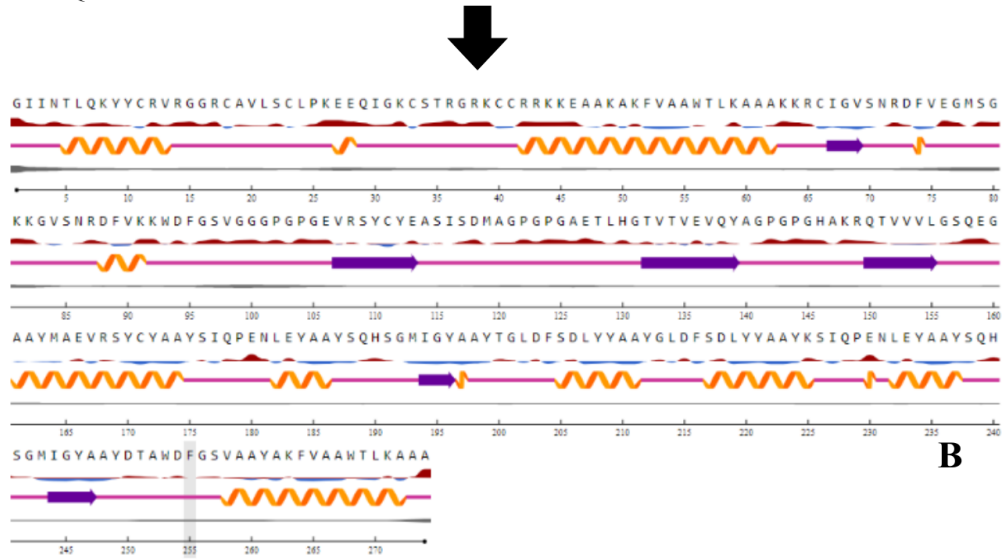
3.4.3) Secondary Structure Prediction of Vaccine Constructs

For the comparative analysis for each vaccine constructed secondary structure was analyzed to see the Helix, coil, disorders, and sheets. For each vaccine construction a 3D image was generated to inspect in depth study of the structure. Moreover, each structure gave a clear representation of the relative surface accessibility and disorders. Based on the structure and the comparative analysis of the physiochemical values of the sequences, 10 best candidates were chosen for further analysis. The primary, secondary and tertiary structure of 19 constructed vaccines is shown below.

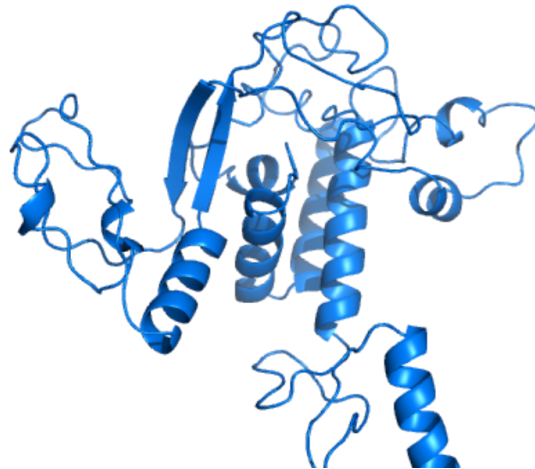
Vaccine -01

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCSTRGRKCCRRKKEAAKAKFVAAWTLK
AAAKKRCIGVSNRDFVEGMSGKKGVSNRDFVKKWDFGVSVGGGPGPGEVRSYCYEASISD
MAGPGPGAETLHGTVTVEVQYAGPGPGHAKRQTVVVLGSQEGAAAYMAEVRSYCYAAYSI
QPENLEYAAYSQHSGMIGYAAYTGLDFSDLYYAAAGLDFSDLYYAAAYKSIQPENLEYAAY
SQHSGMIGYAAYDTAWDFGSVAAYAKFVAAWTLKAAA

A



B



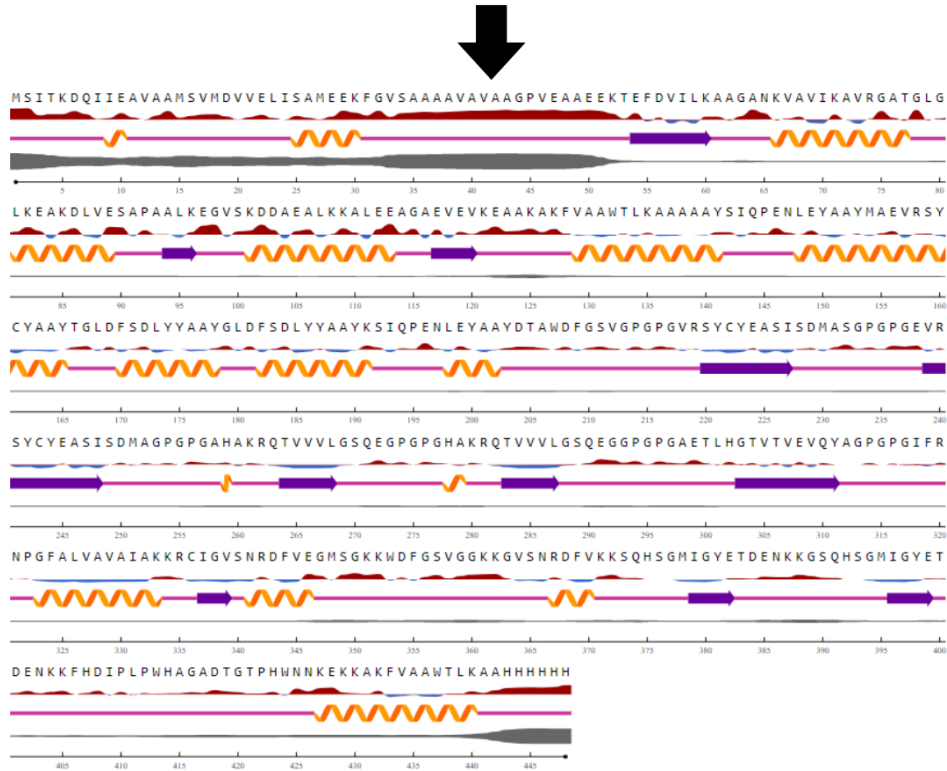
C

Figure 3. 6: Structure of vaccine 1. (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -02

MSITKDQIEEVAAMSVMDEVVELISAMEEKFVGSAAA AVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAAAYSIQPENLEYAAYMAEVRSYCYAAYTGLDFSDLYYAAAY
GLDFSDLYYAAAYKSIQPENLEYAAYDTAWDFGSGVGGPGVRSYCYEASISDMASGPGGPE
VRSYCYEASISDMAGPGPGAHAKRQTVVVLGSEQEGPGPHAKRQTVVVLGSEQEGGPGPGA
ETLHGTVTVEVQYAGPGPGIFRNPGFALVAVAIKKRCIGVSNRDFVEGMSGKKWDFGVS
GGKKGVSNRDFVKKSQHSGMIGYETDENKKGSQHSGMIGYETDENKKFHDIPLPHAG
ADTGTPHWNNKEKKAKFVAAWTLKAAHHHHHH

A



B

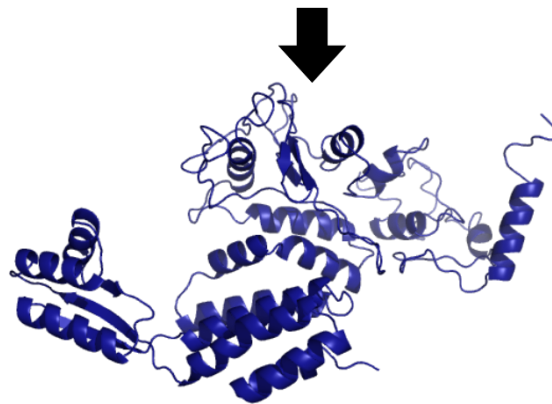


Figure 3. 7: Structure of vaccine 2 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -03

MSITKDQIIEAVAAMSVMDEVVELISAMEEKEFGVSAVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAGGGGSMAEVRSYCYGGGGSSIQPENLEYGGGGSSQHSGMIG
YGGGGSTGLDFSDLYYGGGGGLDFSDLYYGGGGSKSIQPENLEYGGGGSSQHSGMIGY
GGGGSDTAWDFGSVGGPGGEVRSYCYEASISDMAGPGPGAETLHGTVTVEVQYAGPGPG
HAKRQTVVVLGSQEGGPGGAHAKRQTVVVLGSQEGGPGGVRSYCYEASISDMAKKWDF
GSVGGKKRCIGVSNRDFVEGMSGKKGVSNRDFVKKSQHSGMIGYETDENKKFHDIPLPW
HAGADTGTPHWNNKEKKGSGHSQHSGMIGYETDENKKAKFVAAWTLKAAAHHHHHH **A**

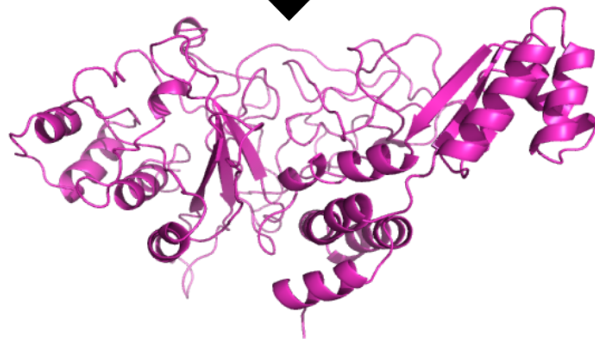
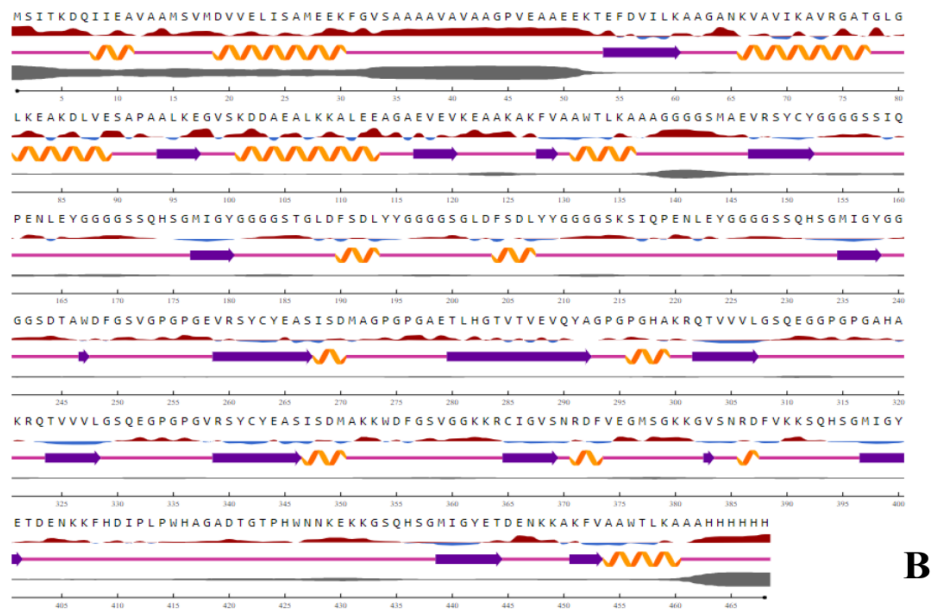


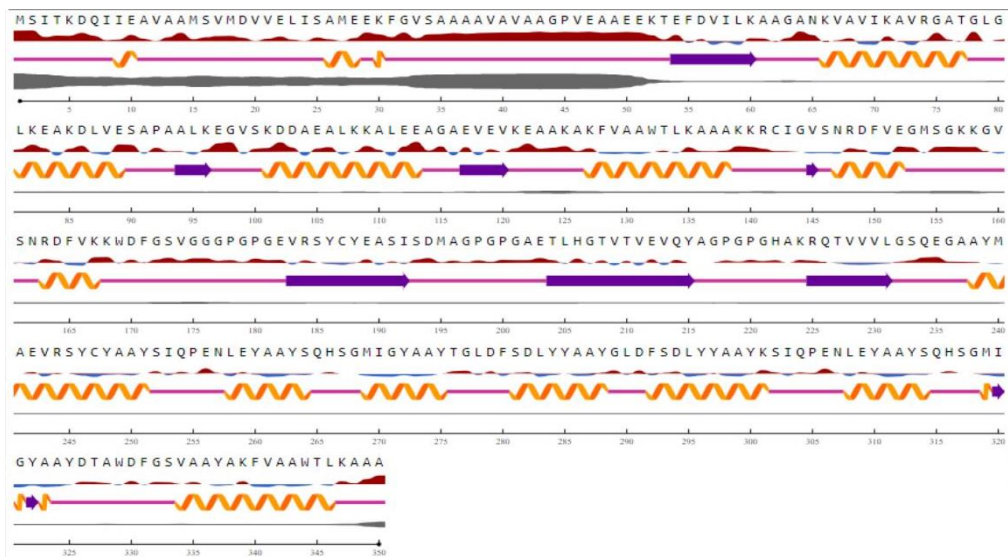
Figure 3. 8: Structure of vaccine 3 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -04

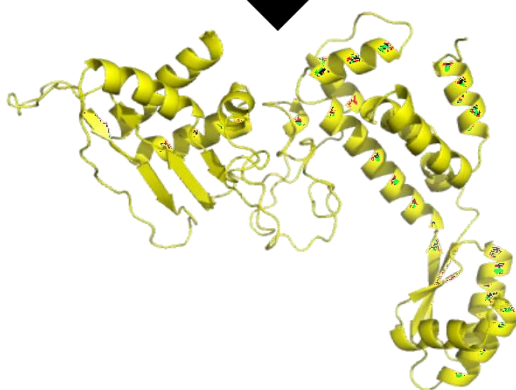
```
MSITKDQIIEAVAAMSVM DVVELISAMEEKFGVSA AAAVAVAAGPVEAAEEKTEFDVILKA  
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV  
KEAAKAKFVAAWTLKAAAKKRCIGVSNRDFVEGMSGKKGVSNRDFVKKWDFGSVGGG  
PGPGEVRSYCYEASISDMAGPGGAETLHGTVTVEVQYAGPGPGHAKRQTVVVLGSQEGA  
AYMAEVRSYCYAAYSIQPENLEYAAYSQHS GMIGYAA YTG LDFSDLYYAAYGLDFSDLYY  
AAYKSIQPENLEYAAYSQHS GMIGYAA YDTAWDFGSVAAYAKFVAAWTLKAAA
```



A



B



C

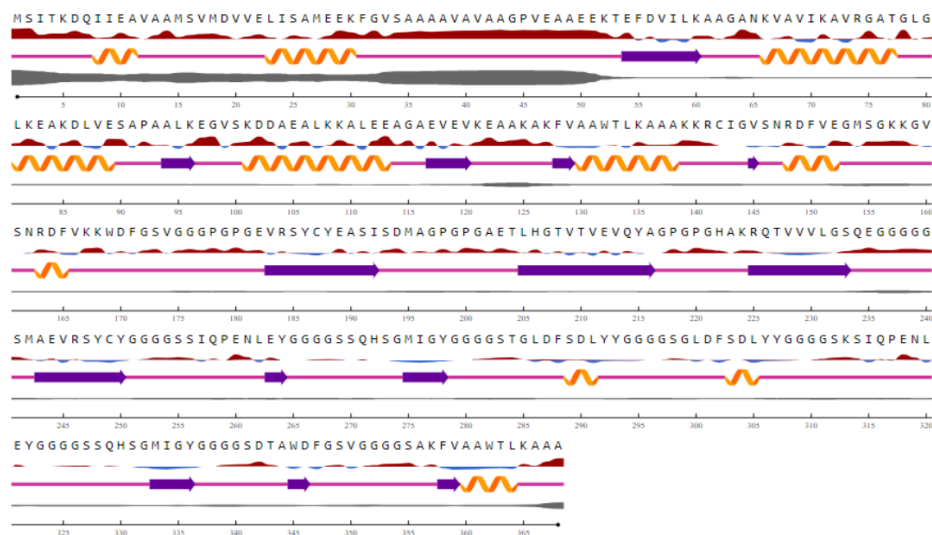
Figure 3. 9: Structure of vaccine 4 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -05

MSITKDQIIEAVAAMSVMDEVVELISAMEEKFVGSAAAAVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAKKRCIGVSNRDFVEGMSGKKGVSNRDFVKKWDFGSVGGG
PGPGEVRSYCYEASISDMAGPGAETLHGTVTVEVQYAGPPGHAKRQTVVVLGSQEGG
GGGSMAEVRSYCYGGGSSIQPENLEYGSGSSQHSGMIGYGGGGSTGLDFSDLYYGGG
GSLDFSDLYYGGGGSKSIQPENLEYGSGSSQHSGMIGYGGGGSDTAWDFGSVGGGG
AKFVAAWTLKAAA



A



B

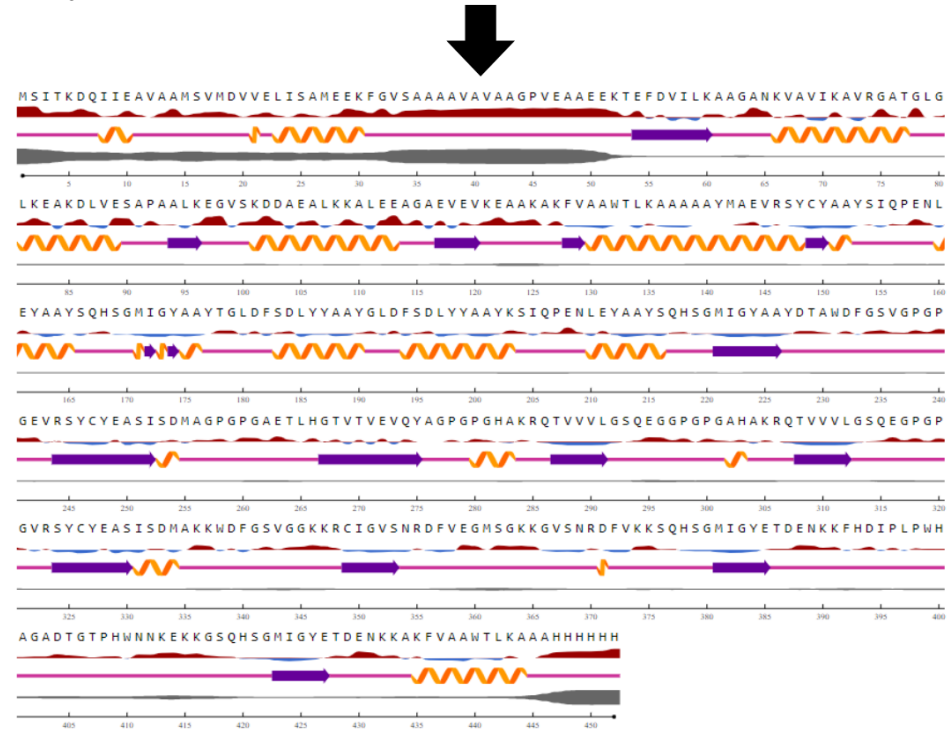


C

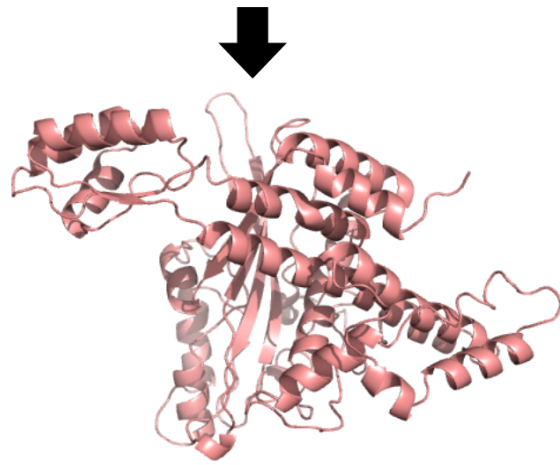
Figure 3.10: Structure of vaccine 5 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

MSITKDQIIEAVAAMSVM DVVELISAMEEKFGVSAAAAVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAAAYMAEVRSYCYAAYSIQPENLEYAAYSQHS GMIGYAAAYTG
LDFSDLYYAA YGLDFSDLYYAA YKSIQPENLEYAAYSQHS GMIGYAAAYDTAWDFGSVGP
GPGEVRSYCYEASISDMAGPGPGAETLHGTVTVVEVQYAGPGPGHAKRQTVVVLGSQEGGP
GPGAHAKRQTVVVLGSQEGPGPGVRSYCYEASISDMAKKWDFGSVGGKKRCIGVSNRDF
VEGMSGKKGVSNRDFVKKSQHS GMIGYETDENKKFHDIPLPHAGADTGTPHWNNKEK
KGSQHS GMIGYETDENKKAKFVAAWTLKAAAHHHHHH

A



B



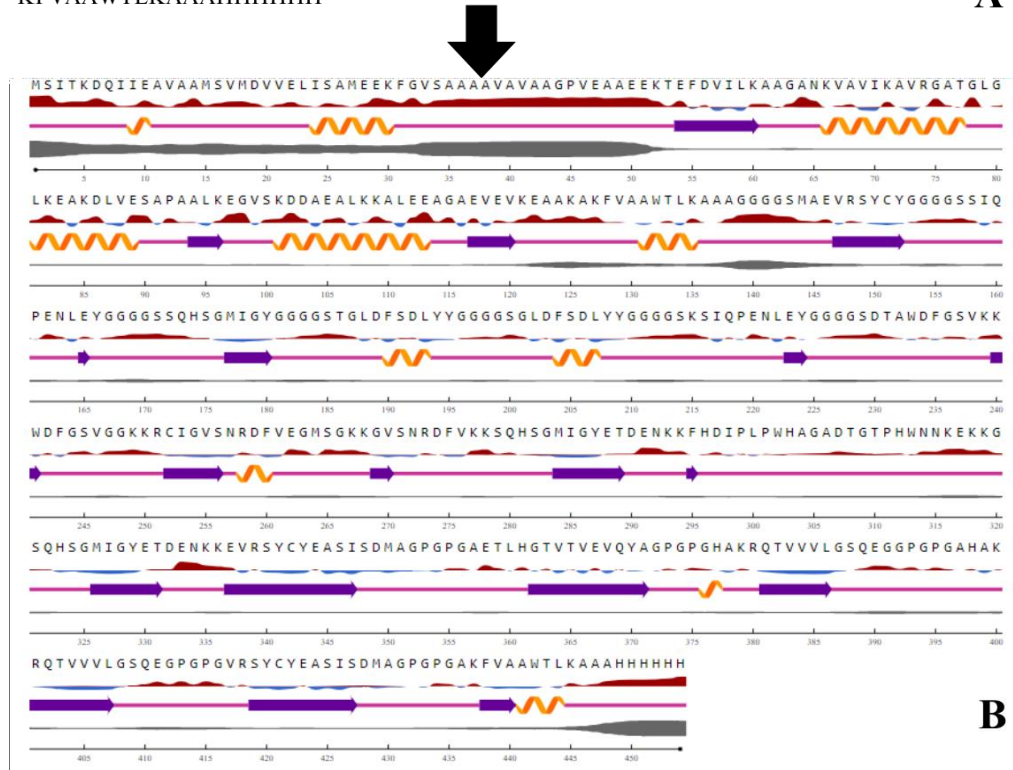
C

Figure 3. 11: Structure of vaccine 6 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

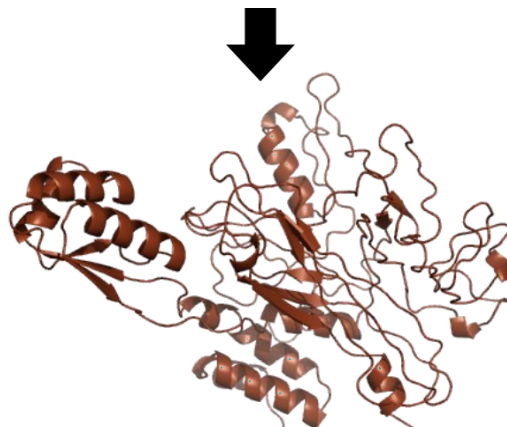
Vaccine -07

MSITKDQIIEAVAAMSVM DVVELISAMEEKFGVSA AAAVAVAAGPVEAAEEKTEFDVILKAAGA
NKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVS KD DAEALKKALEEAGAEVEVKEA AKA
KFVAAWTLKAAAGGGG SMAEVRSYCYGGGGSSIQPENLEYGGGGSSQHSGMIGYGGGGSTGL
DFSDLYYGGGGSLDFSDLYYGGGGSKSIQPENLEYGGGGSDTAWDFGVS VKK WDFGVS VGGK
KRCIGVSNRDFVEGMSGKKGVSNRDFVKKSQHSGMIGYETDENKKFHD IPLPHAGADTGTP
HWNNKEKKGSQHSGMIGYETDENKKEVRSYCYEASISDMAGPGPGAETLHGT VTVVEVQYAGP
GPGHAKRQTVVVLG SQEGGPGPGA HAKRQTVVVLG SQEGGPGPVRSYCYEASISDMAGPGPGA
KFVAAWTLKAAAH HHHHHH

A



B



C

Figure 3. 12: Structure of vaccine 7 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -08

MSITKDQIIEAVAAMSVMDVVELISAMEEKFVGSAAA AVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAKKRCIGVSNRDFVEGMSGKKGVSNRDFVKKWDFGSGGG
PGPGEVRSYCYEASISDMAGPGGAETLHGTVTVEVQYAGPGPGHAKRQTVVVLGSQEGA
AYMAEVRSYCYAAYS IQPENLEYAAYSQHS GMIGYAA YTG LDFSDLYYAAYGLDFSDLYY
AAYKSIQPENLEYAAYSQHS GMIGYAA YDTAWDFGSVAAYAKFVAAWTLKAAAHHHHH
H

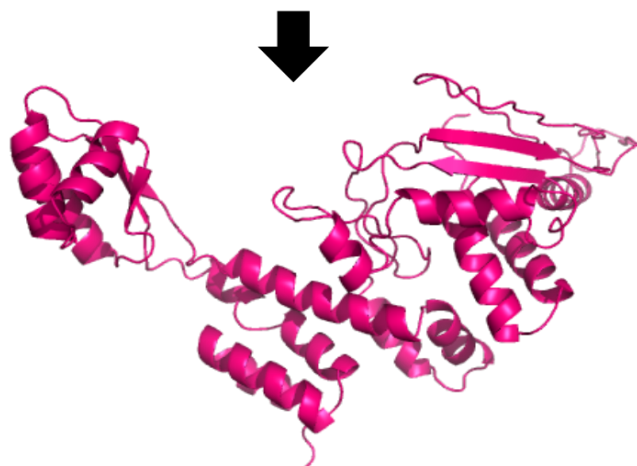
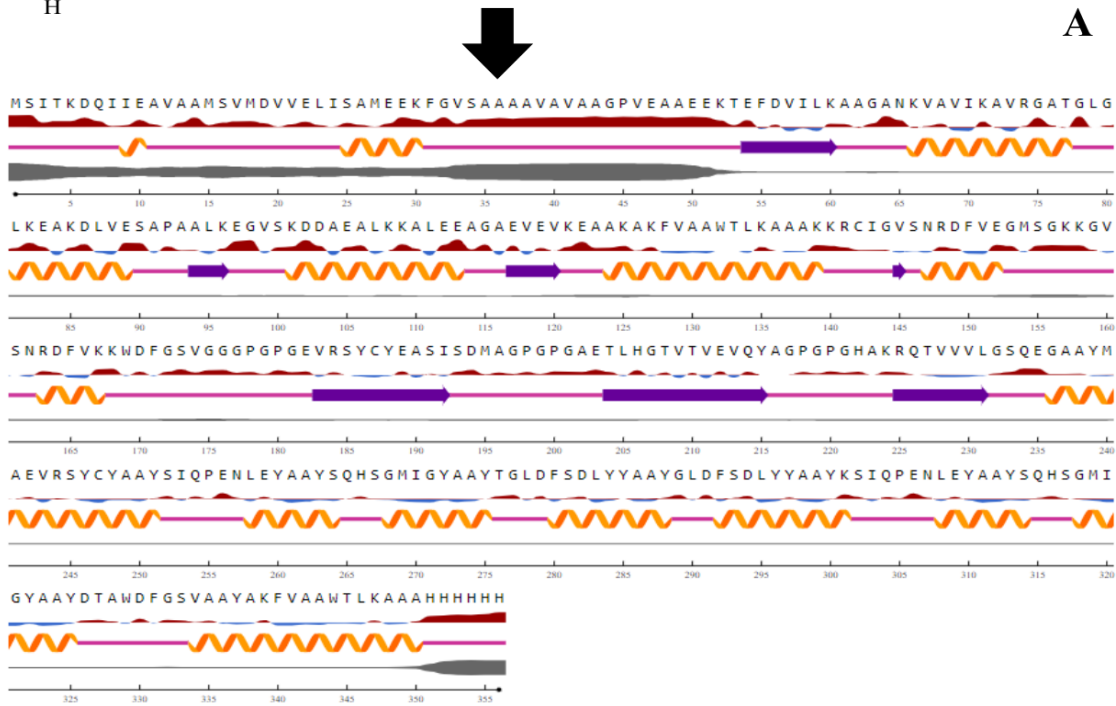


