In Silico Comparative Study of Zika Virus Proteins and Analysis of Membrane glycoprotein M as a Candidate for Vaccine Design

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE IN BIOTECHNOLOGY

Submitted by

AFIA NOWSHEN

Student ID: 13136006

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University, Bangladesh

Submitted on

July 26, 2018

DECLARATION

I hereby declare that the research work embodying the analysis and results reported in the following thesis entitled “In Silico Comparative Study of Zika Virus Proteins and Analysis of Membrane glycoprotein M as a Candidate for Vaccine Design”, submitted by the undersigned has been carried out under the supervision of Dr. M. Mahboob Hossain, Professor, Department of Mathematics and Natural Sciences, BRAC University, Dhaka and Ms. Abira Khan, Lecturer, Department of Genetic Engineering and Biotechnology, University of Dhaka. It is further declared that the research work presented here is original and no part of this thesis has been submitted to any other institution for any degree or diploma.

_____

(Afia Nowshen)

Candidate

Certified:

_____

(Dr. M. Mahboob Hossain)

Supervisor

Professor, Microbiology Program, Department of Mathematics and Natural Sciences

BRAC University, Dhaka

Acknowledgement

I wish to declare my most humble gratitude to the Almighty Allah, who has blessed me with the gift of outmost mercy, and given me the strength, patience and understanding, required for the completion of this thesis work.

I convey my special thanks to Professor A.F.M Yusuf Haider, the Chairperson of the Department of Mathematics and Natural Sciences, late Professor A. A. Ziauddin Ahmed, former Chairperson of the Department of Mathematics and Natural Sciences and Professor Naiyyum Choudhury, former coordinator of the Biotechnology and Microbiology Program, for giving me their outstanding guidance and sharing their valuable experiences during my study at BRAC University. I am grateful to all the faculty members of Department of Mathematics and Natural Sciences for the incredible support and valuable teachings throughout the period of my bachelor's studies. I also thank the authority and management of BRAC University for continuously providing me with organized opportunities and facilities.

I express my gratitude to my supervisor, Professor Dr. M. Mahboob Hossain, Department of Mathematics and Natural Sciences, BRAC University, for believing in me and always giving me immense inspiration for my work. I am indebted to Ms. Abira Khan, Lecture, Department of Genetic Engineering and Biotechnology, Dhaka University for being my guide. Throughout this journey, Ms. Abira has inspired me to explore new ideas and bring out the best results from this research. Her constant help was essential to fulfil the aim of my research. She has my earnest appreciation.

I would like to thank my family for always being there and keeping trust in me. This journey would be incomplete without the presence of some dear friends. They are indeed, very precious gifts BRAC University has given me.

This thesis, my very first research work is an important milestone in my career. I will try to implement the knowledge I have gained in best possible way and continue to work on the improvement of the idea that I carry.

Sincerely,

Afia Nowshen

Department of Mathematics and Natural Sciences, BRAC University

Abstract

Zika virus is a (+) RNA virus that has become a major concern of public health in the last few years. It is known to cause microcephaly, reduction of brain circumference along with many type of brain injury in new-borns at a high rate. An effective vaccine against Zika virus is not yet discovered. In this investigation, a comparative study of Zika virus with its other relative species was performed. In order to do so, polyprotein sequences of 10 viruses (West Nile Virus, Rubella Virus, Dengue Virus 2, Human Immunodeficiency Virus 1 (Partial Polyprotein), Human Immunodeficiency Virus 1 (Partial Polyprotein), Ebola Virus (Structural Polyprotein), Chikungunya Virus (Non-Structural Polyprotein), Hepatitis C virus polyprotein, Cytomegalovirus Polyprotein) were analysed by Multiple Sequence Alignment using T Coffee (EMBL) and COBALT (NCBI) respectively. From the analysis, the most conserved regions among species were further analysed using InterProScan and ScanProSite for finding out their precise function and involvement. The overall biological processes, molecular function, and cell component characteristics were also identified. A phylogenetic tree considering 100% data coverage was also constructed.

As an effort to find out an epitope for a probable peptide vaccine design, ZIKA virus Membrane Glycoprotein (M) was analysed because this was the only complete protein sequence submitted to NCBI so far. Using three distinct antigenicity and epitope prediction software, Vaxijen, BepiPred and BCPREDS five sequences were found to be antigenic. Among the five sequences, a seven amino acid long sequence 'GSSTSQK' and 12 amino acid long ENWIFRNPGFAL were determined to be most antigenic by Vaxijen (Antigenicity of GSSTSQK- 1.1341 and ENWIFRNPGFAL- 0.8161 where threshold value of antigenicity $\geq$ 0.4). These sequences have a high antigenicity, conservancy and surface accessibility suggesting that, these might be epitopes that are most compatible for designing a peptide vaccine.