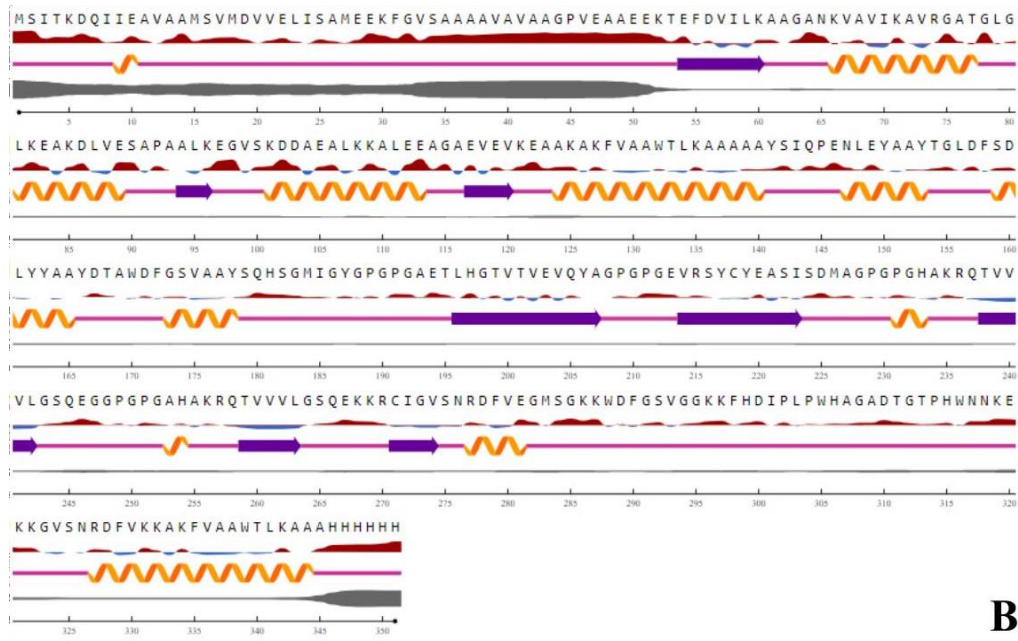
Figure 3. 13: Structure of vaccine 8 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -09

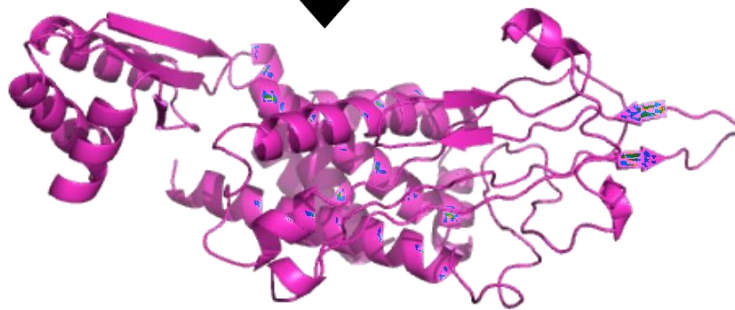
MSITKDQIIEAVAAMSVMDVVELISAMEEKFGVSA AAAAVAVAAGPVEAAEEKTEFDVILKA
AGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEV
KEAAKAKFVAAWTLKAAAAYSIQPENLEYAAYTGLDFSDLYAAAYDTAWDFGSVAAYS
QHSGMIGYGGPGAETLHGTVTVEVQYAGPGPGEVRSYCYEASISDMAGPGPGHAKRQTV
VVLGSQEGGPGGAHAKRQTVVVLGSQEKKRCIGVSNRDFVEGMSGKKWDFGSVGGKK
FHDIPLPHAGADTGTPHWNNKEKKGVSNRDFVKKAKFVAAWTLKAAAHHHHHH



A



B



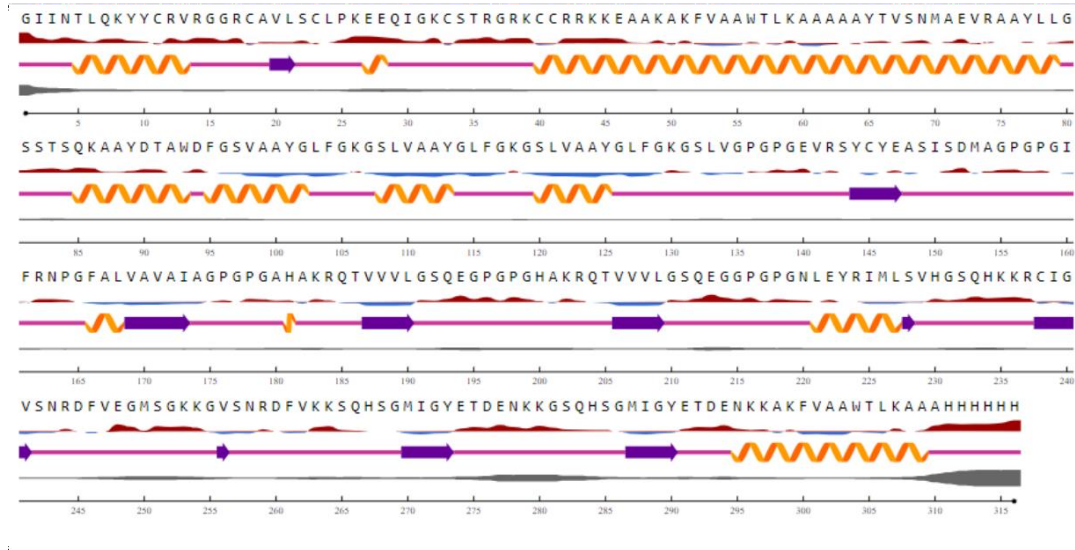
C

Figure 3. 14: Structure of vaccine 9 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

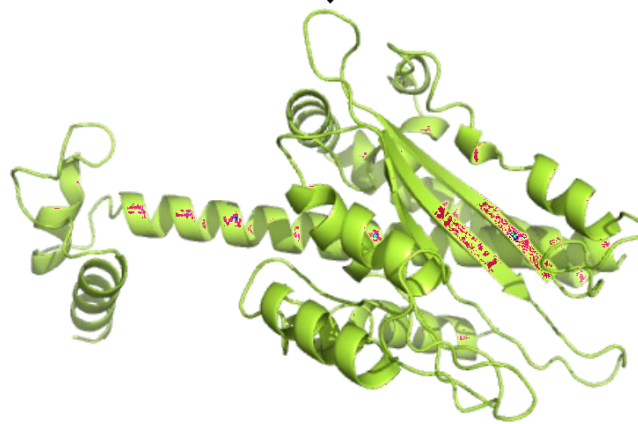
Vaccine -10

GIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCTRGRKCCRKKEAAKAKFVAAWTLK
AAAAAYTVSNMAEVRAAYLLGSSTSQAAYDTAWDFGSVAAAYGLFGKGSVAAAYGLFG
KGSVAAAYGLFGKGSVGPGEVRSYCYEASISDMAGPGPGIFRNPGFALVAVAIAGPGP
GAHAKRQTVVVLGSQEGPGPGHAKRQTVVVLGSQEGGPGPGNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKGVSNRDFVKKSQHSGMIGYETDENKKGSHGMIGYETDENK
KAKFVAAWTLKAAAHHHHHH

A



B



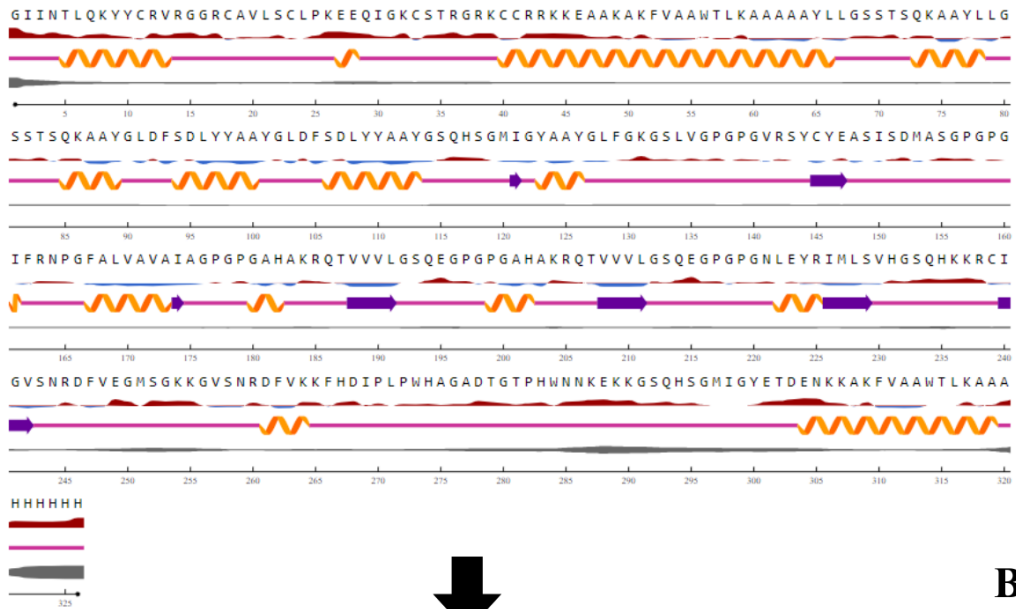
C

Figure 3. 15: Structure of vaccine 10 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

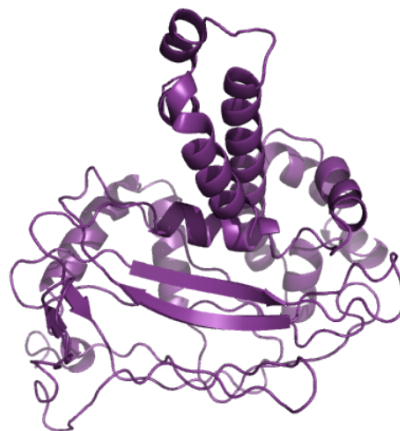
Vaccine -11

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAKAKFVAAWTLK
AAAAAYLLGSSTSQKAAAYLLGSSTSQKAAAYGLDFSDLYYAAAYGLDFSDLYYAAAYGSQHSG
MIGYAAAYGLFGKGSVLPVGGVRSYCYEASISDMASGPGPGIFRNPGFALVAVAIAGPGPG
AHAKRQTVVVLGSQEGPGPGAHAHRQTVVVLGSQEGPGPNLEYRIMLSVHGSQHKKRCI
GVSNRDFVEGMSGKKGVSNRDFVKKFHDIPLPHAGADTGTPHWNNKEKKGSQHSGMI
GYETDENKKAKFVAAWTLKAAAHHHHHH

A



B



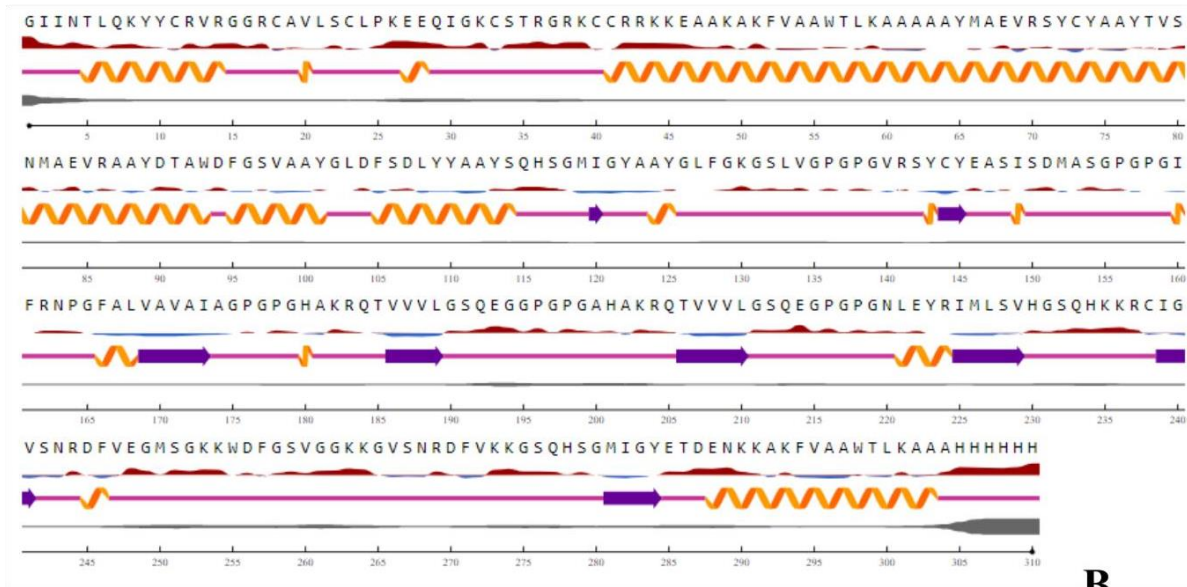
C

Figure 3. 16: Structure of vaccine 11 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

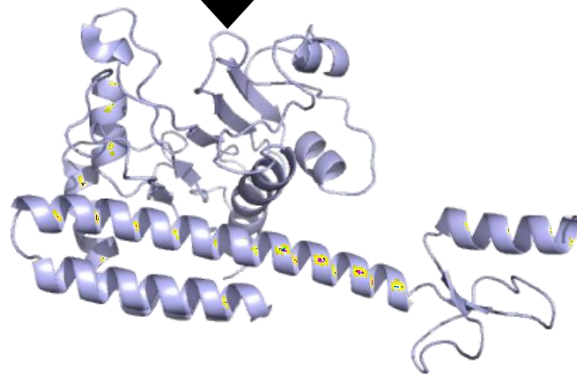
Vaccine -12

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCTRGRKCCRRKKEAAKAFVAAWTLK
AAAAAYMAEVRSYCYAAYTVSNMAEVRAAYDTAWDFGSVAAYGLDFSDLYYAAYSQHS
GMIGYAAAYGLFGKGSVLPVGGVRSYCYEASISDMASGPGPGIFRNPGFALVAVAIAGPGP
GHAKRQTVVVLGSQEGGPGGAHAKRQTVVVLGSQEGGPGNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKWDFGSVGGKKGVSNRDFVKKGSQHSGMIGYETDENKKAKFV
AAWTLKAAAHHHHHH

A



B



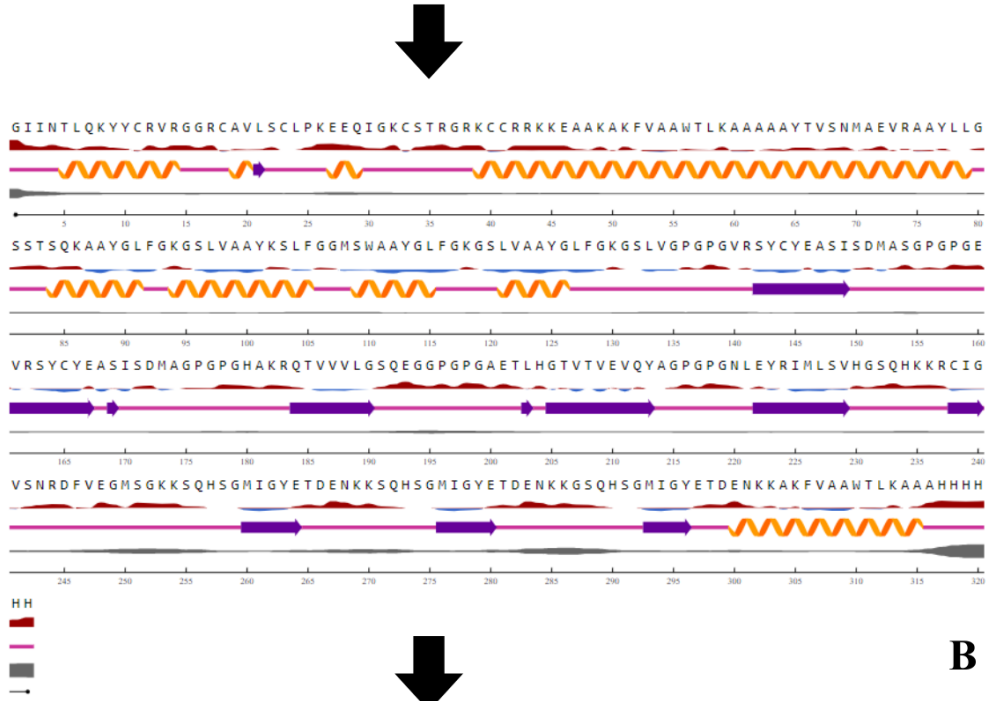
C

Figure 3. 17: Structure of vaccine 12 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -13

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCTRGRKCCRRKKEAAKAFVAAWTLK
AAAAAYTVSNMAEVRAAYLLGSSTSQKAAAYGLFGKGSLVAAAYKSLFGGMSWAAAYGLFG
KGSVAAAYGLFGKGSLVGPGPGVRSYCYEASISDMASGPGPGEVRSYCYEASISDMAGPGP
GHAKRQTVVVLGSQEGGPGPGAETLHGTVTVEVQYAGPGPGNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKSQHSGMIGYETDENKKSQHSGMIGYETDENKKSQHSGMIGY
ETDENKKAKFVAAWTLKAAAHHHHHH

A



B

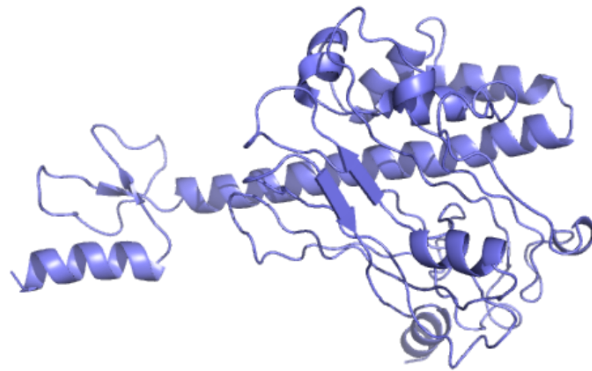
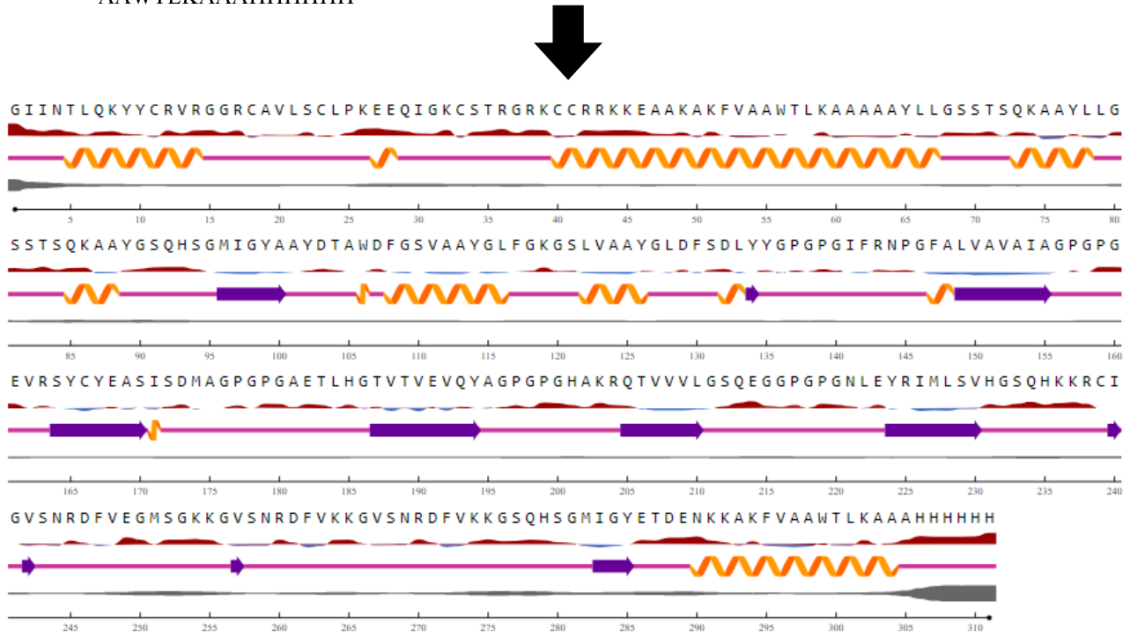


Figure 3. 18 : Structure of vaccine 13 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

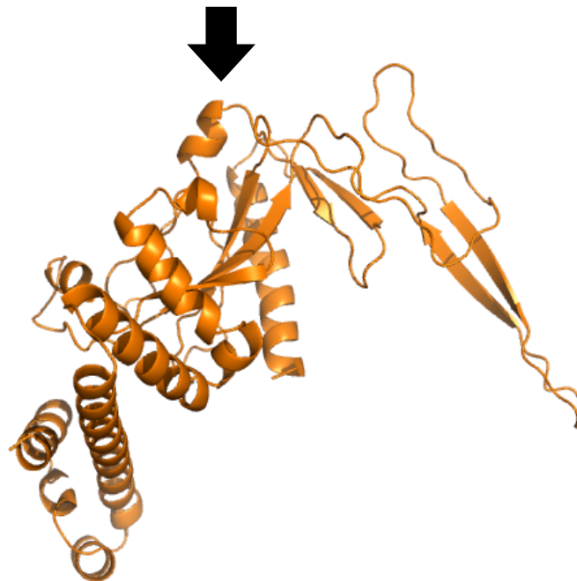
Vaccine -14

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCSTRGRKCCRRKKEAAKAKFVAAWTLK
AAAAAYLLGSSTSQKAAYLLGSSTSQKAAYGSQHSQMIGYAAAYDTAWDFGSVAAYGLFG
KGS LVAAYGLDFSDLYYGGPGGIFRNPGFALVAVAIAGPGGPEVRSYCYEASISDMAGPGP
GAETLHGTVTVEVQYAGPGPGHAKRQTVVVLGSQEGGPGPNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKGVS NRDFVKKGVS NRDFVKKGVS QHSMIGYETDENKKAKFV
AAWTLKAAAHHHHHH

A



B



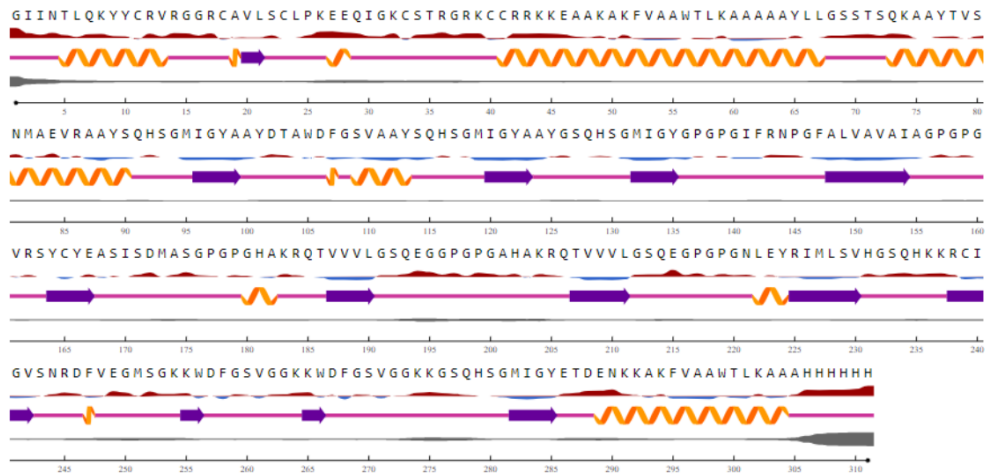
C

Figure 3. 19 : Structure of vaccine 14 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -15

GIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAKAKFVAAWTLK
AAAAAYLLGSSTSQKAAAYTVSNMAEVRAAAYSQHSGMIGYAAAYDTAWDFGSVAAYSQHSG
MIGYAAYSQHSGMIGYGPPGIFRNPGFALVAVAIAGPPGVRSYCYEASISDMASGPPG
GHAKRQTVVVLGSQEGGPPGGAHAKRQTVVVLGSQEGPPGNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKWDFGSVGGKKWDFGSVGGKKGSHSGMIGYETDENKKAKF
VAAWTLKAAAHHHHHH

A



B

C

Figure 3. 20: Structure of vaccine 15 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -16

GIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCTRGRKCCRRKKEAAKAFVAAWTLK
AAAAAYLLGSSTSQKAAYMAEVRSYCYAAYSQHSKMIGYAAAYGLDFSDLYYAAYSIQPEN
LEYAAYSIQPENLEYGPGPGIFRNPGFALVAVAIAGPGPGEVRSYCYEASISDMAGPGPGAH
AKRQTVVVLGSQEGPGGAHAKRQTVVVLGSQEGPGPNLEYRIMLSVHGSQHKKRCIGV
SNRDFVEGMSGKKSQHSKMIGYETDENKKWDFGVSQGGKKSQHSKMIGYETDENKKAK FVAAWTLKAAAHHHHHH

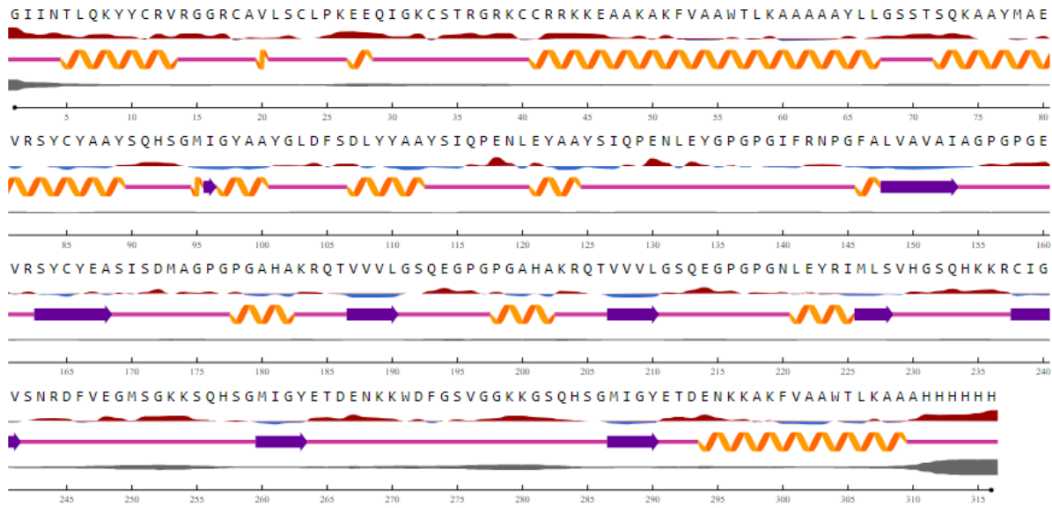


Figure 3. 21: Structure of vaccine 16 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -17

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCTRGRKCCRRKKEAAKAKFVAAWTLK
AAAAAYTVSNMAEVRAAYTVSNMAEVRAAYGLDFSDLYYAAAYKSIQPENLEYAAYKSIQP
ENLEYAAYGLDFSDLYYGPGEVRSYCYEASISDMAGPGPGIFRNPGFALVAVAIAGPGP
GHAKRQTVVVLGSQEGGPGGAHAKRQTVVVLGSQEGPGPNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKGVSNRDFVKKSQHSGMIGYETDENKKGSQHSGMIGYETDENK **A**
KAKFVAAWTLKAAAHHHHHH

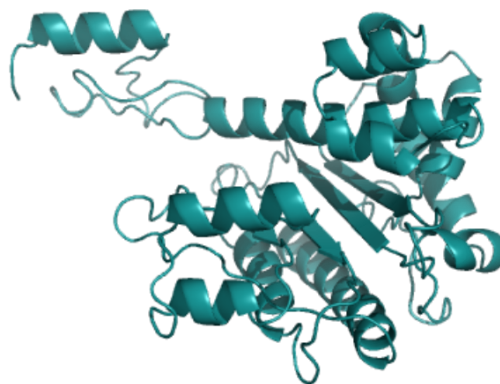
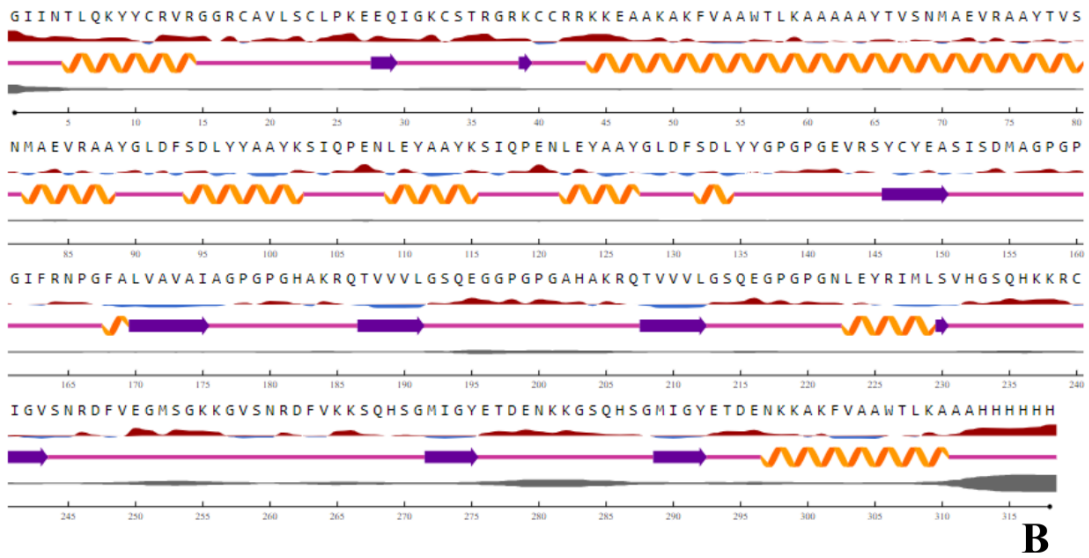
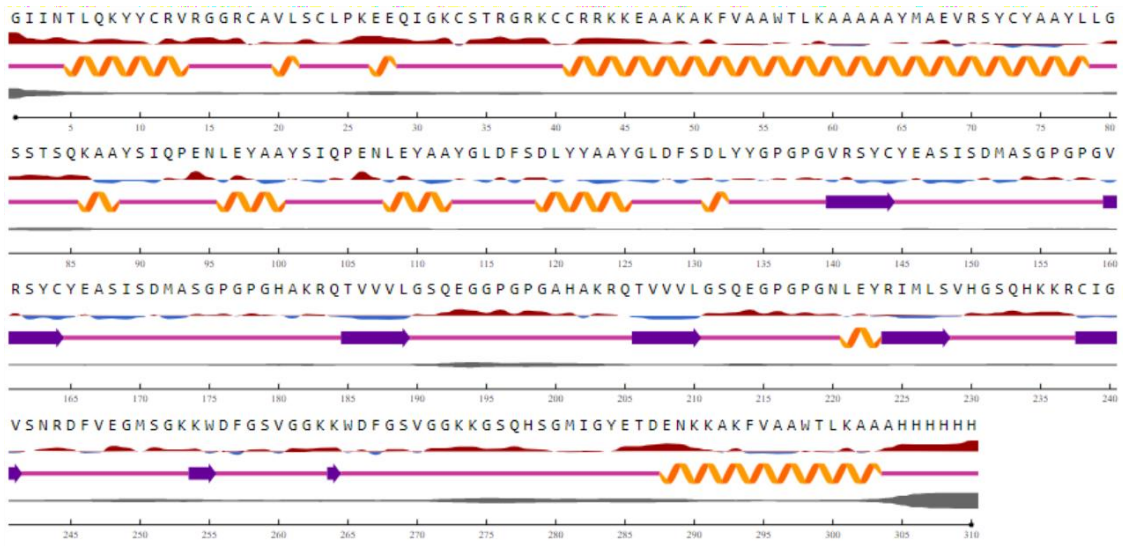


Figure 3. 22: Structure of vaccine 17 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

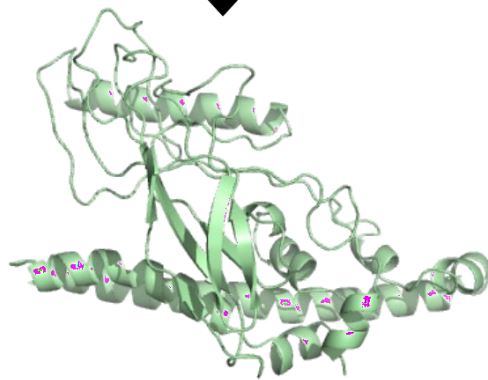
Vaccine -18

GIINTLQKYYCRVRGGRCVLSCLPKKEEQIGKCSTRGRKCCRRKKEAAKAKFVAAWTLK
AAAAAYMAEVRSYCYAAYLLGSSTSQKAAYSIQPENLEYAAYSIQPENLEYAAYGLDFSDL
YYAAYGLDFSDLYYGPVGVRSYCYEASISDMASGPGVRSYCYEASISDMASGPGVGH
KRQTVVVLGSQEGGPGGAHAKRQTVVVLGSQEGPGNLEYRIMLSVHGSQHKKRCIGV
SNRDFVEGMSGKKWDFGVSQGGKKWDFGVSQGGKKSQHSQMIGYETDENKKAKFVAAW
TLKAAAHHHHHH

A



B



C

Figure 3. 23 : Structure of vaccine 18 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

Vaccine -19

GIINTLQKYYCRVRGGRCVLSCLPKEEQIGKICSTRGRKCCRKKEAAKAKFVAAWTLK
AAAAAYMAEVRSYCYAAYTVSNMAEVRAAYSIQPENLEYAAYGLDFSDLYYAAYGLDFS
DLYYAAYGLDFSDLYYGGPGVRSYCYEASISDMASGPGPGEVRSYCYEASISDMAGPGP
GHAKRQTVVVLGSQEGGPGPHAKRQTVVVLGSQEGGPGPNLEYRIMLSVHGSQHKKR
CIGVSNRDFVEGMSGKKWDFGSGVGGKKWDFGSGVGGKKGSHSGMIGYETDENKKAKF
VAAWTLKAAAHHHHHH **A**

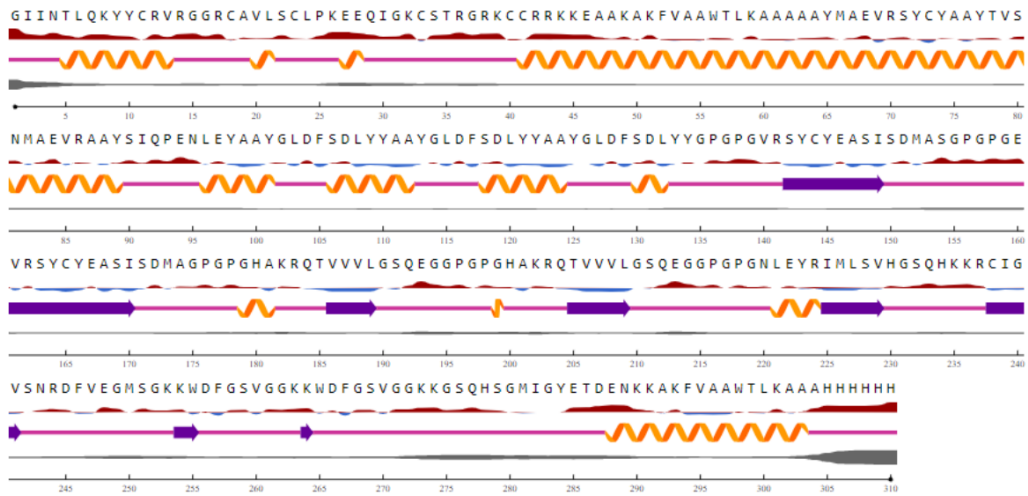


Figure 3. 24: Structure of vaccine 19 (A). Primary structure, (B). Secondary structure, (C). Tertiary structure.

3.5) Selected vaccine analysis

To calculate the accuracy of the constructed vaccine sequence ERRAT and validation were checked by Structure Analysis and Verification Server (saves 2.0) to inspect the quality of the three-dimensional structure for a quantitative analysis to validate the native conformation state of each one and determine the errors. Ramachandran plots were also measured subsequently and fundamentally depict the torsional angles, dihedral angles, and steric hindrance along with the favored region.

3.5.1) Validation, ERRAT and Ramachandran favored and disallowed region

Table 3. 17: Ramachandran plot and ERRAT value of 10 selected vaccines.

Selected Vaccine no	Name	ERRAT value	Ramachandra Chandran favored Region	Ramachandra Chandran Disallowed region
2	A-rsp2	94.279%	98.1%	1.3%
4	B-rso8	83.33%	96.7%	2.3%
5	C-rsp1	94.97%	98.6%	1.4%
6	D-rsn6	89.5928%	97.9%	1%
8	E-rso6	93.1953%	98.4%	1.6%
9	F-rsh1	91.0714%	99.4%	0.3%
11	G- rsh5	86.901%	98.5%	0.4%
12	H-rsh6	87.5839%	98.8%	0.8%
16	I-rsg7	91.1184%	94.6%	1.9%
18	J-rso7	85.4167%	98.5%	1.2%

3.5.2) Ramachandran Plot analysis

According to the PDBsum PROcheck if the protein structure of the selected vaccine has more than 20% R-factor and has over 90% residues in the most favored region, it is recognized and predicted to be a stable protein. The atoms were used for close contact to inspect the structure for each conformation. less than 2% of disallowed regions are expected to be good. The disallowed region produces steric hindrance between chains. All 10 selected vaccines were checked and analyzed

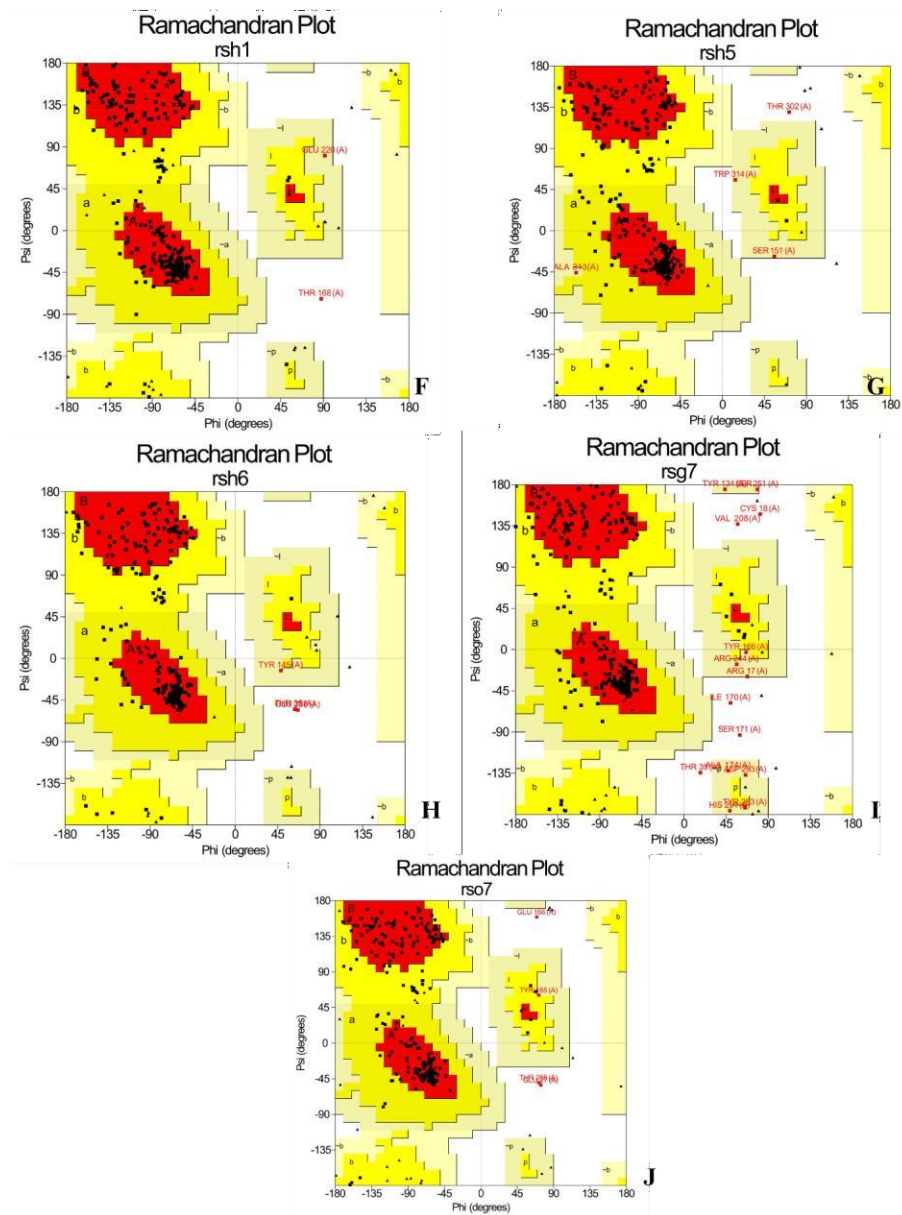


Figure 3. 26: Ramachandran Plot

- A. Vaccine 2 has 98.1% residues in most favored regions and disallowed regions are 1.3%. That makes the Total number of residues are 448 amino acids.
- B. Vaccine 4 has 96.7% of the most favored region which also includes residues in disallowed regions 7 2.3% and the total number of residues are 350 amino acids.
- C. Vaccine 5 has 98.6% residues in most favored regions and disallowed regions 1.4%. Total number of residues 368 amino acids.

- D. Vaccine 6 has 97.9 % favored region along with 1% disallowed region. It has a total number of residues of 452 amino acids.
- E. Vaccine 8 has 98.4% most favored regions and residues in disallowed regions is only 1.6%. Total number of residues 356 amino acids.
- F. Vaccine 9 has 99.4% most favored region and only residues in disallowed regions are 0.6%. Total number of residues 351 amino acids.
- G. Vaccine 11 has the 98.5% most favored region according to the plot and the residues in disallowed regions are 0.4% also the total number of residues 326 amino acids.
- H. Vaccine 12 has the Residues in most favored regions which is 98.8% and residues in disallowed regions are 0.8% and the total number of residues are 310 amino acids.
- I. Vaccine 16 has 94.6% of favored region and Total number of residues 316, The disallowed region includes 1.9%.
- J. The vaccine 18 has 98.5% of the favored region and 1.2% of the disallowed region. It has a total number of residues of 310 amino acids.

Both the Ramachandran and ERRAT values were used to evaluate an overall quality of the construction vaccine structure which was based on atomic interactions, residual errors, and backbone conformation.

3.6) Molecular Docking

The docking server Cluspro 2.0 was employed for molecular docking against TLR4 (Toll Like receptors 4). TLR receptor has the pivotal role in immune response and is frequently expressed. After retrieving the PDB data from UniProt it was used to dock against the ligand (vaccine). It predicted the attachment of the molecules and mathematically calculated the average and the lowest score for each model. The lower docking energy was considered the better and more stable model. However, the tool gave 10 models for each vaccine. Later the prodigy server measured the Binding energy. This server predicted the energy score between the interaction of receptor and the ligand. Lower binding energy contemplated higher stability. The figure below is the docking of 10 selected vaccine models.

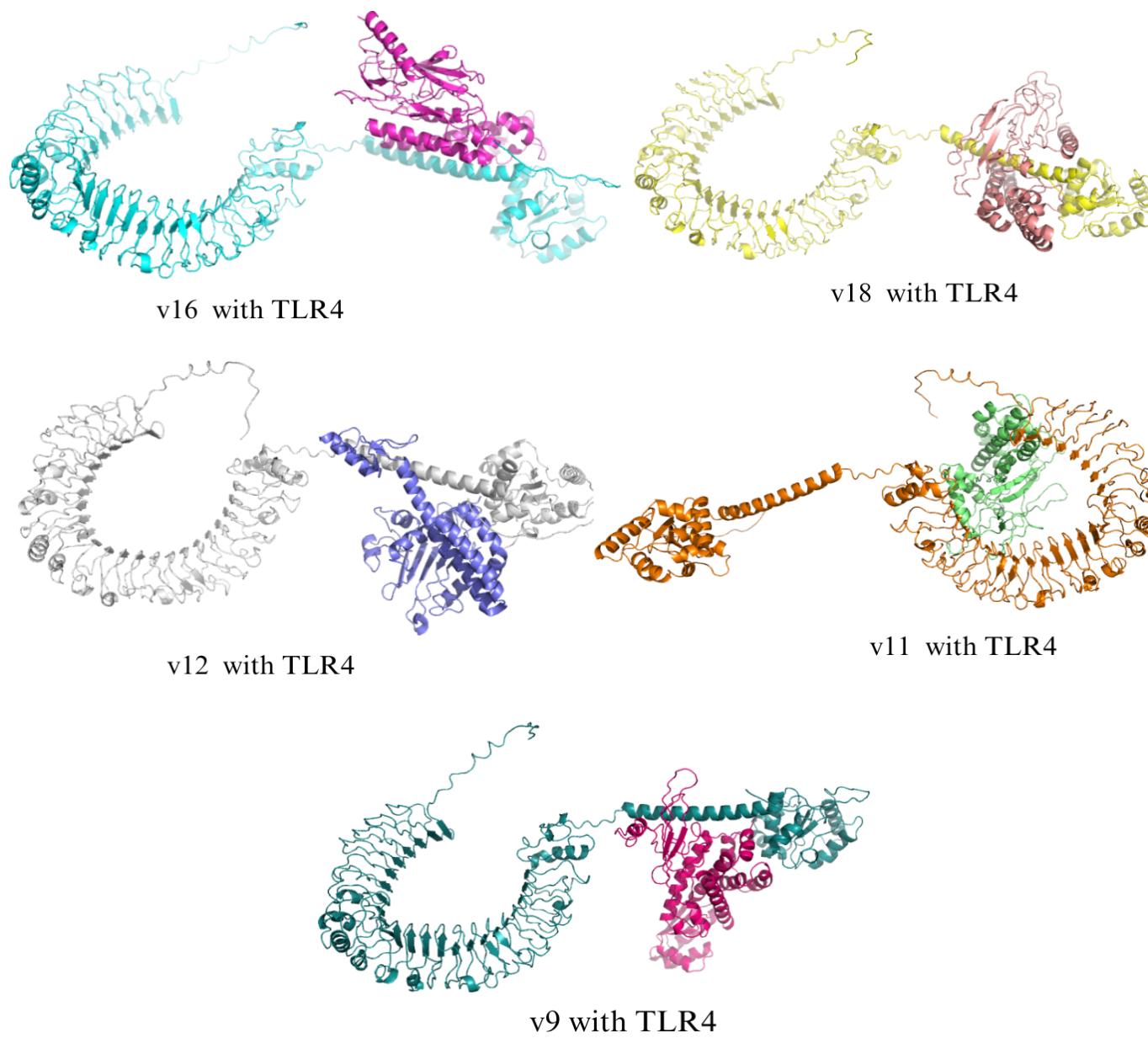


Figure 3. 27: Molecular Docking of Vaccine 16,18,12,11,9 with Receptor TLR4

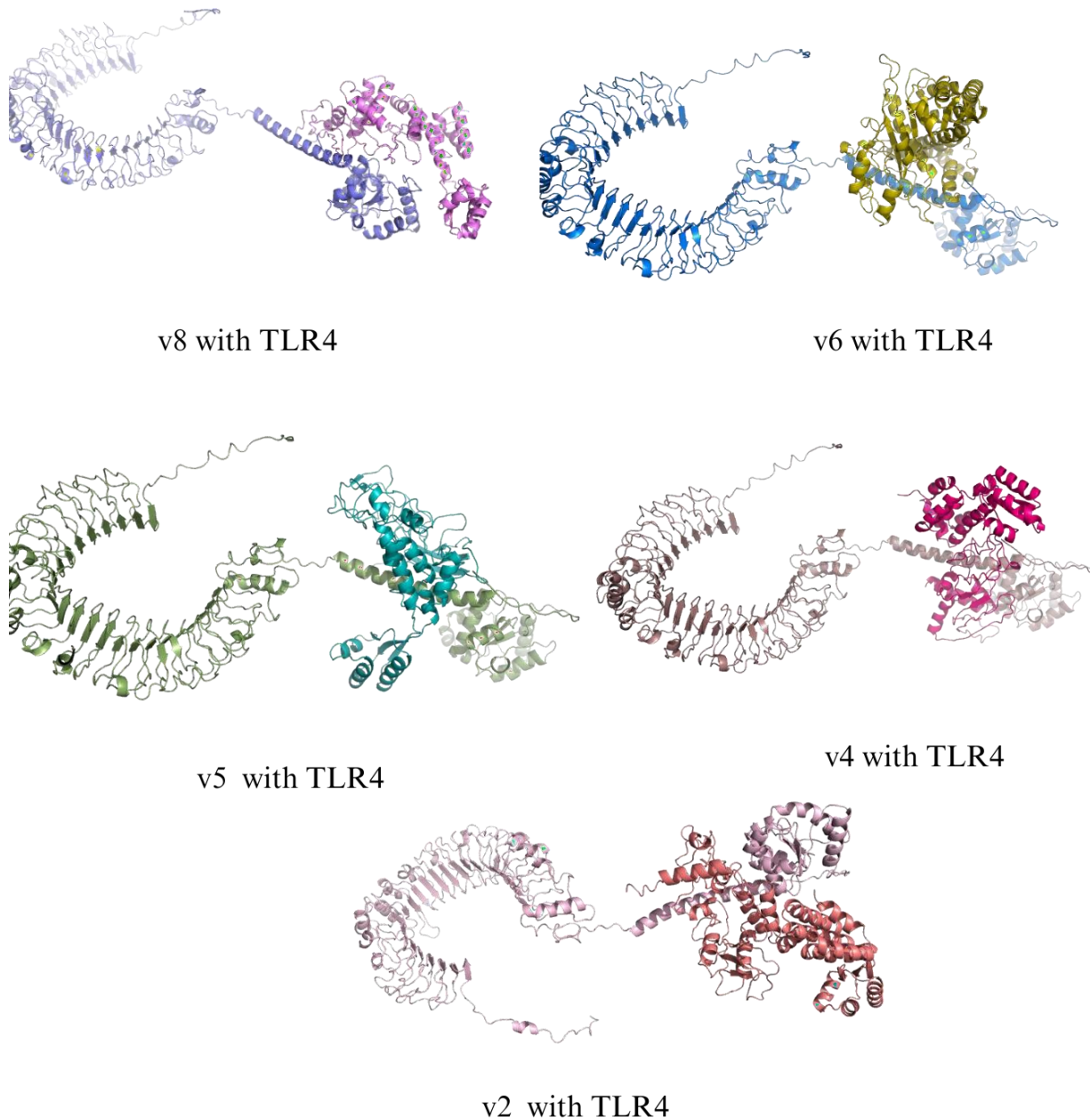


Figure 3. 28: Molecular Docking of Vaccine 8,6,5,4, 2 with Receptor TLR4

3.6.1) Target receptor and vaccine binding energy

Docking and Gibbs free energy was predicted to find out the lowest energy for higher stability among the selected vaccines. After that, intermolecular hydrogen bonds were also calculated as

Table 3. 18: Docking and Binding energy.

Selected Vaccine No	Target	Representative	Docking Score (weighted)	Gibbs Free energy ΔG (kcal mol⁻¹)	Hydrogen bond
2	TLR4	Center	-1166.7	-15.6	18
		Lowest Energy	-1370.5		
4	TLR4	Center	-952.4	-16.3	4
		Lowest Energy	-1270.6		
5	TLR4	Center	-968.7	-10.5	8
		Lowest Energy	-1072.4		
6	TLR4	Center	-1040.2	-9.5	8
		Lowest Energy	-1133.1		
8	TLR4	Center	-1198.3	-15.5	7
		Lowest Energy	-1275.8		
9	TLR4	Center	-1032.2	-16.1	4
		Lowest Energy	-1032.2		
11	TLR4	Center	-1043.7	-14.3	17
		Lowest Energy	-1132.9		
12	TLR4	Center	-989.7	-13.1	7
		Lowest Energy	-1395.8		
16	TLR4	Center	-987.3	-11.9	8
		Lowest Energy	-1111.5		
18	TLR4	Center	-1172.1	-15.6	14
		Lowest Energy	-1260.0		

3.6.2) Surface, Hydrogen, and non-bound Interaction analysis during docking

Molecular dockings predict the surface level interaction of ligand receptors by scoring function and algorithms. Additionally, it predicts the bonded and non-bonded interactions too.

Vaccine 5- bonds and interaction

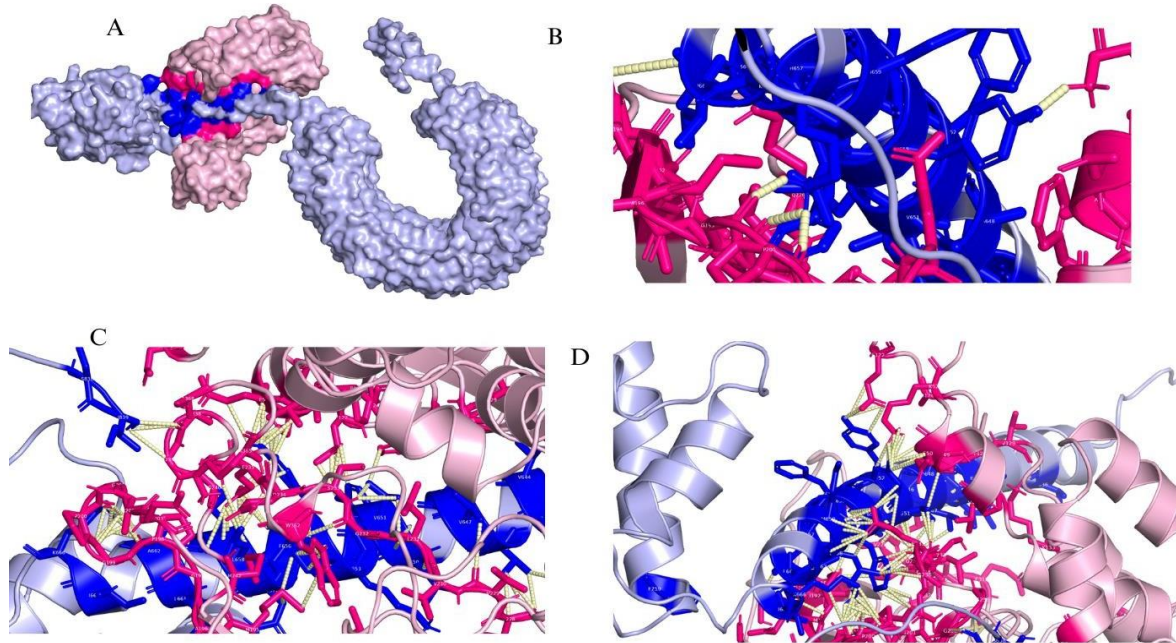


Figure 3. 29: Vaccine5, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction

Vaccine 12 - bonds and interaction

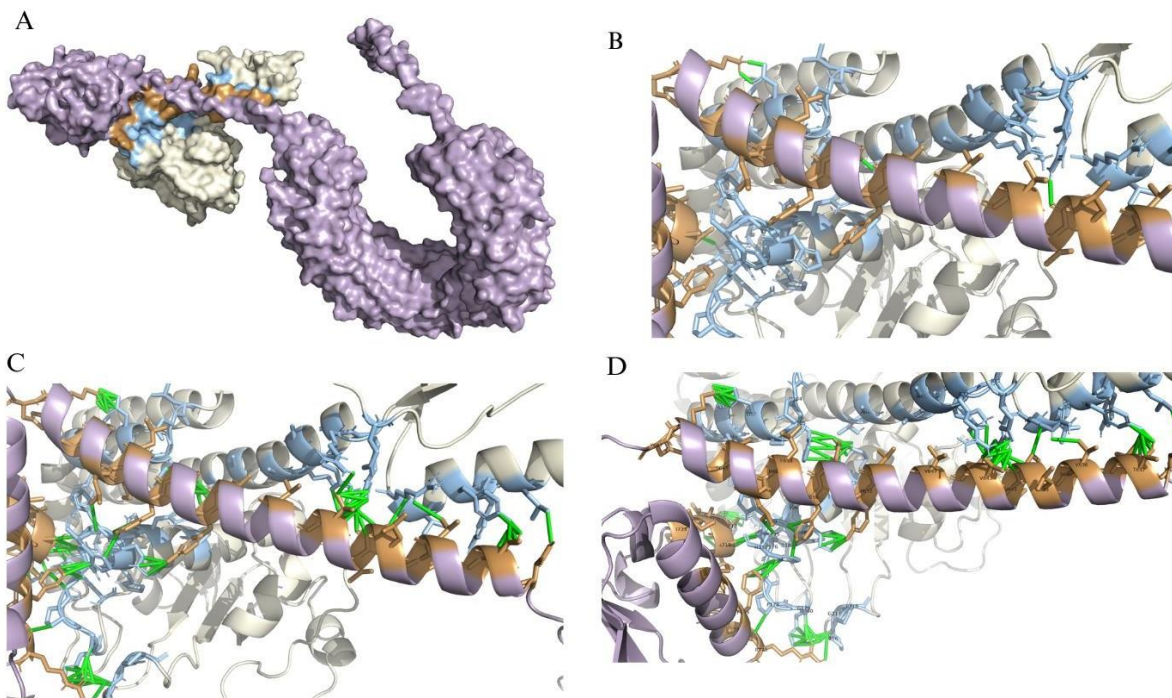


Figure 3.30: Vaccine 12, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction

Vaccine 16 - bonds and interaction

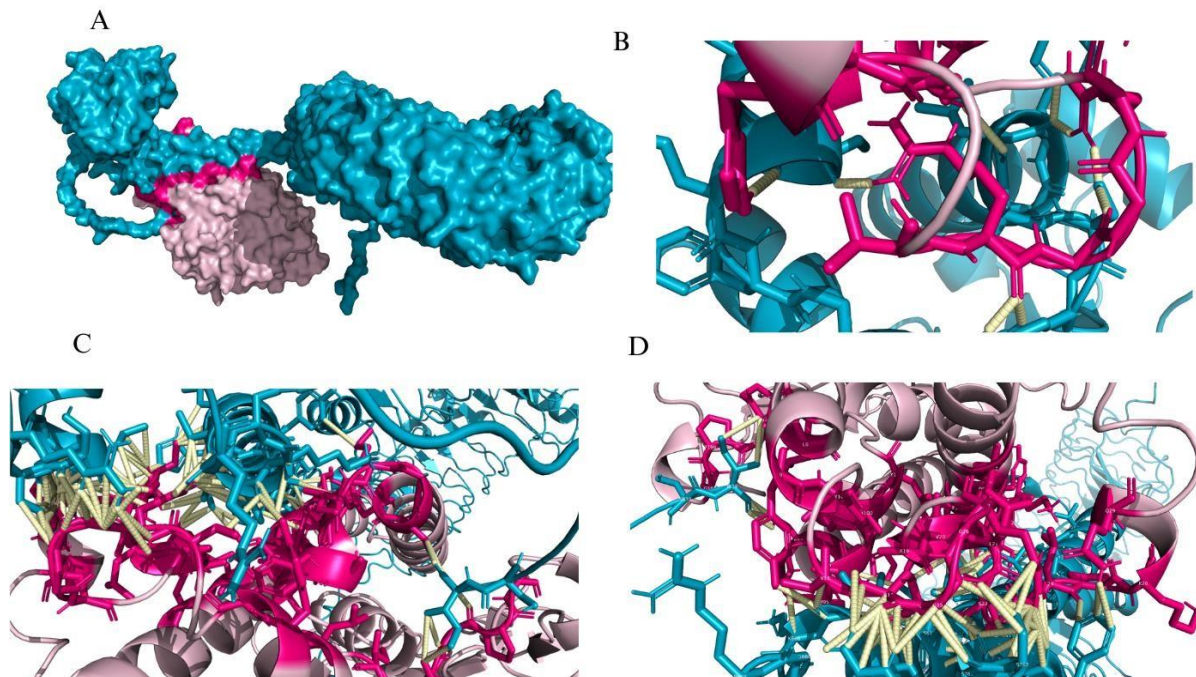


Figure 3.31: Vaccine 16, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction

Vaccine 18 - bonds and interaction

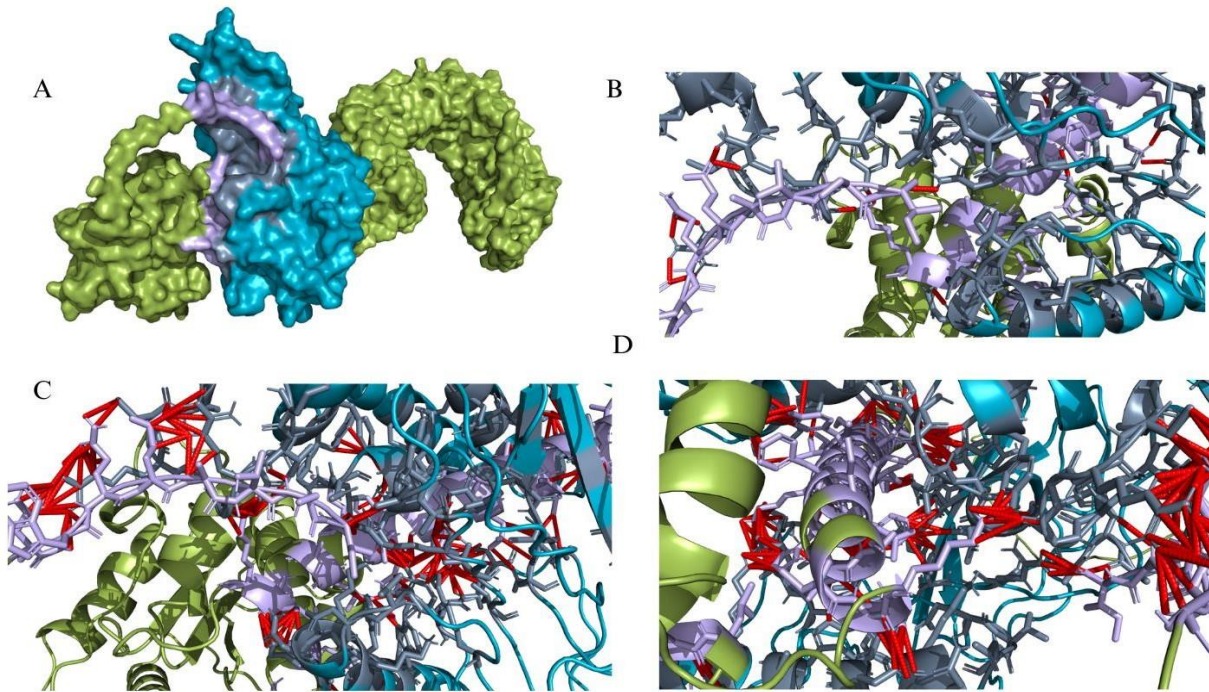


Figure 3.32: Vaccine 18, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction

Vaccine 11 - bonds and interaction

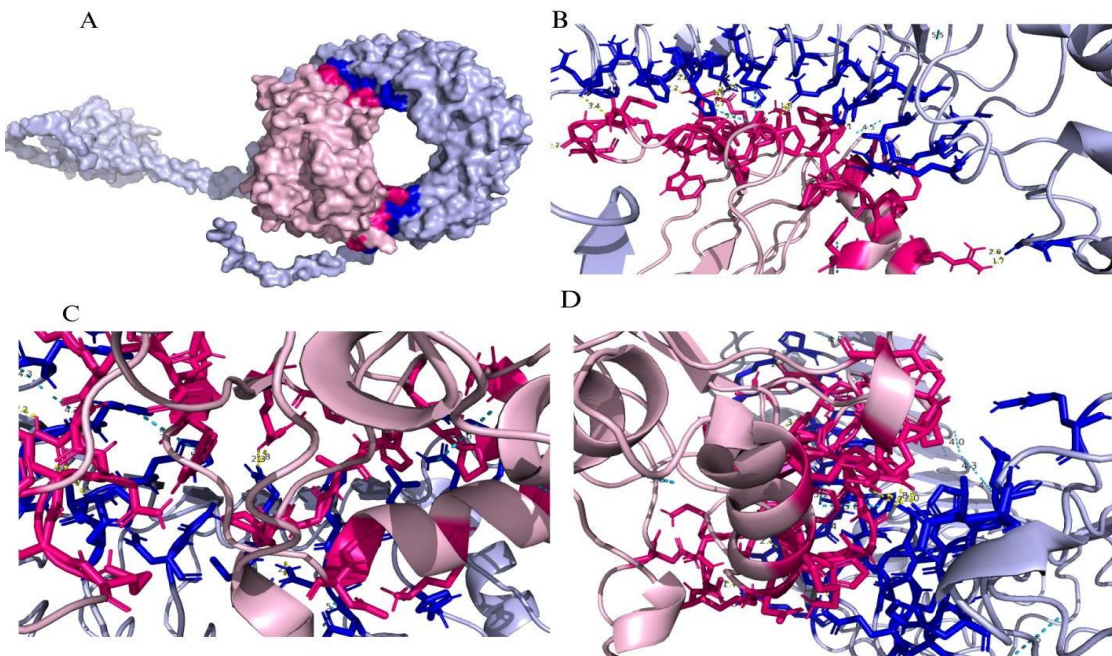
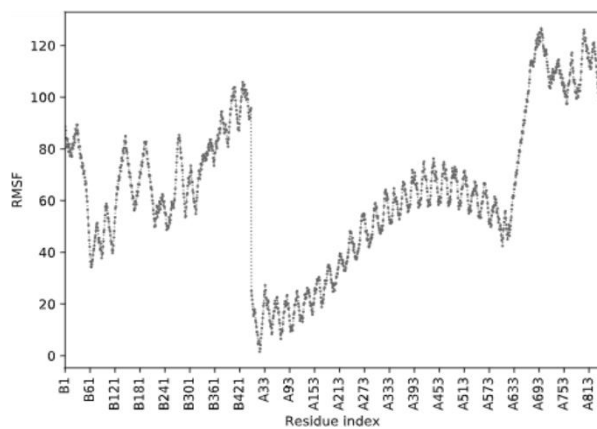


Figure 3.33: Vaccine 11, A) Surface interaction B) Hydrogen Bond C, D) Non-bonded Interaction

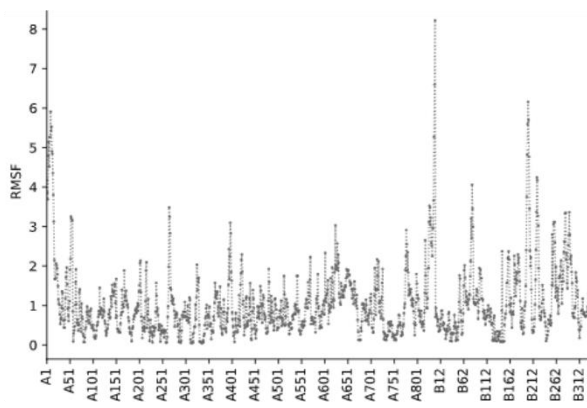
3.7) Structural Simulation

3.7.1) Fluctuation plot analysis

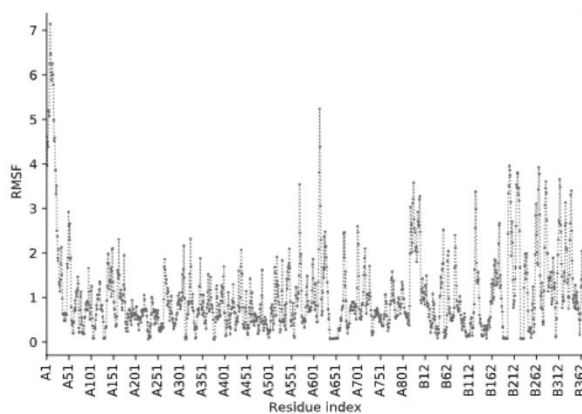
For structural simulation CABS-flex 2.0 was used to model the protein structure and run for efficient simulation that allowed the selected vaccines to have large scale conformational transitions. The Root Mean Square Fluctuation (RMSF) value was measured through the fluctuation graph for each vaccine both for chain A and chain B. The results showed that each residue fluctuates during the simulation. Lower RMSF value most favorably lower than 4 angstrom allows a vaccine structure to be well defined and stable. Fluctuations plot for each vaccine below showed the RMSF values that were plotted for both the chain A and B residue index.



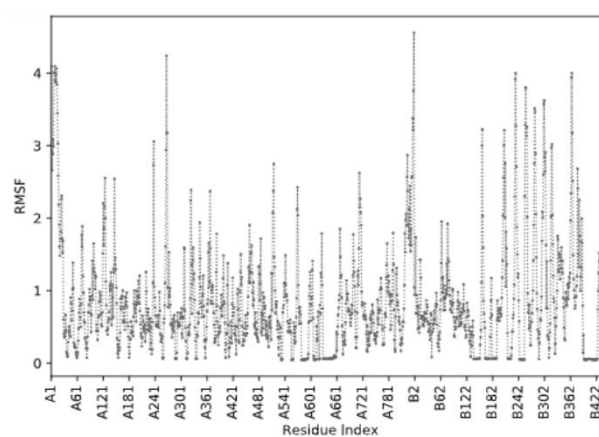
Vaccine 2



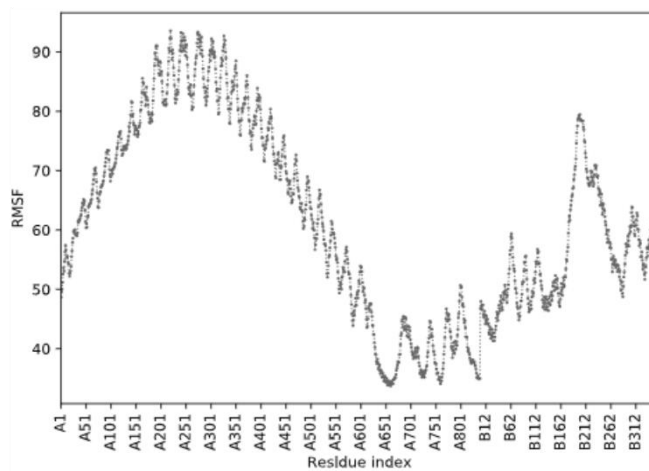
Vaccine 4



Vaccine 5



Vaccine 6



vaccine 8

Figure 3. 34: Fluctuation Plot of vaccine-receptor complex

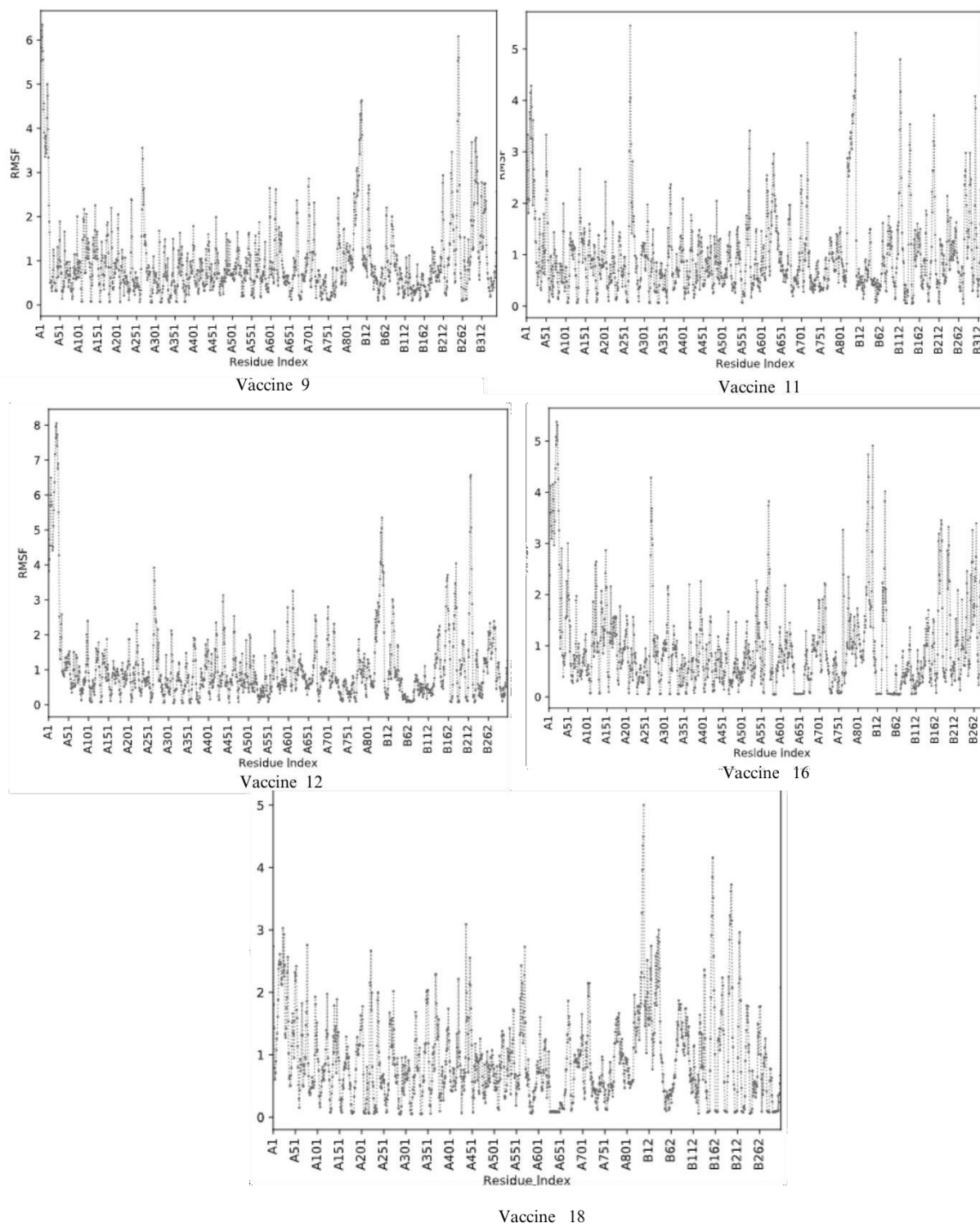


Figure 3. 35: Fluctuation Plot of vaccine-receptor complex

Vaccine 2 – The graph shows that vaccine 2 led to a high fluctuation for both the chain B and A. Hence, it was considered as an unstable protein due to the possibilities of aggregation and misfolding of the structure. Also, the fluctuation crossed 120 angstrom RMSF which means the native conformation had changed.

Vaccine 4 –The residue index of the chain A and B both had a moderate and stable fluctuation, but Chain A residue 301 crossed 5 angstroms, and from Chain A residue 102 to Chain B residue 312, the fluctuation plot was high, which means it could misfold.

Vaccine 5 – The graph shows an incredibly good fluctuation throughout the residue index; only chain A601 to chain A651 were slightly above 5 angstroms.

Vaccine 6 – The graph shows inflated fluctuation points for both chains A and B in the residue index; it signifies that the protein sequence had unstable native conformational change, which led to the possibility of a misfolded sequence.

Vaccine 8 – The RMSF value crossed 90 angstroms for chain A residues, and chain B residues showed high fluctuations, indicating a very unstable protein.

Vaccine 9 – It showed a good plot where all the residues were under 4 angstrom and predicted to have good stability in conformational changes.

Vaccine 11 – Here, maximum residues' RMSF were under 4 angstroms but chain A251, A801 to chain B12, and chain B112 to b312 had inflated fluctuations.

Vaccine 12 – The graph shows that vaccine 12 resulted in great stability and firmness, which was paramount, with less than 4 angstroms, and it is denoted as a good structure.

Vaccine 16 – The result of the fluctuation plot also signifies a very formed and balanced structure due to the low RMSF value for both chains A and B.

Vaccine18 – It showed that a considerable amount of the residues for both chains A and B are well stout and stable, and the sub-intentional proportion had less than 4-angstrom RMSF value, which predicted a good and unchanged native conformational situation.

From the fluctuation plots, vaccines 4, 5,9,11,16, and 18 showed a prominent structure and were predicted to be stable in their native conformation state.

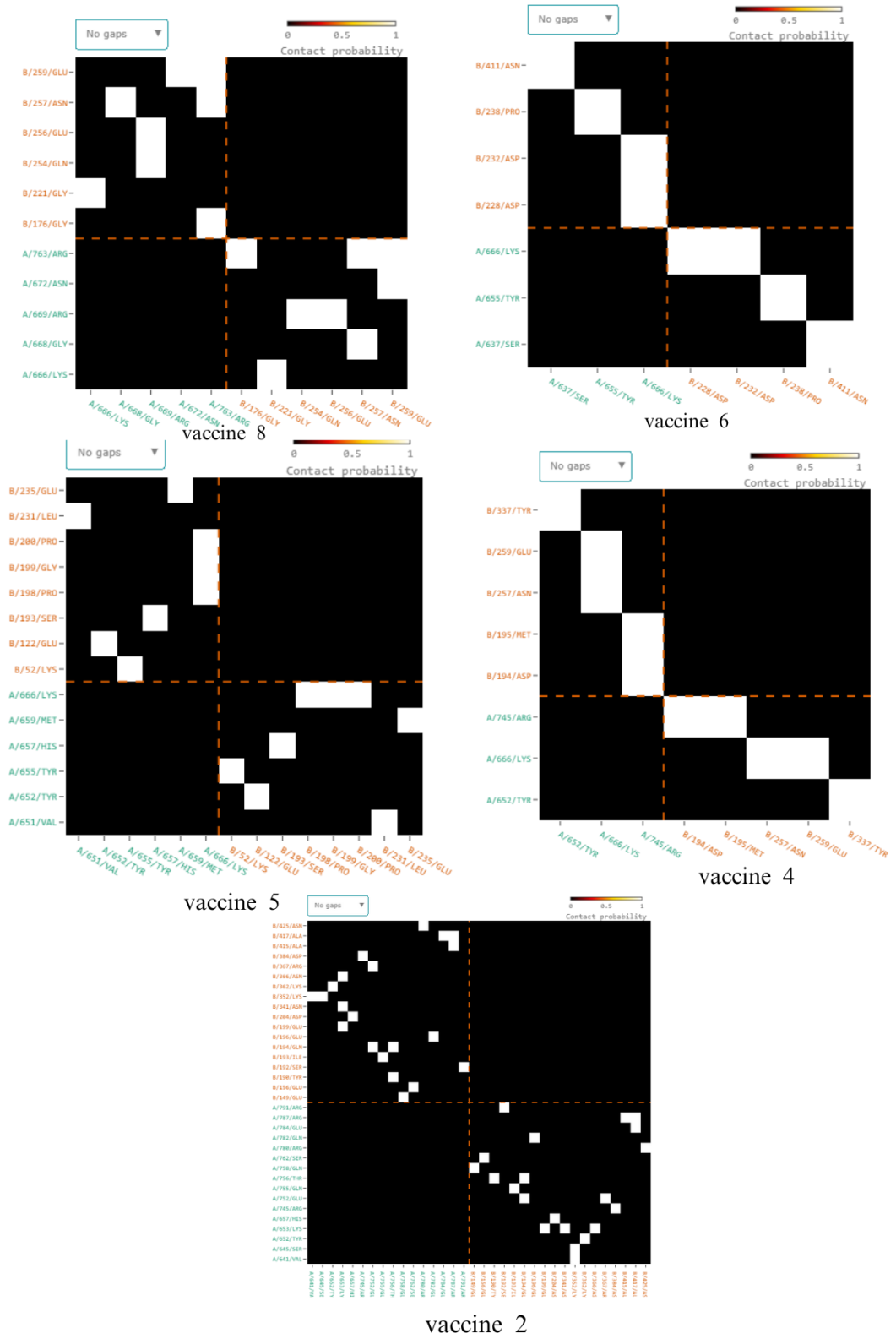


Figure 3. 37: Contact Map of vaccine 8,6,5,4,2.

Table 3. 19: Interactive regions for chain A and B.

Vaccine	Chain A Receptor binding region	Chain B Vaccine binding regions
2	758-GLN,762-SER,756-THR,791-ARG,755-GLN,752-GLU,756-THR,782-GLN,653-LYS,657-HIS,653-LYS,641-VAL,645-SER,652-TYR,653-LYS,752-GLU,745-ARG,787-ARG,84-GLU,787-ARG,780-ARG	149-GLU,156-GLU,190-TYR,192-SER,193-ILE,194-GLN,194-GLN,196-GLU,199-GLU,204-ASP,341-ASN,352-LYS,352-LYS,362-LYS,366-ASN,367-ARG,384-ASP,415-ALA,417-ALA,417-ALA
4	652-TYR, A/666-LYS, A/666-LYS, A/745-ARG, A/745-ARG	257-ASN,259-GLU,194-ASP,195-MET
5	651-VAL,652-TYR,655-TYR,657-HIS,659-MET,666-LYS,666-LYS,666-LYS	231-LEU,122-GLU,52-LYS,193-SER,235-GLU,198-PRO,199-GLY,200-PRO
6	637-SER,655-TYR,666-LYS,666-LYS	411-ASN,238-PRO,228-ASP,232-ASP
8	668-GLY,669-ARG,669-ARG,672-ASN,763-ARG,763-ARG,763-ARG	257-ASN,254-GLN,256-GLU,259-GLU,176-GLY,257-ASN,259-GLU
9	653-LYS,664-CYS,669-ARG,763-ARG	149-GLU,2-SER,10-GLU,7-GLN
11	30-VAL,31-GLU,39-GLN,41-MET,42-GLU,42-GLU,362-LYS,382-ARG,382-ARG,409-ASN,433-ASN,435-LYS,458-HIS,505-GLN,507-GLN,616-GLN	15-GLY,17-ARG,14-ARG,39-LYS,12-ARG,39-LYS,280-GLY,278-ASP,283-HIS,285-ASN,285-ASN,288-GLU,285-ASN,142-ARG,178-GLY,260-ARG
12	650-LEU,654-PHE,666-LYS,5-GLN,75-PH	143-TYR,143-TYR,107-ASP,179-GLY,308-HIS
16	664-CYS,665-ILE,759-PHE,760-LEU,762-SER,762-SER,763-ARG	14-ARG,9-TYR,17-ARG,22-SER,15-GLY,17-ARG,14-ARG
18	645-SER,652-TYR,653-LYS,653-LYS,653-LYS,655-TYR,658-LEU,665-ILE,666-LYS,666-LYS,832-ASN,834-GLN,835-GLU,839-ILE	94-GLU,70-ARG,166-GLU,167-ALA,168-SER,121-TYR,165-TYR,293-LYS,15-GLY,122-TYR,17-ARG,17-ARG,14-ARG,204-ARG

3.8) Immune Simulation

The immune simulation was done by the C-Immsim server via a computational approach to predict the cellular interaction of the selected vaccine in the immune system. However, the selected vaccines were administered in a single dose, and the result was observed for 30 to 35 days. Additionally, the B cell and T helper cell populations were checked via the antibodies, and memory cell production alongside immunoglobulin, immune complexes, cytokines, and interleukins were also checked. Based on the highest production of the antigen in the immune system, the best candidates were selected. All the candidates showed a stranded immune response for the first shot. Hence, we set a threshold value for the selected vaccine, which is that the vaccine must produce more than 10,000 antibodies during the first administration.

Vaccine 2

Figure A) below shows the B-cell population. During the first few days, the number of antibodies was increasing, and during 5 to 10 days, it was the highest, with between 500 to 600 cells, and the memory cells were 200. It was in a stagnant position. In the Next graph, Figure B), the administration of vaccine 2 showed the highest number of T helper cells, which was between 4500-5000, and it slowly degraded after 5 to 10 days after administration, and the Memory cell count was 200 to 250. C) The simulation below of vaccine 2 administration showed that IgM+IgG antigen count was close to 700000 in 15 days. After that, it gradually dropped. The antibody titer was also more than 7000 and close to 8000 of IgM, IgG1, and IgG2. It also showed that the number of D) cytotoxic T was highest, close to 80000 counts, but it lowered after that and was between 400000-450000 after the first injection.

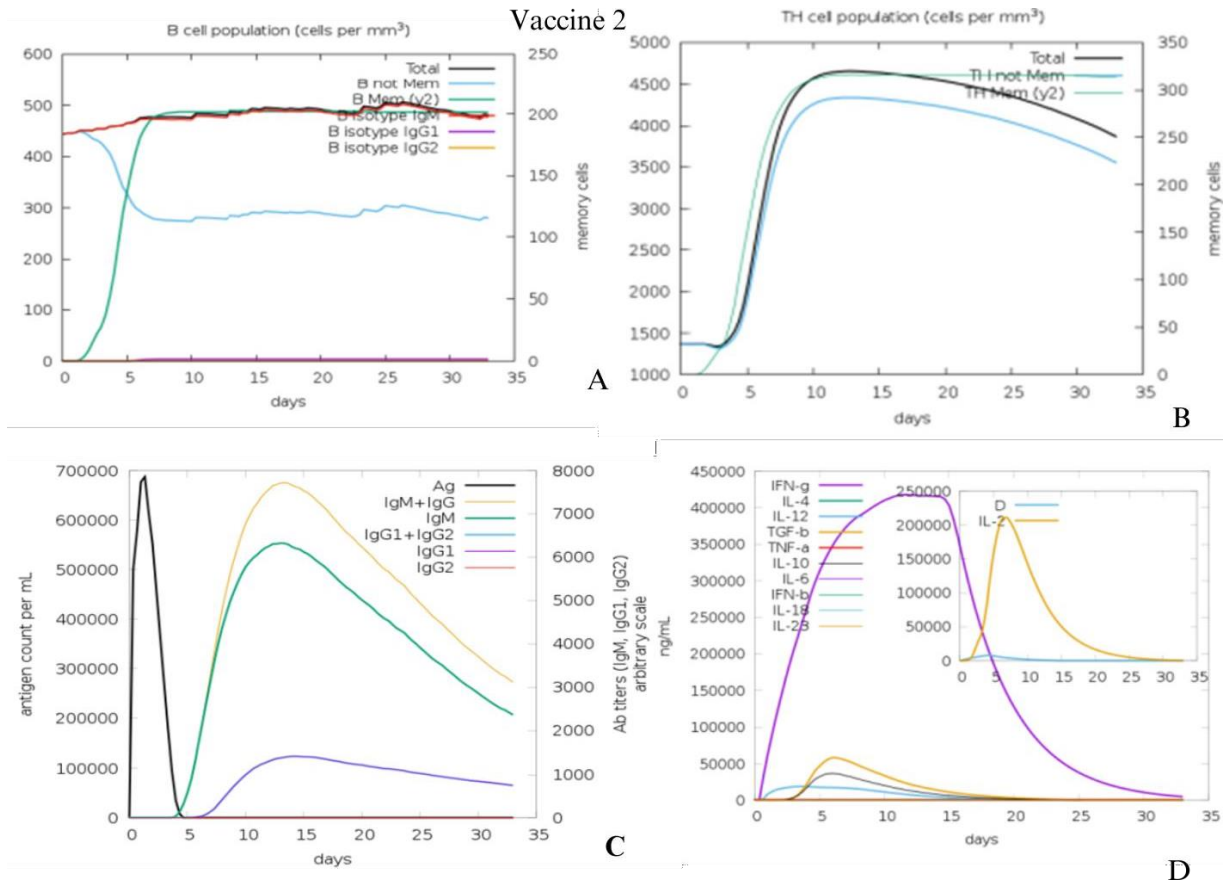


Figure 3.38 : Vaccine 2 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 4

In Figure A) below the b cell population, during the first few days, the number of antibodies was increasing, and during 5 to 10 days, it was the highest, which is between 400 to 500 cells, and the memory cells were more than 200 and close to 300. It was in a stagnant position. In the Next graph, figure B), the administration of vaccine 4 showed the highest number of T cells, which was between 4500-5000, but after 10 days, it was 5000 antibodies for 35 days. And the memory cell count was 350. The simulation below of vaccine 2 administration showed in Figure C) that IgM+IgG antigen count was between 600000-700000 for 10 days, which was the highest after that. It gradually lowered. After 35 days, the antibody peaked at slightly higher but close to 10000 counts. Then it lowered to 5000 to 6000 counts of IgM, IgG1 and IgG2. The next figure D) of

cytokines and interleukins showed the IFN-g was slightly more than 400000 during 10 to 15 days, and after 20 days, it gradually decreased.

Vaccine 4

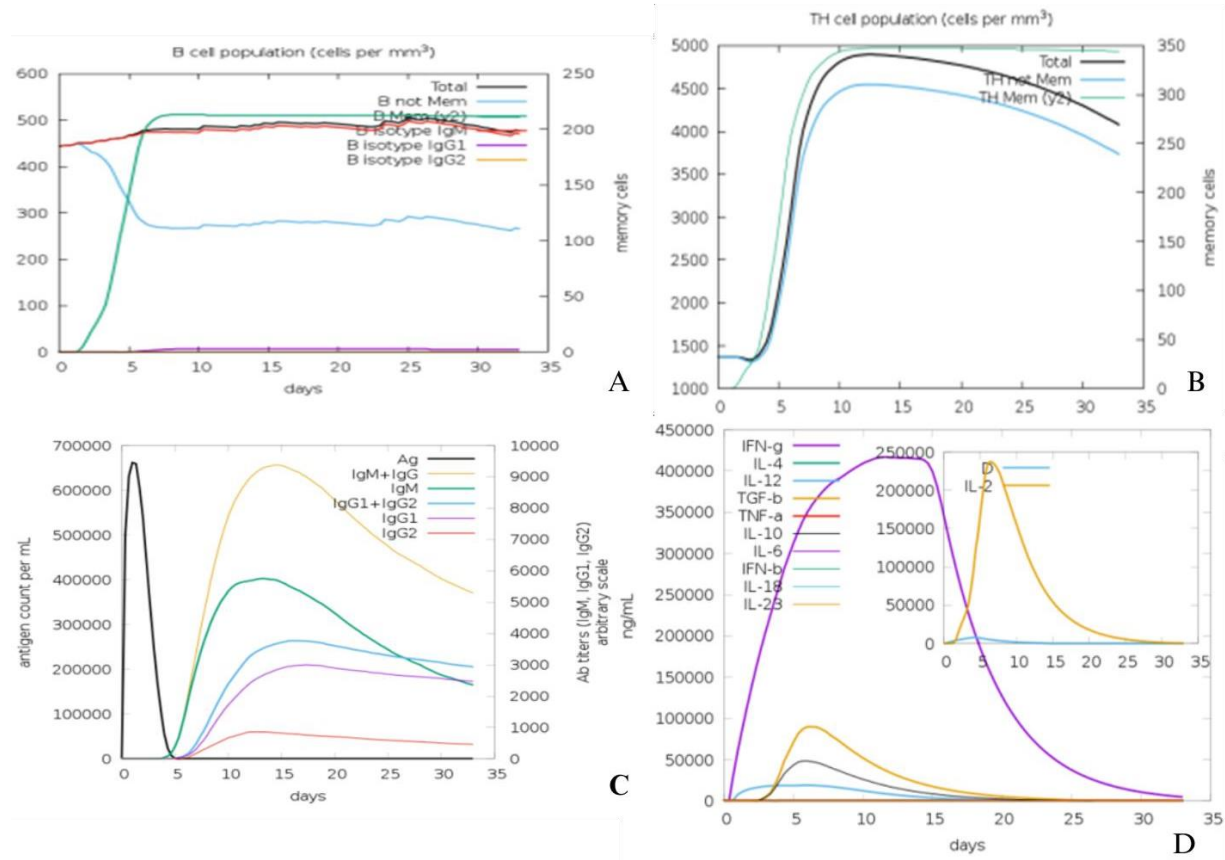


Figure 3.39: Vaccine 4 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 5

Figure A) below shows the b cell population. During the first 4 days the number of antibodies increased and after that, during 5 to 10 days it was the highest, which is between 500 to 600 antibodies and the memory cells were more than 200 and close to 250. It was in a stagnant position. In the Next graph figure B) the administration of vaccine 5 showed the highest number of T helper cell cells which was between 5000-5500 antibodies but after 10 days it was static for memory cells 400 for the remaining 35 days. The simulation below figure C) administration showed IgM+IgG antigen count was between 500000-600000 for 10 days, it was the highest after that it gradually lowered, and the antigen was at peak during the first 5 days which was more than 12000 count. And before 35 days the antibody titer was at peak which is more than slightly more than 12000

counts then after 20 days it reduced to 4000 to 6000 of IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was between 40000-450000 counts during 10 to 15 days and after 20 days it gradually decreased.

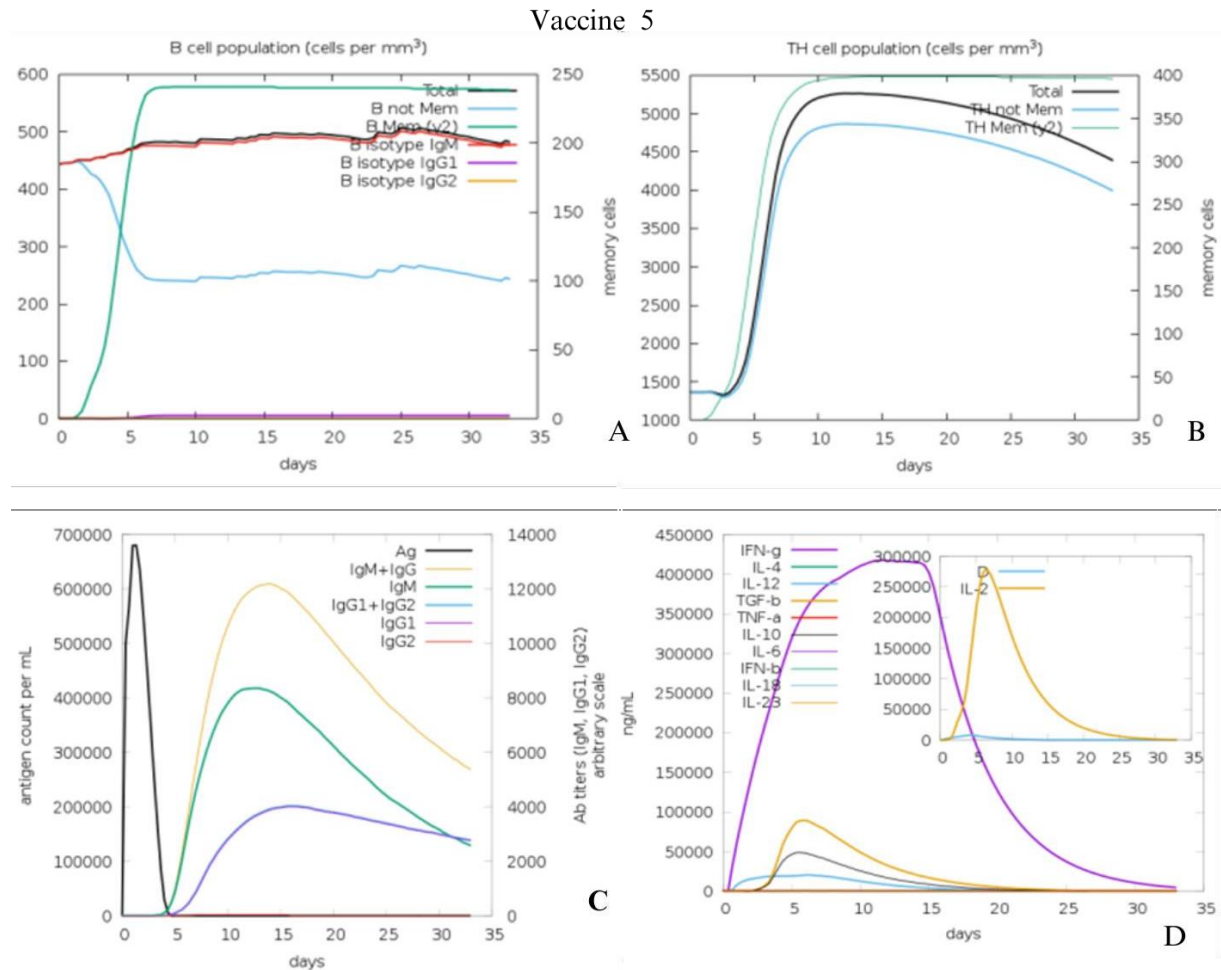


Figure 3. 40: Vaccine 5 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 6

Figure A) below shows the b cell population. During the first 4 days the number of the antibodies increased and after that, during 5 to 10 days it was the highest, which is between 500 antibodies and the memory cells were slightly greater than 200. It was in an inert position. In the Next graph figure B) the administration of vaccine 6 showed the highest number of T helper cell cells which was between 4500-5000 antibodies but after 10 days it was static for memory cells 350 for the remaining 35 days. The simulation below from the figure C) administration showed IgM+IgG antigen count was between 600000-700000 between 10 to 15 days, it was the highest after 16 days

and that it gradually lowered, and the antigen was at peak during the first 5 days which was only 700000 count. And before 35 days the antibody titer was 7000 and lowered to 4000 for IgM+IgG or IgM, IgG1 and IgG2. Besides, the figure D) also showed that the number of IFN-gamma cells was slightly above 400,000 counts during 10 to 15 days and after 20 days it gradually decreased.

Vaccine 6

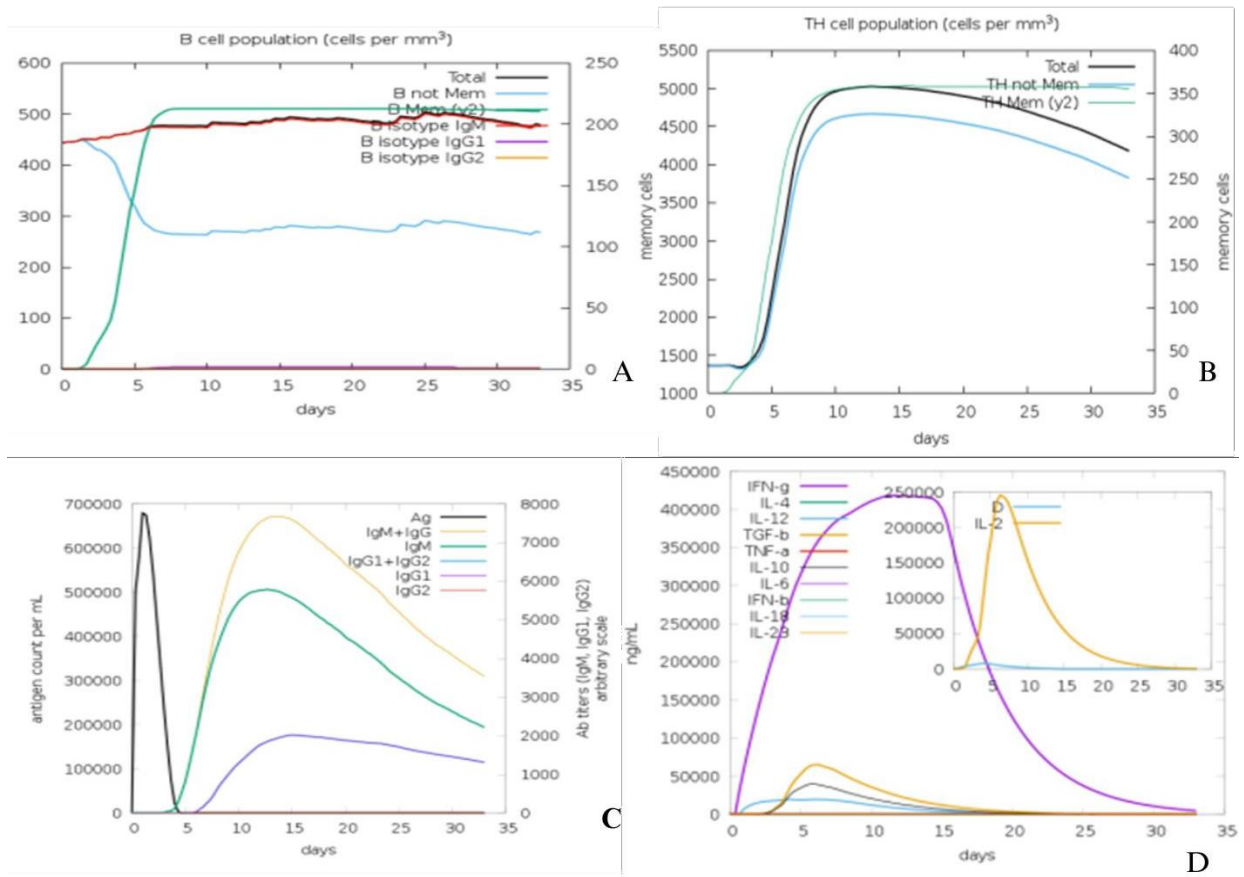


Figure 3.41: Vaccine 6 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 8

Figure A) below shows the b cell population. During the first 4 days the number of the antibodies were increasing and after that, after 5 days of the administration it was the highest, which was standing at 500 antibodies and the memory cells were 200 and in the next figure, the graph figure B) the administration of vaccine 8 showed the gradual increment antibodies and the highest number of T helper cell cells which was a static 5000 antibodies after 10 days but after that it was steady for memory cells which was 400 counts for the remaining days. Figure C) below is showing the simulation of vaccine 8 administration showed IgM+IgG antigen count was between 600000-

700000 between 10 to 15 days, it was the highest then it gradually lowered and the antigen was at peak during the first 5 days which was 9000 and reduced to 5000 counts for IgM+IgG of IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was slightly above 40000 counts during 10 to 15 days and after 20 days it gradually reduced.

Vaccine 8

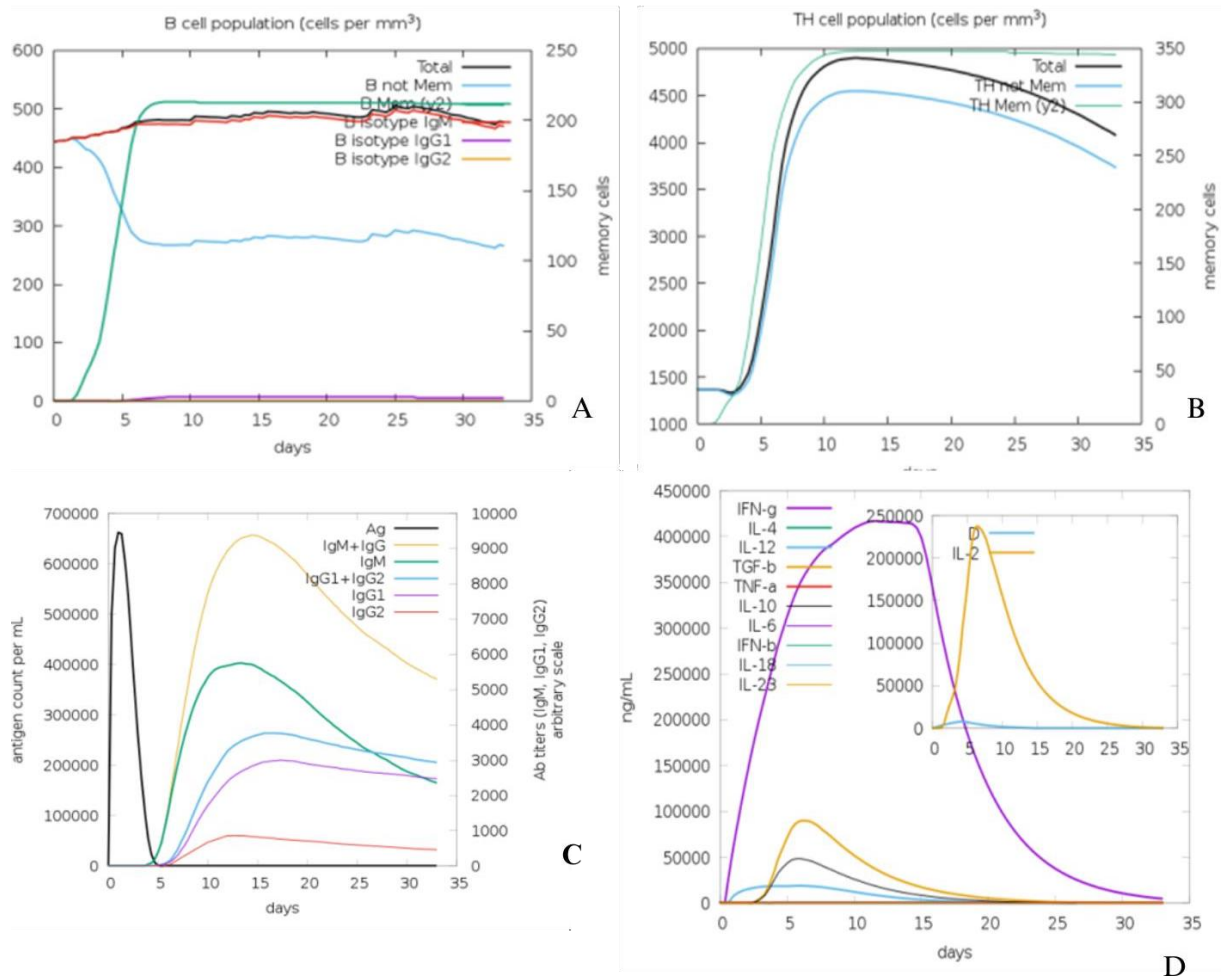


Figure 3. 42: Vaccine 8 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 9

Figure A below shows the b cell population. During the first few 600 antibodies of the antibodies were increasing and after that, after 5 days of the administration it was the highest, which was between 500-600 antibodies and the memory cells were slightly lower than 250 and In the next the graph B) the administration of vaccine 9 showed the gradual increment antibodies and the highest

number of T helper cell cells which was more than 4500 but a static 5000 antibodies after 10 days but after that it was steady for memory cells which was 400 counts for the remaining days. Figure C) below showing the simulation of vaccine 8 administration showed IgM+IgG antigen count was between 600000-700000 between 10 to 15 days, it was the highest then it gradually lowered, and the antigen was at peak during the first 5 days. And before 35 days the antibody titer was more than 8000 but then decreased to 5000 to 6000 counts for IgM+IgG or IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was slightly above 40000 counts during 10 to 15 days and after 20 days it gradually reduced.

Vaccine 9

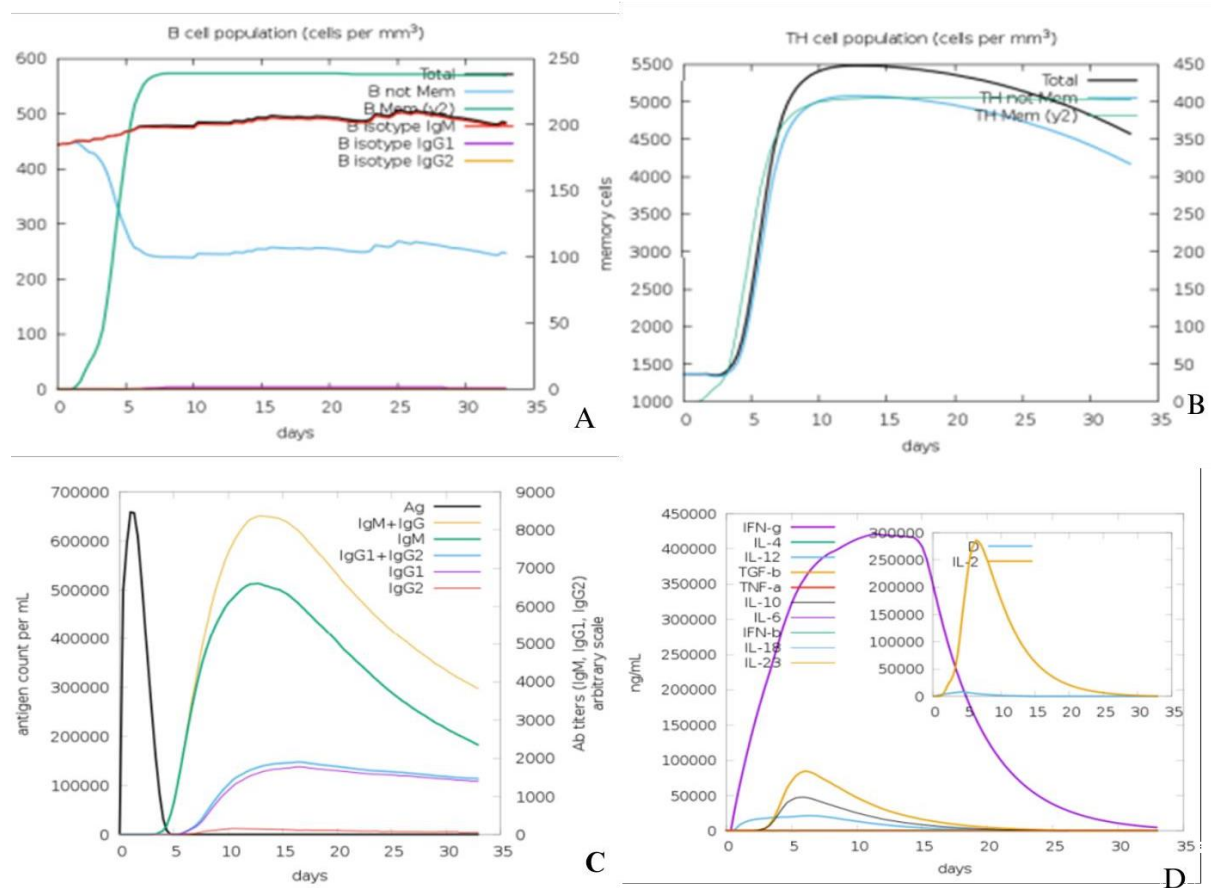


Figure 3.43: Vaccine 10 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 11

Figure A) below shows the b cell population. During the first few days the number of the antibodies increased and after that, it was the highest, which was between 500-600 antibodies and the memory

cells were slightly lower than 250 but more than 200. and In the next figure B), the graph the administration of vaccine 9 showed the gradual increment antibodies and the highest number of T helper cell cells which was between 5000 5500 antibodies after 5 days but after that it was steady for memory cells which little above than 350 and close to 400 counts for the remaining days. Figure C) below is showing the simulation of vaccine 11 administration showed IgM+IgG antigen count was between 600000-700000 between 10 to 20 days, it was the highest on day 15th then it gradually lowered, and the antigen was at peak during the first 5 days which was close to 700000 count. And before 35 days the antibody titer was exceedingly high which was above 10000 and close to 12000 counts then it reduced to 4000 to 6000 counts for IgM+IgG of IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was slightly above 400000 counts and stayed the same during 10 to 15 days, and after 20 days it gradually reduced.

Vaccine 11

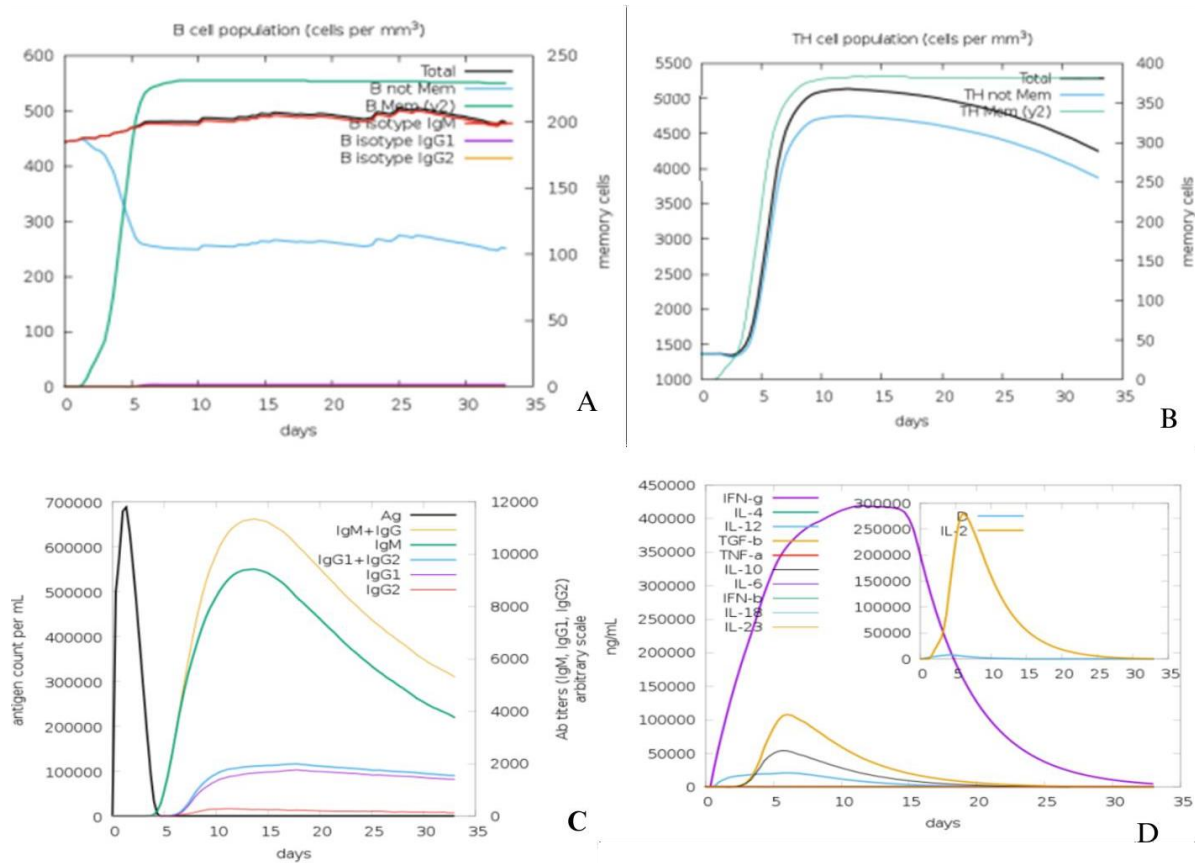


Figure 3. 44: Vaccine 11 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 12

figure A) below shows the b cell population for vaccine 12 administration. During the first few days the number of the antibodies increased and after that it was between 500-600 antibodies but in a stagnant position and the memory cells were slightly lower than 250 but more than 200. and in the next figure B) the graph the administration of vaccine 12 showed the gradual increment of antibodies immediately and the highest number of T helper cell cells which was 5000, it was steady for memory cells which was 400 counts for the remaining days. Figure C) below is showing the simulation of vaccine 12 administration showed IgM+IgG antigen count was between 600000-700000 during the first few days. And before 35 days the antibody titer was more than 12000 but reduced to 6000 for IgM+IgG of IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was moderately above 400000 counts and after 20 days it gradually reduced.

Vaccine 12

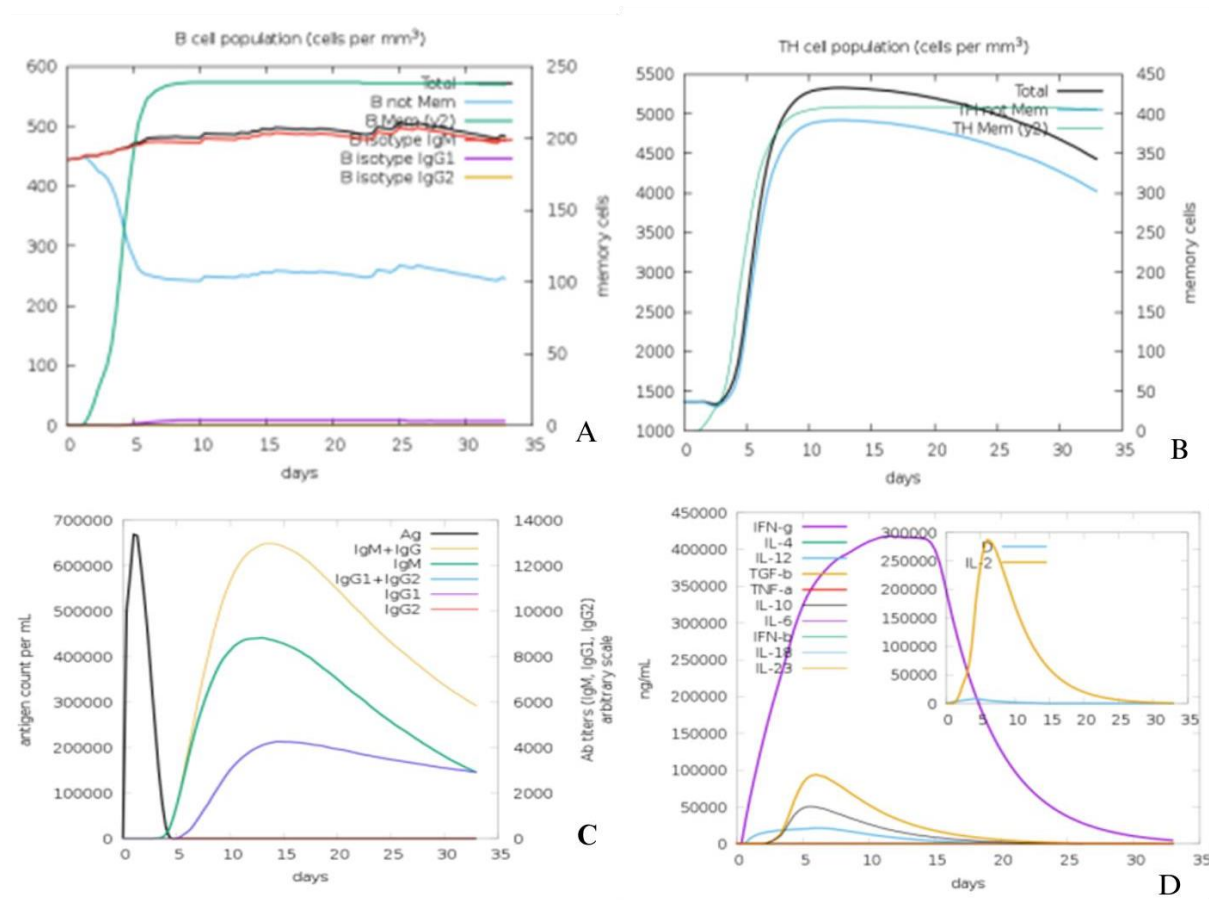


Figure 3.45: Vaccine 12 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 16

Figure A) below shows the b cell population. During the first few days the number of antibodies increased and after that, it was the highest, which was 600 antibodies, and the memory cells were 250 and both were stagnant. In the next figure B), the graph of the administration of vaccine 16 showed the gradual increment of antibodies and the highest number of T helper cell cells which was 5000 antibodies after 5 days but after that it was steady for memory cells which was little above than 400 counts for the remaining days. Figure C) below is showing the simulation of vaccine 16 administration showed IgM+IgG antigen count was between 600000-700000 between 5 to 15 days, it was the highest on day 15th then it gradually lowered, and the antigen was at peak during the first 5 days. And before 35 days the antibody titer was more than 12000 and less than 14000 but then came down to 6000 for IgM+IgG of IgM, IgG1 and IgG2. Besides, the next figure D) also

showed that the number of IFN-gamma cells was slightly above 40000 and little close to 450000 counts and stayed the same during 10 to 15 days, and after 20 days it gradually reduced.

Vaccine 16

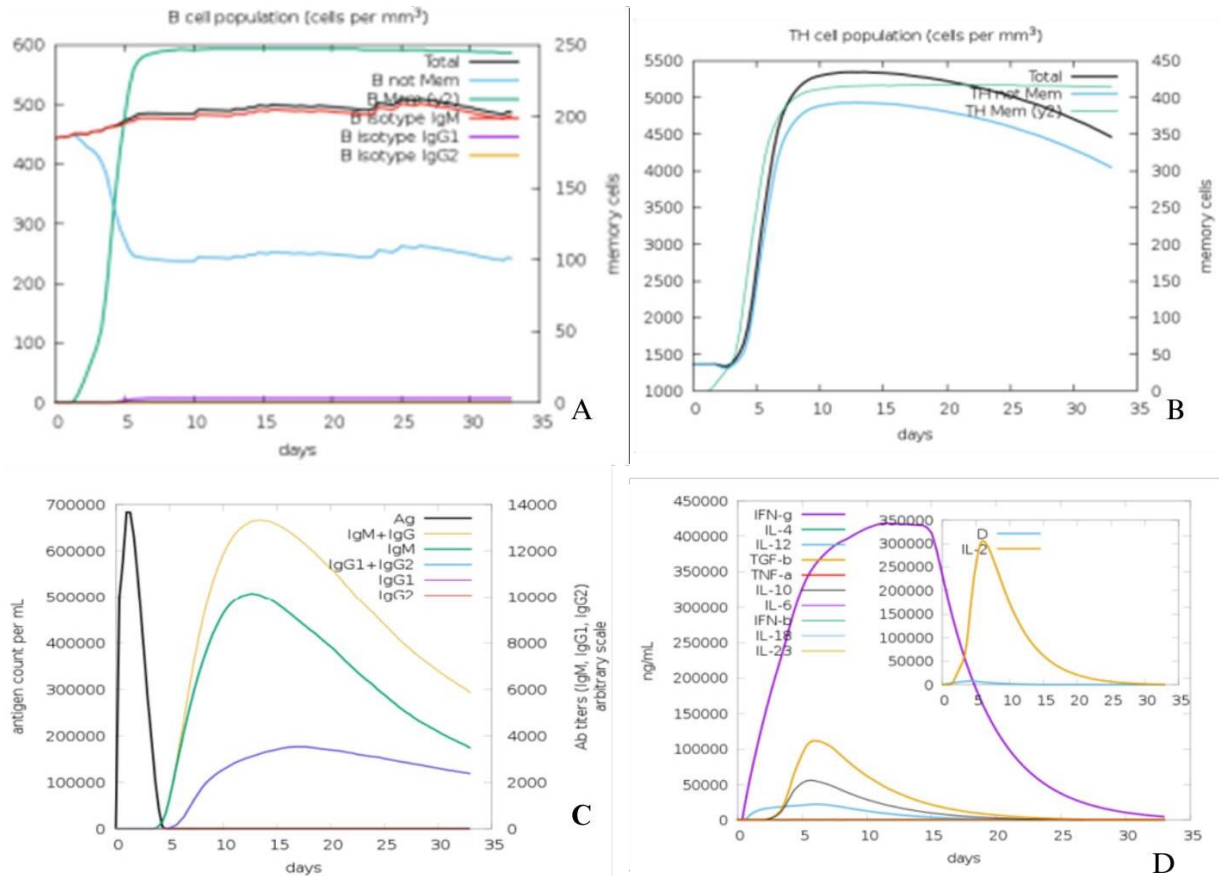


Figure 3.46: Vaccine 16 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Vaccine 18

Figure A) below shows the b cell population. During the first few days the number of the antibodies increased and after that, it was the highest, which was 500 after 5 days and the memory cells were 250. In the next figure B) the graph of the administration of vaccine 18 showed the gradual increment of antibodies and the highest number of T helper cell cells which was 6000 antibodies after 5-6 days but after that it was steady for memory cells which was 450 counts for the remaining days. Figure C) below is showing the simulation of vaccine 16 administration showed IgM+IgG antigen count was increasing after 5 days and it was highest which was between 600000-700000 during 10 to 15 days, and before 35 days the antibody titer was the highest which was 14000 for

IgM+IgG or IgM, IgG1 and IgG2. Besides, the next figure D) also showed that the number of IFN-gamma cells was slightly above 40000 and little close to 450000 counts and stayed the same during 10 to 15 days, and after 20 days it gradually reduced.

Vaccine 18

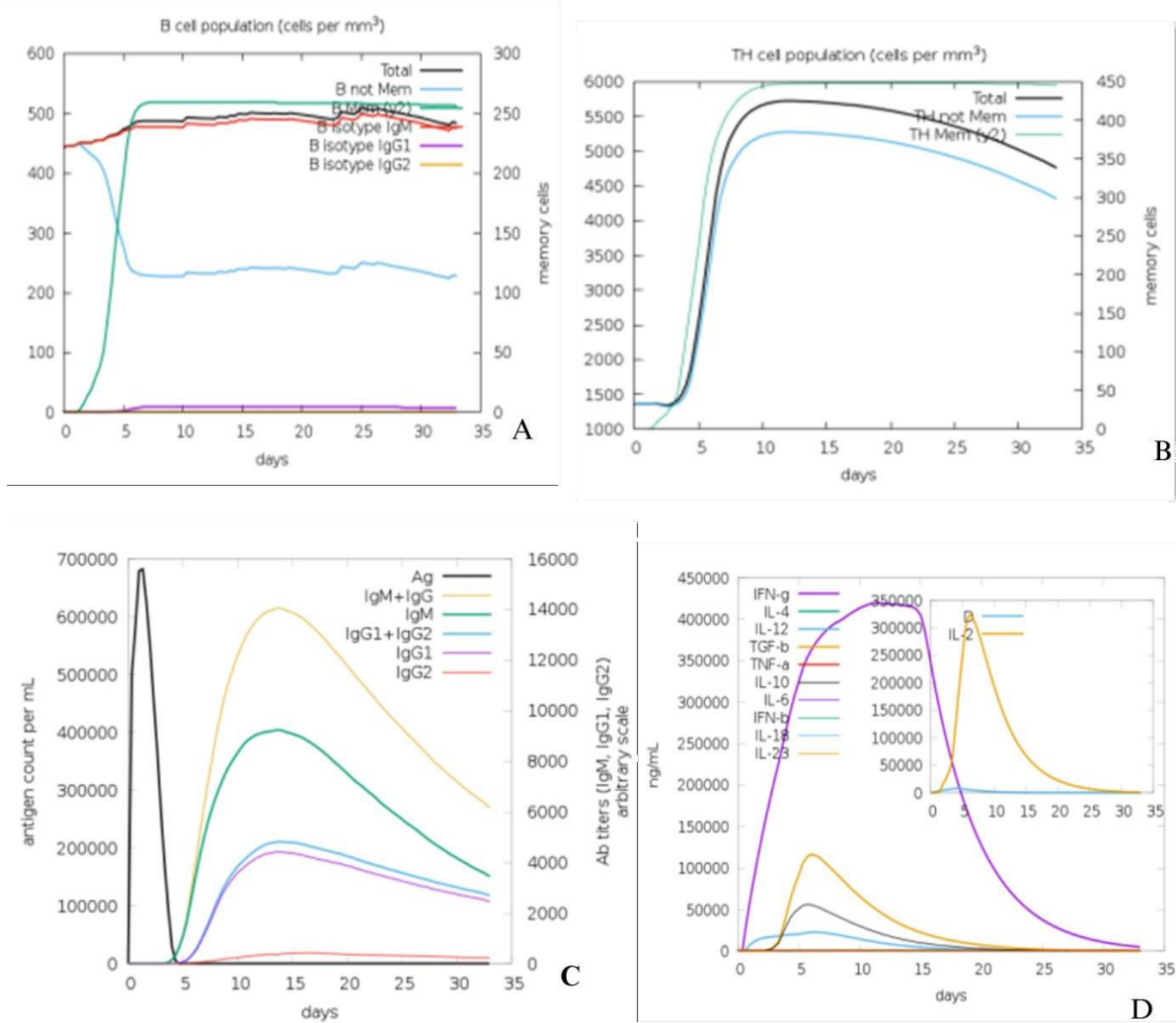


Figure 3.47: Vaccine 18 A) concentration of B cell B) Concentration of helper T cell C) the immunoglobulins and the immune complexes. D) Concentration of cytokines and interleukins

Table below shows the comparative range for the B memory cell, Helper T memory cell, Ag count, IgM+IgG and lastly, IFN-g count for ten selected vaccines.

Table 3. 20: Concentration of B memory cell, Helper T memory cell, Ag count, IgM+IgG and lastly, IFN-g count for ten selected vaccines.

Vaccine number	B mem cell per mm ³		TH mem cell per mm ³		Immunoglobulins and immune complexes count/ml		IFN-g ng/ml
	Antibody count	Memory cell count	Antibody count	Memory cell count	Ag count	IgM+IgG	
2	400-500	200	4500-5000	300-350	700000	7000-8000	400000-450000
4	400-500	200-250	5000	350	600000 - 700000	5000-6000	400000-450000
5	600	200-250	5000-55000	400	700000	12000	400000
6	500	200	5000	350	600000 - 700000	9000	400000
8	500-600	250	4500-5000	400	600000 - 700000	8000	400000
9	500-600	250	5000	400	600000 - 700000	8000-9000	400000-450000
11	500-600	200-250	5000-5500	350-400	700000	10000-12000	400000
12	500-600	200-250	5000	400	600000 - 700000	12000-14000	400000-450000
16	600	250	5000	400-450	700000	12000-14000	400000-450000
18	500	250	6000	450	700000	14000-16000	400000-450000

3.9) Molecular Dynamic Simulation

Molecular simulations were done using GROMAC. Later, molecular dynamic simulations were done for the selected 5 vaccines. Root Mean Square Fluctuation (RMSF), Root Mean Square Deviation (RMSD), Radius of Gyration (RG), Hydrogen bonds and SASA were observed and analyzed to find the best and only candidate among vaccine-5,11,12,16 and 18.

Molecular Dynamic Simulation of vaccine- 11

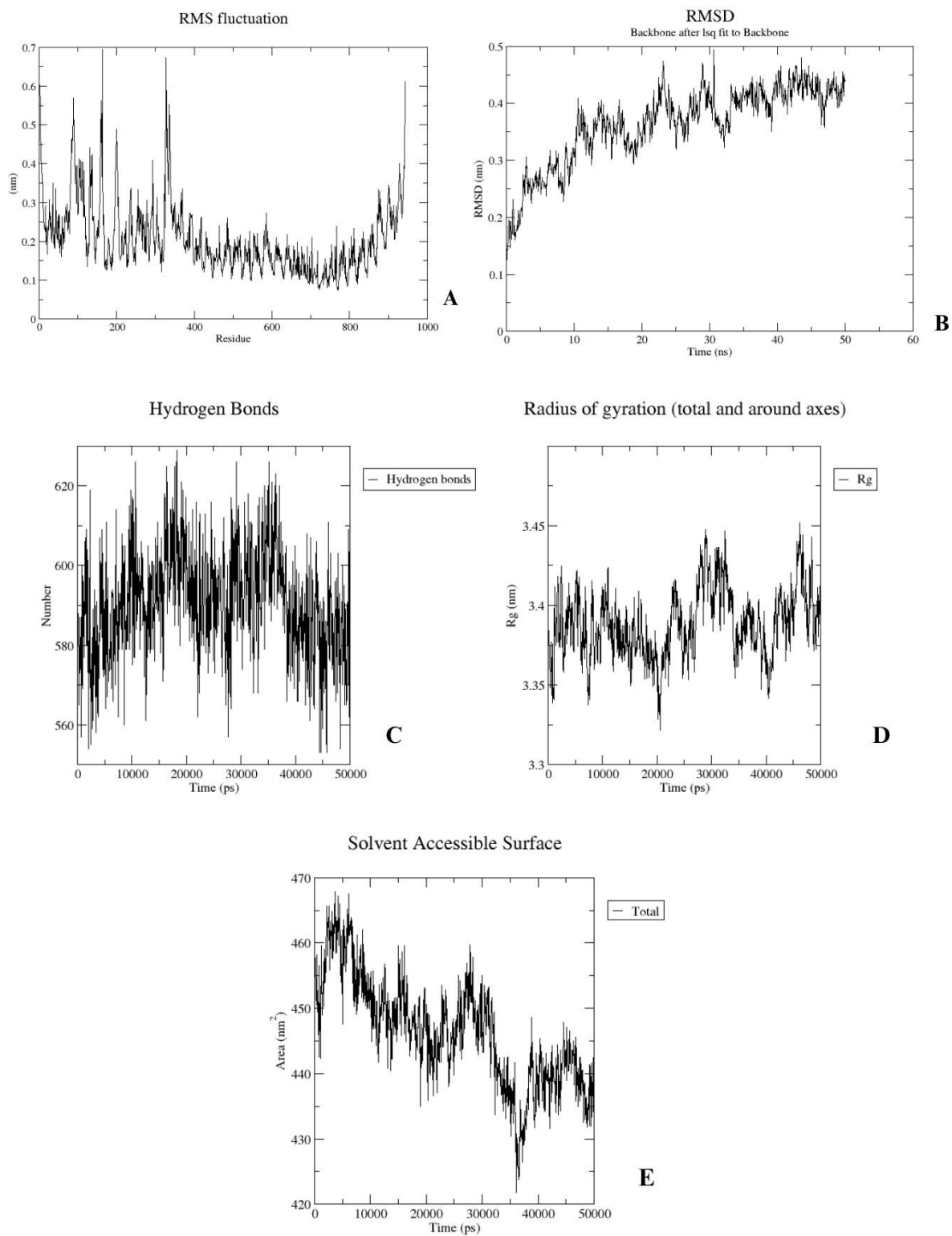


Figure 3. 48: MD Simulation vaccine-11 A) RMSF B) RMSD C)H-Bond D)RG E)SAS

Molecular Dynamic Simulation of vaccine-5

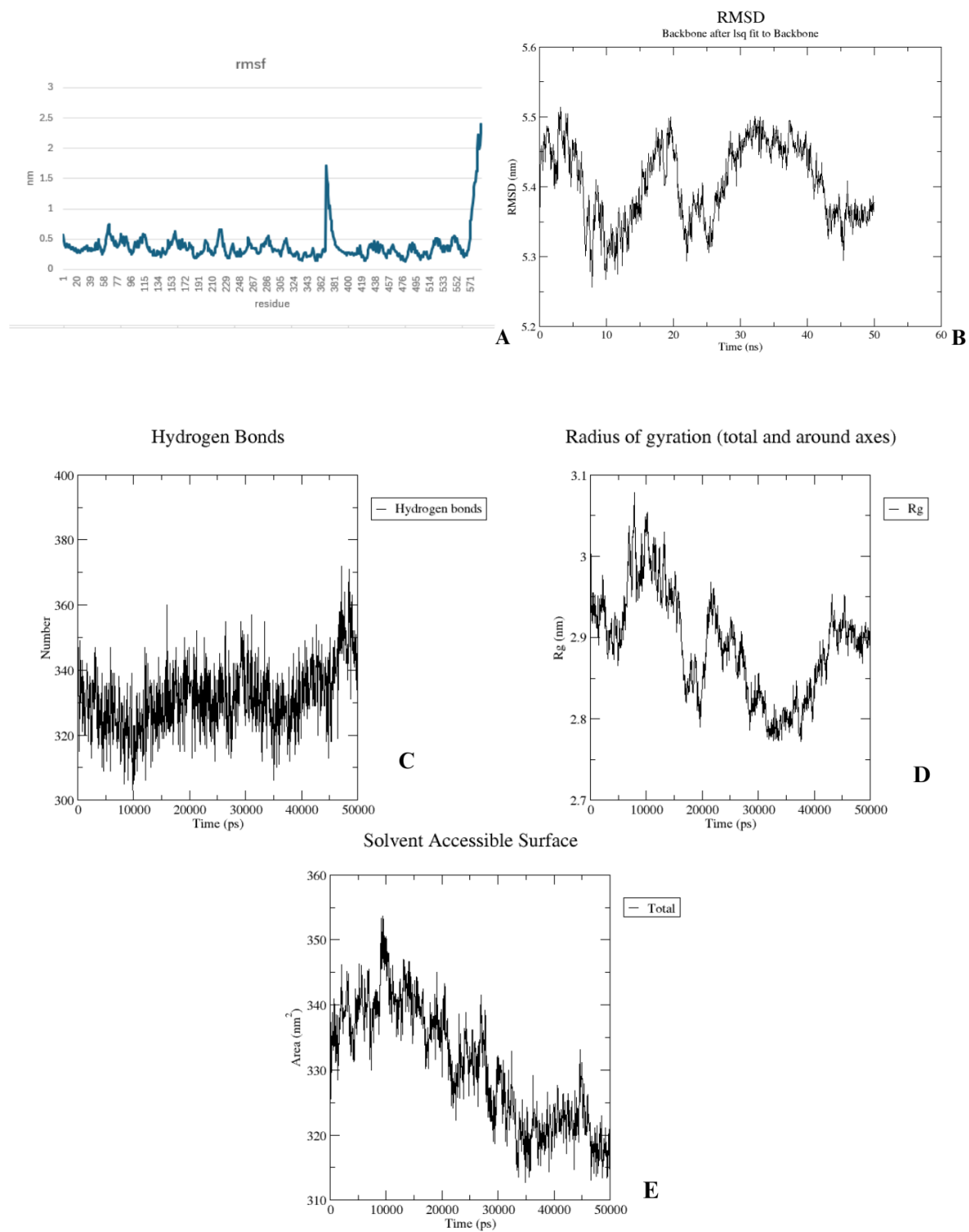


Figure 3. 49: MD Simulation vaccine-5 A) RMSF B)RMSD C)H-Bond D)RG E)SAS

Molecular Dynamic Simulation of vaccine-12

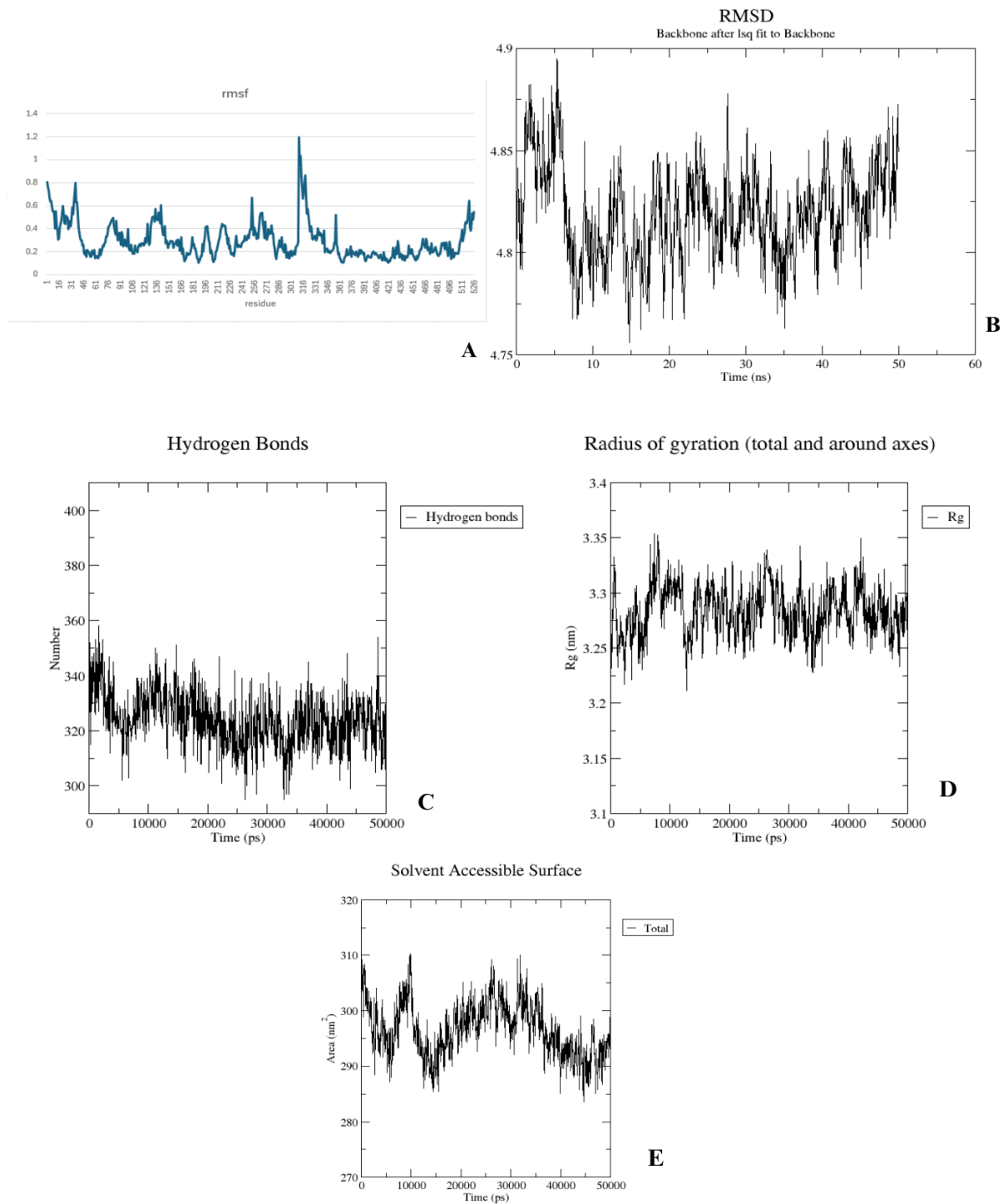


Figure 3.50: MD Simulation vaccine-12 A) RMSF B)RMSD C)H-Bond D)RG E)SAS

Molecular Dynamic Simulation of vaccine-16

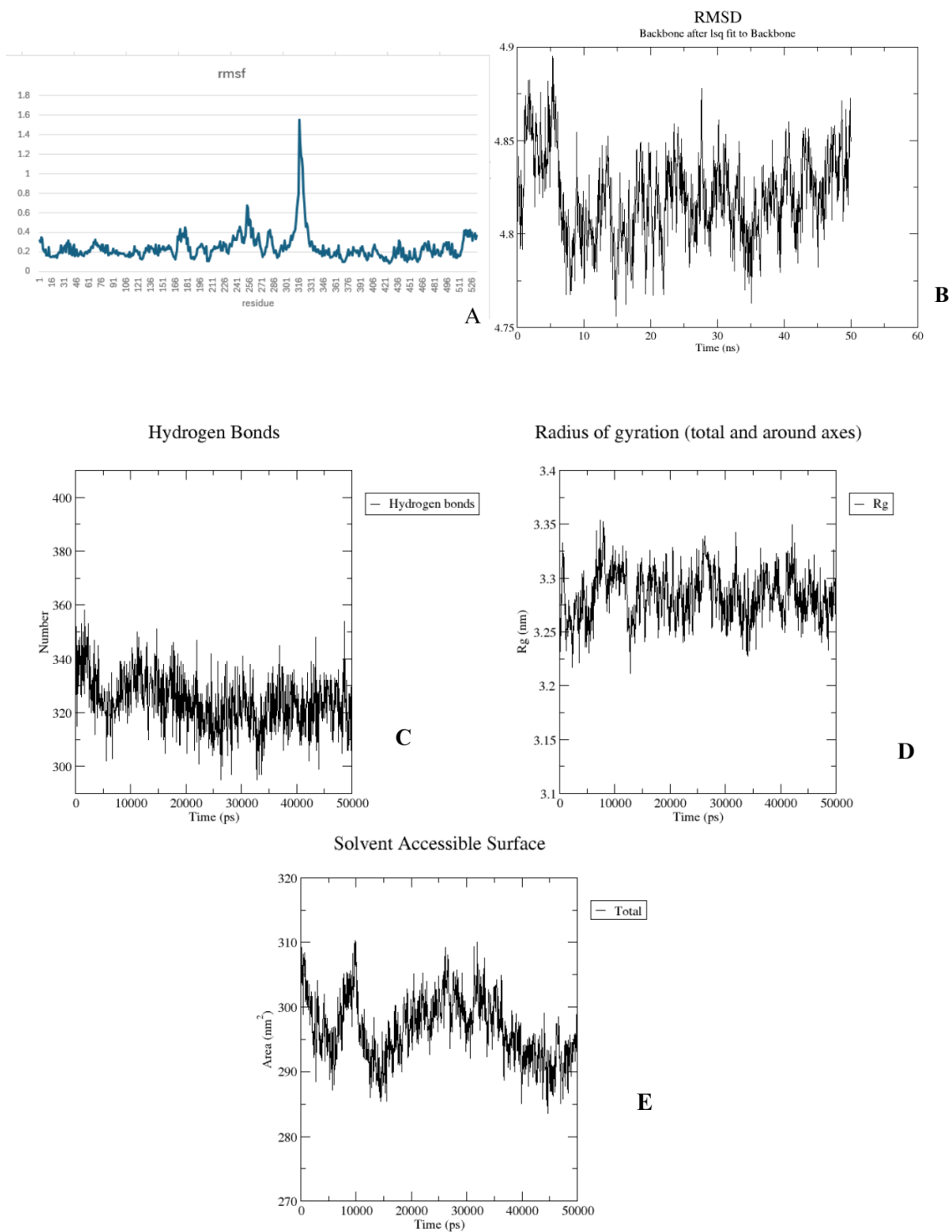


Figure 3. 51: MD Simulation vaccine-16 A) RMSF B) RMSD C) H-Bond D) RG E) SAS

Molecular Dynamic Simulation of vaccine-18

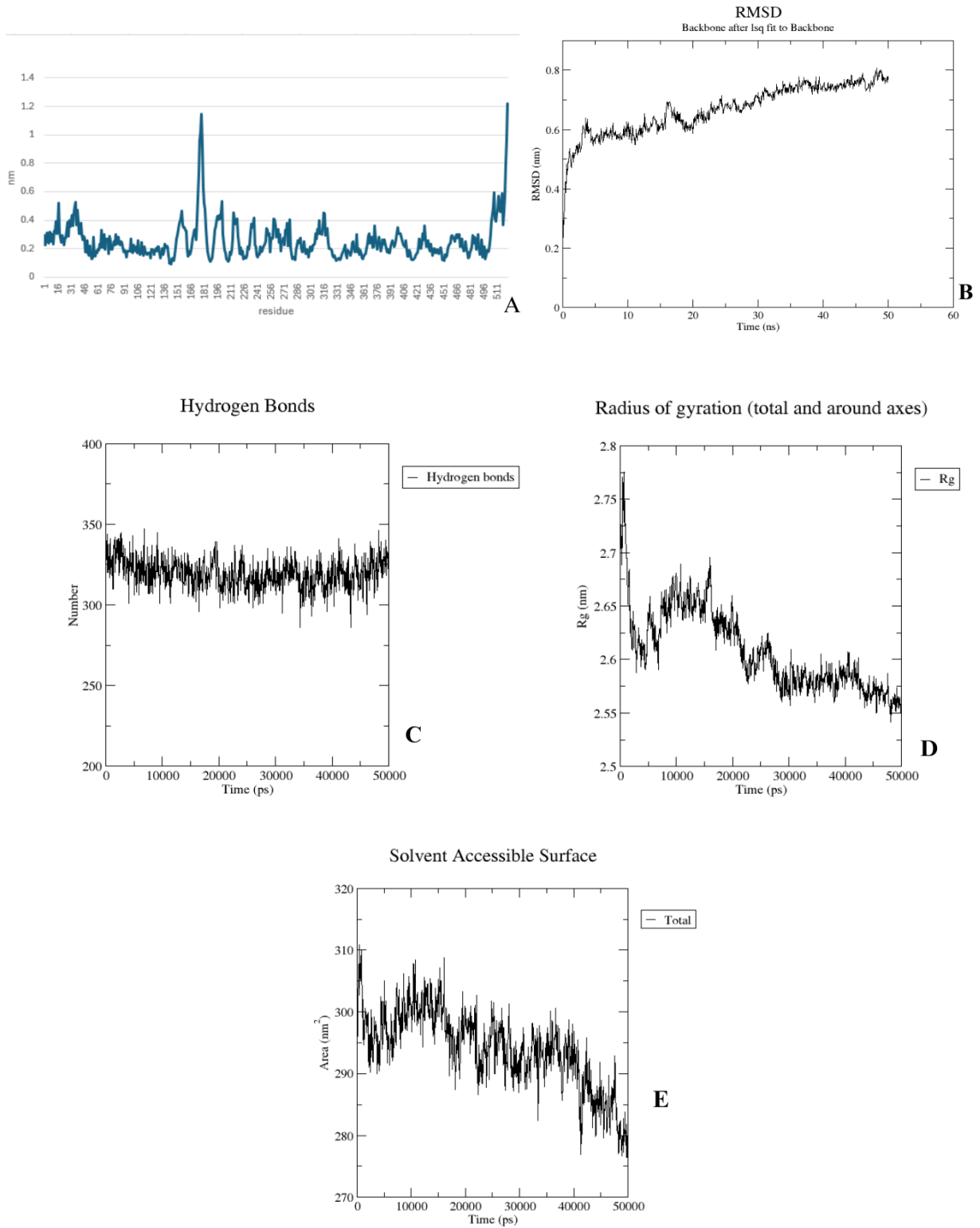


Figure 3. 52: MD Simulation vaccine-18 A) RMSF B) RMSD C) H-Bond D) RG E) SAS

3.9.1) RMS Fluctuation

Vaccine-11 (figure 3.48 A) showed RMS fluctuation less than 0.7 nm or 7Å. Initial residues showed high amount fluctuation and between 300 to 350 residue the peak was high then fluctuation got stable and lower for the flexibility and dynamics between 400 to 800 residues then it again started to increase, several peaks were noticed during the simulation during 100, 200, around 350 and 900 residues. Different types of fluctuated points were shown by vaccine-5 where initial RMSF was lower than 5Å and it didn't increase during the simulation period except for 800 residues where it peaked to 25Å. Vaccine 12 (figure 3.50 A) RMSF analysis showed an increased mark during 300 residues and throughout the simulation it was under 0.5nm or 5Å. Vaccine-16 (figure 3.51 A) showed a similar result that was also between 5Å-6Å. Lastly, Vaccine 18 (figure 3.52A) highlighted the maximum fluctuation under 6Å, although after 199 residues and 800 residues it showed two peaks fluctuating to 10Å. Vaccine 5, instead showed a highly fluctuated RMSF of 30Å made them slightly unstable vaccine constructs than others.

3.9.2) RMSD

vaccine 11 showed conformation changes as it is illustrated in (figure 3.48 B). During the simulation, the average RMSD was initially 2Å. This demonstrated continuous fluctuation from initial structural rearrangements to termination point which was terminally 4.5Å. The vaccine construct started to peak and the highest was close to 4.5Å during 20 to 30 ns, after fluctuation. Vaccine 5 (figure 3.49 B) initially had 55Å, it showed high fluctuation through the simulation and terminally it generated 52Å. Hence, the structure of this construct was distorted. Vaccine 12 (figure 3.50 B) also showed an extremely high RMSD which was within the range of 47.5 to 49Å. The high fluctuation remained in the same range for the whole 50 ns simulation resulting in an unstable structure. Vaccine 16 (figure 3.51B) was also in the same range of 47Å to 49Å. Its initial RMSD was between 48 to 49Å which was the highest during 10 ns. Lastly, vaccines 18 (figure 3.52 B) showed a RMSD less than 8Å, which initiated within the range of 4Å to 6Å and gradually increased to 8Å during the 50ns simulation.

3.9.3) Hydrogen Bond

In the vaccine 11, the illustration (figure 3.48 C) showed the number of the hydrogen bonds forming was almost 620, the highest number was during 20000 picoseconds which had more than

620 bonds. The lowest was seen during 40000 to 50000 picoseconds and terminated within the range between 500 to 600 bonds. Vaccine 5 (figure 3.49 C) started within the range of 320 to 340, the hydrogen bond number was less than 380 in total the lowest was less than 320 during 10000 ps. Terminally it showed little more than 360 hydrogen bonds. Vaccine 12 (figure 3.50 C) initially generated less than 360 hydrogen bonds within ten thousand picoseconds and the lowest was less than 300 which was during 20000 to 30000 picoseconds and it terminated within the range of 320 to 360. Vaccine 16 (figure 3.51 C) showed results closely like vaccine 12 which started within the range of 340 to 360 hydrogen bonds. It formed the lowest number of bonds during 20000 to 30000 picoseconds. vaccine 18 (figure 3. 52 C) generated the same range of hydrogen bond number throughout the 50000 picoseconds simulation, from initial point of to the terminating point which was between 300 to 350 bonds.

3.9.4) RG (Radius of Gyration)

Vaccine 11 (figure 3.48 D) started within the range of 33.5Å to 34Å and the lowest was 33Å during the 20000 picosecond and highest was 34.5Å during 30000 to 40000 picoseconds. It terminated within the same range. Vaccine 5 (figure 3.49 D) the RG was within the range of 29Å to 30Å and it was highest which was 32Å during 10000 PS and lowest during 30000 ps and it terminated with 29Å. Vaccine 12 (figure 3.50 D) fluctuated in the same range which was 32.5Å to 33.5Å. It was highest during 10000 ps and 40000 ps and lowest during 10000-20000 ps which was 32Å. Vaccine 16 (figure 3.51 D) showed the similar result as where the whole simulation of 50000 ns was ranging between the RG of 32.5Å to 33.534Å. Vaccine 18 (figure 3. 52D) initially showed the highest RG which was during 27Å to 28Å and then it decreased gradually and showed lowest RG in the terminating point which was 25.25Å.

3.9.5) SASA

Vaccine 11 (figure 3.48 E) area started within the range 440 to 470 nm² and fluctuated during the whole simulation and lowest fluctuation was during 30000 to 40000 ps and the area was between 420 to 430 nm² and the highest was 470 nm². Vaccine 5 (figure 3.49 E) started within the range of 330 nm² to 350 nm² and gradually decreased to the point of 310 nm² to 320 nm² during 40000 to 50000. Vaccine 12 (figure 3. 50 E) 290 nm² to 310 nm² within 50000 ps. The highest area was 310 nm² during 10000 ps. Vaccine 16 (figure 3.51 E) also started with 300 to

310 nm² area during 10000 ps time. It ended up ranging between 290 to 300 nm². Lastly, the SASA of vaccine 18 (figure 3.52 E) was seen to be at peak during the initial simulation time which was during 10000 ns it was 310 nm² but gradually decreased to 280 nm² at terminating point. It had the lowest peak during 40000 ns that was close to 279 nm² area.

Chapter – 4

Discussion

Zika virus is a global health threat and found in more than 85 countries throughout the world most abundantly in Africa, America, and Asia that declaration of WHO in 2018, it was clearly stated about 159 cases of ZIKV have been detected in the neighboring countries India hence, now Bangladesh, also fell into the category of danger for ZIKV as an emerging threat warning (Hossain et al., 2019). However, according to the WHO landscape the constructed ZIKV vaccines are still on trial, nevertheless, the model vaccine construction would be capable of inducing an immunogenic response against the virus after a single administration among all people and must include safety and highly effective for the pregnant, child and other patients with any medical conditions (Poland et al., 2019). Thus, the study focuses on developing a novel peptide vaccine against the ZIKV virus to prevent deadly infections throughout the world that would work effectively and safely for people.

The Zika virus from the Flaviviridae family is capable of transmitting diseases like microcephaly or other sexual transmission. It undergoes fast mutation due to RNA-dependent-RNA-polymerase, which makes it capable of high adaptation and spreading zoonotic disease, which can cause global emergence. However, the single-stranded RNA-encoded protein of the ZIKV genome is later translated into the polyprotein, which is then divided into structural and nonstructural proteins. Both synonymous and nonsynonymous mutations can occur in the ZIKA, but nonsynonymous mutations were seen to be very prevalent during the outbreak. (Borucki et al., 2019). This is why the emphasis of understanding should project into the evolution and genome study of the virus, which will help to provide a wide perspective of developing a vaccine or any therapeutic targets against it. Moreover, the phylogenetic analysis of the virus gave us an important standpoint of the geographical region that influenced the origin of the ZIKAV. It was seen within the selected most recent 15 variants, only two lineages were observed, 12 of them were Asian, and 3 of them were African. However, the phenomenon of some mutations in those variants or sequences of a particular clade can be a crucial factor for disease control as the connectivity of the pathogenicity of the virus variants with the certain clade is associated, it can also give a potential viewpoint to design vaccines (Seabra et al., 2022).

Additionally, the genes that are subjected to diversification can also be noticed after analyzing the variations within the virus genome. Hence, the initial step in developing an in-silico vaccine against this virus was to retrieve a genome sequence from the BV-BRC server and annotate it to a protein sequence. Later, multiple sequence alignment was done to inspect the conserved region,

which was later cut to extract the desired proteins according to the start and end position of the capsid, envelope, and membrane (table 3.2). The structural proteins were selected due to their significant role of being abundant and easy target antigens in the body. Also, the appropriate organization of conserved structural proteins like capsid, membrane, or envelope can vastly induce an immune response (Sanami et al., 2021).

The sequences are assessed computationally for the physicochemical properties (Table 5). Epitopes were predicted from the IEDB server. Cytotoxic CD8+ T lymphocytes or CTL epitopes are memory responses that help to render the immune reaction persistent. If the vaccine stimulates a regulatory CD8+ cell in an effective manner against ZIKA, it can induce a good antibody reaction and deliver them for protection. In consequence therapeutic vaccines (Kumaraguru & Woolard, 2010). The selection of Helper T cells is one of the most important criteria because it plays a key role in inducing humoral and cell-mediated immune responses. Therefore, to develop a multi-epitope-based vaccine, the selection of HTL receptor-specific epitopes was especially important. For further screening, B cell epitopes were selected, which showed broad-spectrum function for generating antibodies. It is also known as a determinant of antigen where a B cell receptor recognizes a specific antigen, it has two types which are discontinuous or conformational and linear or continuous (Ras-Carmona et al., 2022). However, the main purpose of B cell prediction for the vaccine construction was to induce specific antibodies; also, it was formulated as peptides that synthesize antigen-specific antibody prediction against the ZIKAV. Based on the highest antigenicity, T cell epitopes were selected for capsid, envelope, and membrane, but along with that, the toxicity, allergenicity, and threshold score were checked simultaneously. Three-dimensional image prediction (table 3.14) provides a clear view of the epitopes, which helps to enable the construction of a multi-epitope vaccine that consists of the most effective.

A comparative Population coverage study is done with epitopes based on their binding with the human leukocyte antigen (HLA) alleles to see the highest hits and coverage for certain geographical areas and the study widens the analysis of the coverage rate of immunogenic peptides or epitopes that rendered immunogenic responses in substantial portions among the population in the world along with selected geographical areas hence, identifying the importance of conceivable factors and for developing our multi epitope-based vaccine (Mukherjee et al., 2020). The highest coverage for MHC-I was seen in Europe with 5.26 average hits, and the world has an 88.09% coverage with 5.07 hits (Table 3.15); on the other hand, MHC-II has also

shown the highest coverage in Europe and South Asia which was 99.9% and 99.22%. The selected promiscuous epitope sequences of class I and class II were exerted with their corresponding HLA alleles to find the coverage, as the type of HLA differs from ethnicity to ethnicity, and it is challenging to predict complete consummate coverage due to the extreme polymorphism against antigen (Misra et al., 2011).

For developing a vaccine, selected epitopes were used along with the best-fitted adjuvants and linkers. Notwithstanding, designing a multi-epitope-based vaccine is one of the promising advancements in medicine, and the Research purpose of designing a multi-epitope-based vaccine is for a better comprehensive immunoinformatic analysis that can avert and fight against the virus by generating a large immune response from cytotoxic T cells, helper T cells and B cells. These types of vaccines are comparatively more effective due to their stability and potential to produce both humoral and cellular immune system responses. Additionally, the *in-silico* design of such a structure is considered to have firm interactions with receptors like TLR4 (Alibakhshi et al., 2024). Beta defensin-3, a 50s ribosomal protein (table-1), is used as the adjuvant for the constructed vaccines. However, the fusion protein technology to develop such vaccines can improve immunogenicity at a high rate. A particularly suitable selection of linkers can enhance the affinity and efficacy of the peptides. In this case, β -defensin was used for the vaccine construction for its crucial immune efficacy and demonstrated as a potent immune adjuvant for vaccine development as well as It is an antimicrobial peptide that can play a role as a first line of defense (Zheng et al., 2023). Another adjuvant that was used was the 50s ribosomal protein L7/L12. It is an immunoadjuvant widely used for vaccine construction due to high recognition by TLR4 and its capacity to induce cellular immunity. (Lee et al., 2014).

Empirical Linkers such as EAAAK hold the importance of constructing stable because of the functional and efficient vaccines to make and they often accept linkers for fusion (Chen et al., 2013). Proteins are used to connect the adjuvant with the rest of the protein regions. It is appended to the C-terminal of the adjuvants which are Beta-defensin and 50s ribosomal protein. Also, the GGGGS linker has shown immense potential for the stability of the fusion protein technique (Shamriz et al., 2016) for the CTL linker. Along with this, the AAY linkers also showed an exciting potential to generate a flexible binding site to enhance epitope presentation for joining cytotoxic T epitopes (Dong et al., 2020). The helper T cells were combined with each other with the GPGPG linker, and KK was used as B cell linkers for the epitope linkers to develop immunogenic multi-

epitope-based vaccines. Lastly, 16 of the vaccines were designed to have a Poly histidine tag at the c-terminal of the residues because it is used for its small size, and it confers good affinity and stability (Skiba et al., 2018). Non-specific binding towards untagged protein residues is one of the drawbacks of using a poly-his tag (Bornhorst & Falke, 2000). Among 19 vaccines 3 of the constructs did not have a poly-hist tag adjoined at its c-terminal.

Physicochemical properties (table 3.17) are the ones that influence the vaccine's values and overall efficiency. Allergenicity, antigenicity, solubility, and hydrophobicity were the parameters that were used to find better candidates among primary ones. Moreover, the solubility of a vaccine is a crucial factor when it comes to hypo sensitization or inducing the tolerance of autoantigenic properties (Sestak et al., 2014). Often, vaccines are used to make the body identify and respond to a specific antigen through the immune system, and the size and particles of the vaccines are typically quite small. Therefore, the circulation of these particles from the injection site to the entire system depends a lot on the solubility of the vaccine's peptide sequences. All the constructed vaccines had a solubility of more than 0.45, which means all of them were soluble. Nevertheless, the allergenicity and antigenicity were checked thoroughly so it turned out all the primary constructs were non-allergenic and antigenic. These all are commendatory properties for vaccines. Next, The GRAVY values from the designed vaccines were all negative and ranged between -0.01 to -0.45, this observation indicates that the proteins are very hydrophobic and present a high interaction with water molecules (Droppa-Almeida et al., 2018).

Secondary and tertiary structures of constructed 19 vaccines were predicted to analyze the helices, strands, coils in the protein sequences of vaccines. The advantages of helices are they can provide stability through hydrogen bonds. (Emberly et al., 2002). On the other hand, coil structures are formed due to neutral amino acid residues which do not form the hydrogen bonds; hence, the protein structure loses its structural rigidity. Random coils of the vaccine's tertiary structure were generated so that all 19 of them could observe the helix, strands, and extended strands closely.

From the observations, vaccine-2,4,5,6,8,9,11,12,16 and 18 showed moderate antigenicity with favorable solubility and GRAVY value along with the stable tertiary structure. To understand further, all selected 10 vaccines were used for in-depth analysis by generating a Ramachandran plot via the PDBSUM tool (table 3.18). The ERRAT values were also checked for the validation of the constructs. However, ERRAT is determined by crystallography, and it is used to find the errors in the residues by using the 9-residue window (Pandey et al., 2016). Overall quality depends

on how close the value is to 100. Vaccine-2,5,8,9 and 16 showed more than 90% ERRAT value. However, the Ramachandran plot was again checked for further observation of the distribution of the dihedral phi/Psi angles and the stereochemical quality of the proteins (Elslinger & Wilson, 2012). It envisioned both the disallowed region and the allowed region to distinguish between the poor and stable quality of the homology model. The mediocre quality is usually found more in the forbidden region, in the fourth quadrant, due to its steric clashes, structural problems, and bad conformation. On the other hand, the favored region or allowed region includes the amino acids that do not have steric hindrance and dihedral confirmation of both alpha-helix and beta-sheet (Aarthy & Singh, 2022). In the Ramachandran plot, the red regions between the yellow regions of the second and third quadrants represent the favored region of the plot, and it resembles conformations, which have stability and better dihedral angles. Most of the constructed vaccines, like vaccines 2,5,8,9,11,12, and 18, showed more than 98% of the allowed region and less than 2% of the disallowed region.

Then, molecular docking was done on the vaccines using CLusPro2.0 against TLR4. TLR4 was selected as the receptor for the docking analysis and plays a key role in the immune response. (Koike, 2018) TLR4 works as a host immune system and first-line defensive molecule, and it assists immunity by enlarging the number of both antigen-processing and antigen-presenting cells. Hence, TLR4 gives more stability and better scores during the molecular docking simulation than other receptors by allowing better RMSD tolerance (Yahaya et al., 2021). Nevertheless, ligand-receptor docking identifies the structural and predominant binding models by successfully providing a three-dimensional structure based on the docking score function for each candidate of vaccines. (Morris & Lim-Wilby, 2008). From (table 3.19), the docking score is represented by the center and lowest energy with the target receptor TLR4. Approximate prediction of both receptor-ligand models included the conformation orientation, and interaction of potential sites decoded the stability of your vaccine-receptor models. Vaccines 2,4,8,9,18 consecutively showed the lowest Gibbs free energy, which was respectively -15.6, -16.3, -15.5, -16.1, and -15.6 kcal/mol calculated from the Prodigy web server. However, if the ΔG values which are within the range of -5 to -15 kcal/mol are commonly perceived as marginally stable that dominate the folding mechanism of protein (Kar et al., 2020).

However, TLR4 can identify the PMPs, and the principal role is to lead the innate to arrange the adaptive immune response (Sorokina et al., 2022). Therefore, when the interaction occurs the dock.

scoring function predicts the binding affinities of the two molecules based on models, clusters, and properties.

The presence of hydrogen bonds between the vaccine and the receptor enhances the binding affinity of the ligand, representing the intramolecular strength and its direct association of inducing the immune response because of guiding the conformation of helix or sheets of the peptide sequences that prompt various immune responses (Reyes et al., 2017). These maximum and consistent hydrogen bonds between the vaccine and TLR4 provide a wider inspection of better exhibition of residual interaction and affinity (Moin et al., 2023).

A fluctuation plot was generated for the structural simulation of the designed vaccine for predicting the conformational changes and residues. The RMSF for each residue was represented, and residual-fluctuation profiles were envisioned through a graph. Fluctuation occurs due to the distances of the residues of the amino acids of the protein, the greater the distances between the amino acids, the more distortion, hence, the less native conformation state and instability. The vaccine constructed with TLR4 receptor generated several types of fluctuation, nevertheless, both rigidity and flexibility of a protein are necessary in terms of understanding the wide view of structural fold, dynamic functions, and alterations (Sljoka, 2021). Vaccines 8 and 2 showed a high fluctuation for both chain A and B with high degree of RMSF. Their performances showed that the native conformations were disrupted and. The individual residue flexibility was absent during the simulation and resulted in dissociation. On the other hand, vaccines 5, 19, 12, 16, 18 showed less quantity of structural fluctuation. It means, the atoms of those vaccines had similarities to the reference structure and less displacement that allowed them to be stable. The fluctuated structure of the residues of chain A and B involved in rigid-body alignment which includes the native conformational changes (Martínez, 2015).

To calculate the distances and interaction between the amino acids of the protein sequences, analyze the network of the tertiary and secondary interaction of the chain, ten contact maps were generated via the RING server. Additionally, interpretation of the residue-residue contact can play a crucial role in predicting accuracy of protein threading contact (Bhattacharya & Bhattacharya, 2020). From the generated contact map for 10 selected vaccines, it was seen that amino acids between the area of chain A and chain B interact with the adjacent line. The Y axis was the intersection line and the second and fourth quadrant represented the interaction of both Chain A and B. (table-3.20) showed the residues of amino acids of chain A and B.

Immune simulation was also done by computation approach using C-immsim server and the importance of inspecting the immunological process onset of the constructed vaccine is exceedingly high. However, the concentration of B cell, helper T cell, immunoglobulins, and immune complex and lastly, concentration of cytokines and interleukins were independently generated to observe the agents and populations in the server generated and simulated biological environment of the body (Rapin et al., 2011). Only one dose of vaccines was selected as a vaccine schedule to simulate.

The vaccine must produce a humoral immune system along with antibody and antibody mediated cells. It was seen that B cell antibody count was highest 600 cells/mm³ and memory cell count was highest 250 cells per mm³. For every vaccine, the antibody count and the memory count were the same average count, which was between 400 and 600 antibodies and 200 to 250 memory cell counts. Development of such consistent concentration depicts a stable concentration for B cell population. Among vaccine 5, 6, 11, 12, 16 and 18 was observed to generate steady results for the population and induced the same count for B cell population from day 5 to days 35 holding the closest 600 count/ml for vaccine 5 and respectively 500, between (500-600), Between (500-600), 600, 500 cells/mm³ for rest five vaccines and it was the highest. And memory cell count for the B memory cell was also very consistent to generate quick antigen identification and induce fast response.

An injection of a single dose vaccine can trigger the immune system by inducing helper T cells and can be observed by the adaptive immunity by number of cells count and the memory cell induction. The highest helper T cell count was around 6000 cell/mm³ and the lowest was 4500. Vaccine 18 only developed a maximum concentration of helper T cells which was 6000 cells/mm³. Among 10 vaccines 5, 6, 12, 16 and 18 showed a fixed and strong count of Helper T cell and the for the rest 35 days and was around 5000 and 6000 cells/mm³. Other vaccines provoked the cell count to be a very stable condition too. The memory cell production was high for the vaccine 16 and 18 which was 450 cells/mm³.

Immunoglobulins and the immune complexes play a significant role in the body. It binds to the invader and fastens the immune response by advancing the opsonization, activation and neutralization. It was observed that the IgM+IgG population count was up to 700000 cells/mm³ for the antigen count for 12000. All 10 vaccines had generated the highest amount of

immunoglobulins IgM+IgG count between 600000-700000 cells/mm³ count/ml. The antigen count was highest for the vaccine 5,11,12,16 and 18 that crossed 10,000 cells/mm³. Hence, it was considered as the most effectful candidate. Also, the all ten vaccines that have been selected can be selected only if it produced relatively high antigen dose to efficient immune response and optimum antibody stators to prevent the virus attack and prevent it from transmitting cell to cell (Billeskov et al., 2019) The antigen count was set to identify the best candidate due to the ability of the simulation of vast immune response by inducing macrophages, T cells, B cells along with this, it generate a wide range of chemical substances like cytokines, interleukins immunoglobulin like IgG, IgM and other immune complexes (Chuekwon & Cheng, 2023). The IFN-gamma was also checked because it signals instantly balance the spectrum of CD8+ T cells and cytolytic capability (Alspach et al., 2019) and the (table 3.21)showed almost all vaccine showed a perfect and moderate amount of IFN- γ that could produce enough amount of helper T cell as immune response and the count was between 400000-450000 cells/mm³.

Based on the result, the best 5 vaccine candidates were chosen according to the validation, immune and structural simulation results. Vaccine 5,11,12,16 and 18 were selected. That included ERRAT, Ramachandran plot favored region, docking score, Gibbs free energy and hydrogen bond. The vaccine 5 had an antigen count of 12000 cells/mm³ count and IgG and IgM count was 700000 cells/mm³. It also had a moderate RMS fluctuation of amino acids residues index with less than 5 angstroms. The average center docking energy was -968.7 kcal/mol and Gibbs free energy was -10.5 Kcal/mol. It had 8 hydrogen bonds for intramolecular bindings. The vaccine had more than 98% of the favored region and ERRAT value of 94.97%. Vaccine 11 RMSF was also less than 4 angstroms with -1132.9 kcal/mol docking score and Gibbs free energy of -14.3 kcal/mol with 17 hydrogen energy. It has 98.5% allowed region in the Ramachandran plot and 86.9% ERRAT value. However, it generated close to 12000 cells/mm³ antigen count. Vaccine 12 resulted in 12000-14000 cells/mm³ antigen count with 98.8% favored region of Ramachandran plot. It has the docking energy of -1395.8 kcal/mol and -13.1kcal.pol Gibbs free energy. It had 7 hydrogen bonds. It showed great stability in the fluctuation plot with less than 4 angstrom RMSF during structural simulation. Both 16 and 18 signified higher stability in fluctuation plot and balanced structure by generating less than 4 angstrom RMSF that demonstrated unchanged native conformational changes. Vaccine 8 had crossed 14000 cells/mm³ antigen count and 98.5% favored region in Ramachandran plot with 85.4167% ERRAT value. Additionally, the docking score was -1260.0

kcal/mol and the Gibbs free energy was -15.6kcal/mol with 14 hydrogen bonds. However, vaccine 16 has 94.6% favored region and 91.11% ERRAT value. It had -1111.5 kcal/mol docking score with 11.9 kcal/mol Gibbs free energy with 8 hydrogen bonds.

Root means square fluctuation is analyzed for the effect of the amino acid residues and protein flexibility and rigidity function under specific temperature, pressure, and volume over the simulation period (Likić et al., 2005). Around 5 peaks were noticed during the simulation under 7Å. Vaccine 5 peaked at only terminal point, Vaccine 12 (figure 3.50 B) and 16 (figure 3.51 B) showed closely similar results with 2 peaks during the range of 200 to 400 residue. (Kalita et al., 2020) Average rms fluctuation value 3.90Å were considered to be stable in subsequent result analysis. vaccine 18 showed only a fluctuating peak that crossed 10Å during the simulation, but other residues were compact and stable under 5Å, and it can be concluded, vaccine 11 and 18 had better rms fluctuation and exhibited better structural stability.

To find the stability of the complex of the vaccine with the receptor TLR4, RMSD was conducted to determine the different variation of their conformational changes when they undergo interactions (Zeyauallah et al., 2023). Higher RMSD value demonstrated the greater deviation in protein conformation in the md simulation. Usually greater than 10Å indicates deformation or unstable native conformation structure due to small molecules, and structure instability. vaccine 11 (figure 3.48 B) demonstrated continuous fluctuation from initial structural rearrangements to termination point which was terminally 4.5Å. Vaccine 5 structure was distorted due to extremely elevated levels of RMSD. Vaccine 12 generated a high RMSD of 49Å, with. The high fluctuation remained in the same range for the whole 50 ns simulation resulting in an unstable structure. Vaccine 16 (figure 3.51 B) showed the same result hence from the trajectories, it had structural distortions. Lastly, vaccines 18 (figure 3.52 B) showed a RMSD less than 8Å. Nevertheless, vaccine-11 rms deviation of structure showed a stable state and better protein folding performance. The RMSD of md simulation emphasize on the deviation for the atoms and molecules of the vaccines and logan complex, hence, a high value would be inferencing as highly unstable structure and deficient amount in the md simulation through the time (Al-Karmalawy et al., 2021).

Evaluate the stability and depict the degree of strength and balance of intermolecular and intramolecular bindings H-bond is calculated. In the vaccine 11, the illustration (figure 3.48C) showed the number of the hydrogen bonds forming was almost 620, the highest number was during 20000 picoseconds which had more than 620 bonds. The lowest was seen during 40000 to 50000

picoseconds and terminated within the range between 500 to 600 bonds. Vaccine 5 (figure 3.49 C) started within the range of 320 to 340, the hydrogen bond number was less than 380 in total the lowest was less than 320 during 10000 ps. Terminally it showed little more than 360 hydrogen bonds. Vaccine 12 (figure 3.50 C) initially generated less than 360 hydrogen bonds within 10000 picoseconds and the lowest was less than 300 which was during 20000 to 30000) picoseconds and it terminated within the range of 320 to 360. Vaccine 16 (figure 3. 51 C) showed results closely like vaccine 12 which started within the range of 340 to 360 hydrogen bonds. It formed the lowest number of bonds during 20000 to 30000 picoseconds. Vaccine 18 generated the same range of hydrogen bond number throughout the 50000-picosecond simulation, from initial point of to the terminating point which was between 300 to 350 bonds.

Radius of gyration is the compactness of the vaccine construct and calculated by the protein backbone during the md simulation. It was compact during the simulation when the RG was found to be ranging between 21 Å to 22 Å (Rampogu et al., 2022) Vaccine 11 (figure 3.48 D) highest was 34.5Å during. However, Vaccine 5 was slightly less, and the RG was within the range of 29Å to 30Å. Vaccine 12 (figure 3.50 D) and 16 (figure 3.51 D) fluctuated in the same range which was 32.5Å to 33.5Å. Vaccine 18 (figure 3.52 D) initially showed the highest RG which was during 27Å to 28Å. The less the RG, the more the structural compactness and more the RG, less the structural compactness. Comparing the radius to find the more compactness during simulation vaccine 18 showed a considerable result.

The extracellular area on the protein molecule surface which has the most proximal distance to access solvent, hence, accessibility of the solute surface to interact the solvent exhibits the degree of interactions by calculating SASA (solvent accessible surface area). It is considered to indicate the protein stability by predicting the centralized solvent sphere interacting with van der Waals connected with the surface (Borjian Boroujeni et al., 2021). Vaccine 11 (figure 3. 48 E) area started within the range 440 to 470 nm². However, vaccine 5 (figure 3.49 E) started within the range of 330 nm² to 350 nm². Vaccine 12 (figure 3.50 E) 290 nm² to 310 nm² within 50000 ps, hence, it occupied less area than vaccine 11. Vaccine 16 (figure 3.51 E) also started with 300 to 310 nm² area during 10000 ps time. Lastly, the SASA of vaccine 18 was generated high but over the time it decreased to a considerate level. Standard SASA for the molecular simulation is considered for better compactness and stability of the vaccine when it is ranging less than 200 nm². The SASA

profile of the constructed vaccine showed less flexibility and stability but vaccine 18 achieved a better rigidity of binding with ligand molecules than others.

Based on the results of RMSD, RMSF, Rg, Hydrogen bond and SASA, vaccine 11 and vaccine 18 showed prominent results. Vaccine 11 had highest hydrogen bonds established as both intra and intermolecular implanting better strength. The number was 620 and for vaccine 18 hydrogen bond number 400. Vaccine 11 did not cross 7Å rms fluctuation and had less than 5Å rmsd. However, vaccine 18 had peaks that crossed 10Å in RMSF calculation during the simulation and had 8Å RMSD. The SASA and Rg were respectively 34.5Å and 28Å; 470nm² and 310 nm² for the vaccine 11 and 18. The structural rigidity thus can be considered as better and stable for the variants of unchanged native conformations and dynamics for vaccine 11. All the proposed vaccine candidates were observed to be simulated for 50ns. The results were satisfactory for some parameters among all 5 and few showed pitfalls due to some reasons. The proposed approach must include longer simulation time, advanced empirical force-field parameterization required as water model structure could be employed at advanced level for the complex constructs. However, the MD simulation is computationally intensive so whole protein conformation cannot be sampled in one space hence, the result depends on the total simulation time. The improvisation could bring better results of the vaccine during md simulation unless, of course, the protein properties do not entirely depend on the total time although a large number of MD run can bring better reproducibility of molecular system by tracking down the trajectories and individual atoms and molecule for more in depth and dynamic experimental data (Cavalcanti da Silveira et al., 2020).

Chapter – 5

Conclusion

Zika virus, a member of the Flavivirus genus, is a potential threat to humans. After several outbreaks quickly transmitted among many continents, it has become an international concern, and Bangladesh is at risk due to its long persistence and expanded host range. Hence, the study, an *in-silico* approach to developing a multi-peptide vaccine against the deadly ZIKV, aims to suggest the potential adequacy of the constructed vaccine as a novel vaccine candidate against all the recent variants of ZIKV worldwide. The unavailability of Zika virus vaccines and the lack of approved vaccines after years of research make the study more relevant and important. Structural proteins like membranes, envelopes, and capsids were selected to generate the epitopes due to the more accessible and efficacious target in the body, resulting in the epitopes being highly antigenic and non-allergic. The highest population coverage was also worldwide, covering the maximum threshold level. The previous study used the most potent adjuvants and linker molecules to adjoint the constructs; all 19 primarily developed vaccines later showed standard physiological properties. Furthermore, the secondary and tertiary structure was analyzed for each one to select them as suitable candidates based on antigenicity and other physiological factors; consequently, ten favored vaccine constructs were selected. In addition to that, the validation ERRAT and Ramachandran plots were checked respectively for each of the ten vaccines to compare. After the validation, molecular docking was employed with the TLR4 receptor, and each of them showed a hand docking score and high binding affinities. Afterward, structural simulations were used to study the key function of the protein's interaction and evaluate its stability. A contact map and fluctuation plot were generated to analyze each vaccine-receptor complex's atomic motion, residual distance, and substantial integrity. Besides, immune stimulation ascertained the vaccines that could generate both humoral and cell-mediated immune responses at a high rate and in an extended manner by administering only one dose. Finally, based on that, vaccines 5,11,12,16, and 18 picked as the strong candidates. Lastly, the MD simulation was run. However, 5,12, and 16 showed high instability, whereas vaccines 11 and 18 showed potential dominance in structural integrity over others. After observing rigidity, flexibility, and integrity parameters, vaccine 11 was chosen as the best candidate and strongly recommended as the most suitable vaccine construct. Nevertheless, predicting the characteristics and properties of the construct computationally can only be concluded to have explanatory efficiency if it is analyzed further in vitro and in-vivo assessment for the several phases of clinical approval. Only then can it be employed for humans, and the study would be a handful.

Reference:

1. Aarthy, M., & Singh, S. K. (2022). Envisaging the conformational space of proteins by coupling machine learning and molecular dynamics. In *Advances in Protein Molecular and Structural Biology Methods*. <https://doi.org/10.1016/B978-0-323-90264-9.00028-3>
2. Al-Karmalawy, A. A., Dahab, M. A., Metwaly, A. M., Elhady, S. S., Elkaeed, E. B., Eissa, I. H., & Darwish, K. M. (2021). Molecular Docking and Dynamics Simulation Revealed the Potential Inhibitory Activity of ACEIs Against SARS-CoV-2 Targeting the hACE2 Receptor. *Frontiers in Chemistry*. <https://doi.org/10.3389/fchem.2021.661230>
3. Alibakhshi, A., Alagheband Bahrami, A., Mohammadi, E., Ahangarzadeh, S., & Mobasheri, M. (2024). In-silico design of a new multi-epitope vaccine candidate against SARS-CoV-2. *Acta Virologica*. <https://doi.org/10.3389/av.2023.12481>
4. Alspach, E., Lussier, D. M., & Schreiber, R. D. (2019). Interferon γ and its important roles in promoting and inhibiting spontaneous and therapeutic cancer immunity. *Cold Spring Harbor Perspectives in Biology*. <https://doi.org/10.1101/cshperspect.a028480>
5. Baek, M., Dimaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., Degiovanni, A., Pereira, J. H., Rodrigues, A. V., Van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). transformer改进1 : Accurate prediction of protein structures and interactions using a 3-track network. *BioRxiv*.
6. Bhattacharya, S., & Bhattacharya, D. (2020). Evaluating the significance of contact maps in low-homology protein modeling using contact-assisted threading. *Scientific Reports*. <https://doi.org/10.1038/s41598-020-59834-2>
7. Billeskov, R., Beikzadeh, B., & Berzofsky, J. A. (2019). The effect of antigen dose on T cell-targeting vaccine outcome. In *Human Vaccines and Immunotherapeutics*. <https://doi.org/10.1080/21645515.2018.1527496>
8. Borjian Boroujeni, M., Shahbazi Dastjerdeh, M., Shokrgozar, M. A., Rahimi, H., & Omidinia, E. (2021). Computational driven molecular dynamics simulation of keratinocyte growth factor behavior at different pH conditions. *Informatics in Medicine Unlocked*. <https://doi.org/10.1016/j.imu.2021.100514>
9. Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. In *Methods in Enzymology*. [https://doi.org/10.1016/s0076-6879\(00\)26058-8](https://doi.org/10.1016/s0076-6879(00)26058-8)
10. Borucki, M. K., Collette, N. M., Coffey, L. L., Van Rompay, K. K. A., Hwang, M. H., Thissen, J. B., Allen, J. E., & Zemla, A. T. (2019). Multiscale analysis for patterns of Zika virus genotype emergence, spread, and consequence. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0225699>
11. Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-7-153>

12. Bullard-Feibelman, K. M., Govero, J., Zhu, Z., Salazar, V., Veselinovic, M., Diamond, M. S., & Geiss, B. J. (2017). The FDA-approved drug sofosbuvir inhibits Zika virus infection. *Antiviral Research*. <https://doi.org/10.1016/j.antiviral.2016.11.023>
13. Byler, K. G., Ogungbe, I. V., & Setzer, W. N. (2016). In-silico screening for anti-Zika virus phytochemicals. *Journal of Molecular Graphics and Modelling*. <https://doi.org/10.1016/j.jmglm.2016.08.011>
14. Cavalcanti da Silveira, C. O., Gonçalves, A. da S., Costa Franca, T. C., & Silva Filho, E. A. (2020). Computational studies of mucin 2 and its interactions with thiolated chitosans: a new insight for mucus adhesion and drug retention. *Journal of Biomolecular Structure and Dynamics*. <https://doi.org/10.1080/07391102.2019.1610499>
15. Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. In *Advanced Drug Delivery Reviews*. <https://doi.org/10.1016/j.addr.2012.09.039>
16. Chou, P. Y., & Fasman, G. D. (2006). Prediction of the Secondary Structure of Proteins From Their Amino Acid Sequence. In *Advances in Enzymology and Related Areas of Molecular Biology*. <https://doi.org/10.1002/9780470122921.ch2>
17. Chuekwon, K., & Cheng, L. T. (2023). Immune Response to the Dual Antigen Vaccine of *Actinobacillus pleuropneumoniae* in a Mouse Model. *Trends in Sciences*. <https://doi.org/10.48048/tis.2023.6798>
18. Clementel, D., Del Conte, A., Monzon, A. M., Camagni, G. F., Minervini, G., Piovesan, D., & Tosatto, S. C. E. (2022). RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac365>
19. Dar, H., Zaheer, T., Rehman, M. T., Ali, A., Javed, A., Khan, G. A., Babar, M. M., & Waheed, Y. (2016). Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An *in silico* approach. *Asian Pacific Journal of Tropical Medicine*. <https://doi.org/10.1016/j.apjtm.2016.07.004>
20. Desta, I., Porter, K., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and Its Limits in Rigid Body Protein-Protein Docking. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3537797>
21. Dikhit, M. R., Ansari, M. Y., Vijaymahantesh, Kalyani, Mansuri, R., Sahoo, B. R., Dehury, B., Amit, A., Topno, R. K., Sahoo, G. C., Ali, V., Bimal, S., & Das, P. (2016). Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus: A first step towards vaccine development. *Infection, Genetics and Evolution*. <https://doi.org/10.1016/j.meegid.2016.08.037>
22. Dong, R., Chu, Z., Yu, F., & Zha, Y. (2020). Contriving Multi-Epitope Subunit of Vaccine for COVID-19: Immunoinformatics Approaches. *Frontiers in Immunology*. <https://doi.org/10.3389/fimmu.2020.01784>
23. Doytchinova, I. A., & Flower, D. R. (2007). Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine*. <https://doi.org/10.1016/j.vaccine.2006.09.032>

24. Droppa-Almeida, D., Franceschi, E., & Padilha, F. F. (2018). Immune-informatic analysis and design of peptide vaccine from multi-epitopes against *Corynebacterium pseudotuberculosis*. *Bioinformatics and Biology Insights*. <https://doi.org/10.1177/1177932218755337>
25. Elsliger, M. A., & Wilson, I. A. (2012). Structure validation and analysis. In *Comprehensive Biophysics*. <https://doi.org/10.1016/B978-0-12-374920-8.00110-7>
26. Emberly, E. G., Wingreen, N. S., & Tang, C. (2002). Designability of α -helical proteins. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.162105999>
27. Gupta, A. K., Kaur, K., Rajput, A., Dhanda, S. K., Sehgal, M., Khan, M. S., Monga, I., Dar, S. A., Singh, S., Nagpal, G., Usmani, S. S., Thakur, A., Kaur, G., Sharma, S., Bhardwaj, A., Qureshi, A., Raghava, G. P. S., & Kumar, M. (2016). ZikaVR: An Integrated Zika Virus Resource for Genomics, Proteomics, Phylogenetic and Therapeutic Analysis. *Scientific Reports*. <https://doi.org/10.1038/srep32713>
28. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0073957>
29. H. Bekker, H. Berendsen, E.J. Dijkstra, S. Achterop, R. Drunen, & D. Van Der Spoel. (1993). Gromacs: a parallel computer for molecular dynamics simulations – *ScienceOpen. Physics Computing*.
30. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: A web tool for predicting protein solubility from sequence. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btx345>
31. Høie, M. H., Kiehl, E. N., Petersen, B., Nielsen, M., Winther, O., Nielsen, H., Hallgren, J., & Marcatili, P. (2022). NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac439>
32. Honorato, R. V., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. J. J. (2021). Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Frontiers in Molecular Biosciences*. <https://doi.org/10.3389/fmolb.2021.729513>
33. Hossain, M. G., Nazir, K. H. M. N. H., Saha, S., & Rahman, M. T. (2019). Zika virus: A possible emerging threat for Bangladesh! In *Journal of Advanced Veterinary and Animal Research*. <https://doi.org/10.5455/javar.2019.f385>
34. Kalita, P., Padhi, A. K., Zhang, K. Y. J., & Tripathi, T. (2020). Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microbial Pathogenesis*. <https://doi.org/10.1016/j.micpath.2020.104236>
35. Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*. <https://doi.org/10.1038/s41598-020-67749-1>
36. Koike, N. (2018). The role of stem cells in the hepatobiliary system and in cancer development:

- A surgeon's perspective. In *Stem Cells and Cancer in Hepatology: From the Essentials to Application*. <https://doi.org/10.1016/B978-0-12-812301-0.00011-6>
37. Kumaraguru, U., & Woolard, S. N. (2010). Viral vaccines and CTL response. In *Journal of Biomedicine and Biotechnology*. <https://doi.org/10.1155/2010/141657>
 38. Kuriata, A., Gierut, A. M., Oleniecki, T., Ciemny, M. P., Kolinski, A., Kurcinski, M., & Kmiecik, S. (2018). CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky356>
 39. Lazear, H. M., & Diamond, M. S. (2016). Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. *Journal of Virology*. <https://doi.org/10.1128/jvi.00252-16>
 40. Lee, S. J., Shin, S. J., Lee, M. H., Lee, M. G., Kang, T. H., Park, W. S., Soh, B. Y., Park, J. H., Shin, Y. K., Kim, H. W., Yun, C. H., Jung, I. D., & Park, Y. M. (2014). A potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 enhances dendritic cells-based tumor immunotherapy. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0104351>
 41. Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkab301>
 42. Louten, J. (2016). Virus Structure and Classification. In *Essential Human Virology*. <https://doi.org/10.1016/b978-0-12-800947-5.00002-8>
 43. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac240>
 44. Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0119264>
 45. Misra, N., Panda, P. K., Shah, K., Sukla, L. B., & Chaubey, P. (2011). Population coverage analysis of T-Cell epitopes of *Neisseria meningitidis* serogroup B from Iron acquisition proteins for vaccine design. *Bioinformatics*. <https://doi.org/10.6026/97320630006255>
 46. Moin, A. T., Rani, N. A., Ullah, M. A., Patil, R. B., Robin, T. B., Nawal, N., Zubair, T., Mahamud, S. I., Sakib, M. N., Islam, N. N., Khaleque, M. A., Absar, N., & Shohael, A. M. (2023). An immunoinformatics and extended molecular dynamics approach for designing a polyvalent vaccine against multiple strains of Human T-lymphotropic virus (HTLV). *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0287416>
 47. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-59745-177-2_19
 48. Mukherjee, S., Tworowski, D., Detroja, R., Mukherjee, S. B., & Frenkel-Morgenstern, M. (2020). Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets. *Vaccines*.

<https://doi.org/10.3390/vaccines8020290>

49. Nguyen, M. N., Krutz, N. L., Limviphuvadh, V., Lopata, A. L., Gerberick, G. F., & Maurer-Stroh, S. (2022). AllerCatPro 2.0: a web server for predicting protein allergenicity potential. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac446>
50. Nicolson, G. L., Lotan, R., & Rios, A. (1981). Heterogeneous in vitro sensitivities of metastatic B16 melanoma sublines and clones to retinoic acid or BCNU. *Cancer Treatment Reports*. [https://doi.org/10.1016/s0022-5347\(17\)52699-3](https://doi.org/10.1016/s0022-5347(17)52699-3)
51. Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., ... Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac1003>
52. Pandey, R. K., Prajapati, P., Goyal, S., Grover, A., & Prajapati, V. K. (2016). Molecular Modeling and Virtual Screening Approach to Discover Potential Antileishmanial Inhibitors Against Ornithine Decarboxylase. *Combinatorial Chemistry & High Throughput Screening*. <https://doi.org/10.2174/1386207319666160907100134>
53. Poland, G. A., Ovsyannikova, I. G., & Kennedy, R. B. (2019). Zika Vaccine Development: Current Status. In *Mayo Clinic Proceedings*. <https://doi.org/10.1016/j.mayocp.2019.05.016>
54. Ramacciotti, E., Agati, L. B., Aguiar, V. C. R., Wolosker, N., Guerra, J. C., de Almeida, R. P., Alves, J. C., Lopes, R. D., Wakefield, T. W., Comerota, A. J., Walenga, J., & Fareed, J. (2019). Zika and Chikungunya Virus and Risk for Venous Thromboembolism. *Clinical and Applied Thrombosis/Hemostasis*. <https://doi.org/10.1177/1076029618821184>
55. Rampogu, S., Lee, G., Park, J. S., Lee, K. W., & Kim, M. O. (2022). Molecular Docking and Molecular Dynamics Simulations Discover Curcumin Analogue as a Plausible Dual Inhibitor for SARS-CoV-2. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms23031771>
56. Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0009862>
57. Rapin, N., Lund, O., & Castiglione, F. (2011). Immune system simulation online.

Bioinformatics. <https://doi.org/10.1093/bioinformatics/btr335>

58. Ras-Carmona, A., Lehmann, A. A., Lehmann, P. V., & Reche, P. A. (2022). Prediction of B cell epitopes in proteins using a novel sequence similarity-based method. *Scientific Reports*. <https://doi.org/10.1038/s41598-022-18021-1>
59. Rawal, G., Yadav, S., & Kumar, R. (2016). Zika virus: An overview. *Journal of Family Medicine and Primary Care*. <https://doi.org/10.4103/2249-4863.197256>
60. Reyes, C., Moreno-Vranich, A., & Patarroyo, M. E. (2017). The role of pi-interactions and hydrogen bonds in fully protective synthetic malaria vaccine development. *Biochemical and Biophysical Research Communications*. <https://doi.org/10.1016/j.bbrc.2017.01.077>
61. Ruivinho, C., & Gama-Carvalho, M. (2023). Small non-coding RNAs encoded by RNA viruses: old controversies and new lessons from the COVID-19 pandemic. In *Frontiers in Genetics*. <https://doi.org/10.3389/fgene.2023.1216890>
62. Sager, G., Gabaglio, S., Sztul, E., & Belov, G. A. (2018). Role of host cell secretory machinery in Zika virus life cycle. In *Viruses*. <https://doi.org/10.3390/v10100559>
63. Sanami, S., Alizadeh, M., Nosrati, M., Dehkordi, K. A., Azadegan-Dehkordi, F., Tahmasebian, S., Nosrati, H., Arjmand, M. H., Ghasemi-Dehnoo, M., Rafiei, A., & Bagheri, N. (2021). Exploring SARS-COV-2 structural proteins to design a multi-epitope vaccine using immunoinformatics approach: An in silico study. *Computers in Biology and Medicine*. <https://doi.org/10.1016/j.compbimed.2021.104390>
64. Seabra, S. G., Libin, P. J. K., Theys, K., Zhukova, A., Potter, B. I., Nebenzahl-Guimaraes, H., Gorbalenya, A. E., Sidorov, I. A., Pimentel, V., Pingarilho, M., De Vasconcelos, A. T. R., Dellicour, S., Khouri, R., Gascuel, O., Vandamme, A. M., Baele, G., Cuypers, L., & Abecasis, A. B. (2022). Genome-wide diversity of Zika virus: Exploring spatio-temporal dynamics to guide a new nomenclature proposal. *Virus Evolution*. <https://doi.org/10.1093/ve/veac029>
65. Sestak, J. O., Fakhari, A., Badawi, A. H., Siahaan, T. J., & Berkland, C. (2014). Structure, size, and solubility of antigen arrays determines efficacy in experimental autoimmune encephalomyelitis. *AAPS Journal*. <https://doi.org/10.1208/s12248-014-9654-z>
66. Shamriz, S., Ofoghi, H., & Moazami, N. (2016). Effect of linker length and residues on the structure and stability of a fusion protein with malaria vaccine application. *Computers in Biology and Medicine*. <https://doi.org/10.1016/j.compbimed.2016.06.015>
67. Sharma, A., & Lal, S. K. (2017). Zika virus: Transmission, detection, control, and prevention.

- In *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2017.00110>
68. Shragai, T., Tesla, B., Murdock, C., & Harrington, L. C. (2017). Zika and chikungunya: mosquito-borne viruses in a changing world. In *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.13306>
 69. Skiba, M. A., Maloney, F. P., Dan, Q., Fraley, A. E., Aldrich, C. C., Smith, J. L., & Brown, W. C. (2018). PKS–NRPS Enzymology and Structural Biology: Considerations in Protein Production. In *Methods in Enzymology*. <https://doi.org/10.1016/bs.mie.2018.01.035>
 70. Sljoka, A. (2021). Structural and Functional Analysis of Proteins Using Rigidity Theory. In *Sublinear Computation Paradigm: Algorithmic Revolution in the Big Data Era*. https://doi.org/10.1007/978-981-16-4095-7_14
 71. Sorokina, I., Mushegian, A. R., & Koonin, E. V. (2022). Is Protein Folding a Thermodynamically Unfavorable, Active, Energy-Dependent Process? In *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms23010521>
 72. The PyMOL Molecular Graphics System. (2010). The {PyMOL} Molecular Graphics System, Version~1.3. In There is no corresponding record for this reference.
 73. Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L. C., Vanden Eynden, E., Vandamme, A. M., Deforche, K., & De Oliveira, T. (2019). Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty695>
 74. Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky1006>
 75. Wang, Y., Ling, L., Zhang, Z., & Marin-Lopez, A. (2022). Current Advances in Zika Vaccine Development. In *Vaccines*. <https://doi.org/10.3390/vaccines10111816>
 76. Wang, Z., & Ma'ayan, A. (2016). An open RNA-Seq data analysis pipeline tutorial with an example of reprocessing data from a recent Zika virus study. *F1000Research*. <https://doi.org/10.12688/F1000RESEARCH.9110.1>
 77. Wiwanitkit, V. (2016). Zika virus infection: Control and prevention. In *Journal of the Chinese Medical Association*. <https://doi.org/10.1016/j.jcma.2016.04.002>
 78. Yahaya, M. A. F., Bakar, A. R. A., Stanslas, J., Nordin, N., Zainol, M., & Mehat, M. Z. (2021). Insights from molecular docking and molecular dynamics on the potential of vitexin as an

antagonist candidate against lipopolysaccharide (LPS) for microglial activation in neuroinflammation. *BMC Biotechnology*. <https://doi.org/10.1186/s12896-021-00697-4>

79. Zeyauallah, M., Khan, N., Muzammil, K., AlShahrani, A. M., Khan, M. S., Alam, M. S., Ahmad, R., & Khan, W. H. (2023). In-silico approaches for identification of compounds inhibiting SARS-CoV-2 3CL protease. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0284301>
80. Zheng, J., Yang, J., Zhang, Z., Liang, X., Liu, S., Pan, Y., Wei, J., Huang, Y., Huang, X., & Qin, Q. (2023). An improved oral vaccine with molecular adjuvant β -defensin protects grouper against nervous necrosis virus infection. *Fish and Shellfish Immunology*. <https://doi.org/10.1016/j.fsi.2023.108709>