Here a complete comparative study of Zika virus with its relatives were conducted to get insights about how Zika virus works. Alongside, identification of peptide vaccine candidate epitopes was done using bioinformatic tools.

Table of Contents

Table of Contents

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Sl	Short form	Abbreviation
1	ZIKV	Zika Virus
2	CHIKV	Chikungunya Virus
3	HIV	Human Immunodeficiency Virus
4	HCV	Hepatitis C Virus
5	CMV	Cytomegalo Virus
6	WNV	West Nile Virus
7	DENV	Dengue Virus
8	RNA	Ribonucleic Acid
9	BLAST	Basic Local Alignment Search Tool
10	MSA	Multiple Sequence Alignments
11	WHO	World Health Organization
12	CDC	Centre for Disease Control
13	EMBL	European Molecular Biology Laboratory
14	NCBI	National Centre for Biotechnology Information
15	FDA	Food and Drug Administration
16	IEDB	Immune Epitope Database and Analysis Resources

INTRODUCTION & LITERATURE REVIEW

Introduction and Literature Review

1.1 Overview:

A virus is a very small infectious particle that has no biological activity of its own. However, when it enters a susceptible host cell, it can reproduce by directing the host cell machinery and exert pathogenic stress on the host. Most of the virus has a small piece of DNA or RNA as its genetic material that can be single of double stranded. The nucleic acid is wrapped inside a protein shell called capsid. The entire infectious virus particle is called a virion. Some viruses have an envelope that covers the entire virus particle (Lodish, 2000).

The group of cell types that a virus can infect is known as its host range. A certain virus's host range is generally restricted, and it serves as a basis for classification of viruses. According to host, virus is of three types,

- Bacteriophages, that infect bacteria
- Plant viruses that infect plants
- Animal viruses that use animal as its host

The host-cell range of some animal viruses is further restricted because only a limited number of cell types have the appropriate surface receptors for allowing the virus to enter (Lodish, 2000). Still there are so many viruses that infects almost every animal on this planet.

Zika virus is an animal virus that belongs to the genus Flavivirus under the family Flaviviridae. It is an arthropod- borne virus whose main carrier is the mosquito *A. ageptie.* ZIKV is closely related to other members of the family such as WNV, DENV, YFV, CHIKV, HIV, Ebola virus etc. Zika virus has two lineages, The African strain and The Asian strain. Zika virus is known to be asymptomatic in 80% of the cases (Yadav, 2016, Sharma, 2017).

Zika virus (ZIKV) dragged the attention of the world back in 2015 when an outbreak occurred in Brazil causing massive microcephaly in new-borns. This epidemic quickly spread through South and North America as well as several islands in the pacific and Southeast Asia. A situation of panic was raised when a 20-fold increase in parental microcephaly and 19% increase in Guillain Barre syndrome (GBS) was reported in 2015 compared to the previous year (Sharma, 2017). In February 2016, the World Health Organization (WHO) declared Zika virus a public health emergency owing to its association with such congenital deformation (Rawal, 2016). Though the 'Emergency' status was lifted in November 2016, WHO is committed for a long-term response against Zika Virus.

This severe outbreak was caused by the Asian lineage. It is postulated that this strain emerged due to accumulation of mutations in ZIKV genome therefore, changing its molecular interaction pattern that eventually changed its pathogenicity, vector competence and epidemic potential.

Vaccines against multiple flavivirus such as YFV, TBEV, JEV, DNEV has been successfully developed so far. To date, no vaccine against ZIKA virus has entered clinical stage (Sharma, 2017). The only FDA approved drug proven to inhibit ZIKA virus infection in non-human model is Sofosbuvir (Bullard-Feibelman, 2017). However, this drug is not enough to face the challenge raised by ZIKV today. A comparative study of Zika virus with its relative species may provide valuable insight while developing an effective ZIKA virus vaccine.

1.2 Zika Virus: Genetic Basis of Classification and Replication Cycle

Depending on viral mRNA synthesis, viruses are classified in seven major classes. This classification is called Baltimore Classification. Since the stage 'Transcription' or mRNA synthesis brings in the major differences among virus life cycles it is very important to pin down how the virus has accomplished it. The classes in Baltimore Classification (Catherine, 2018) are:

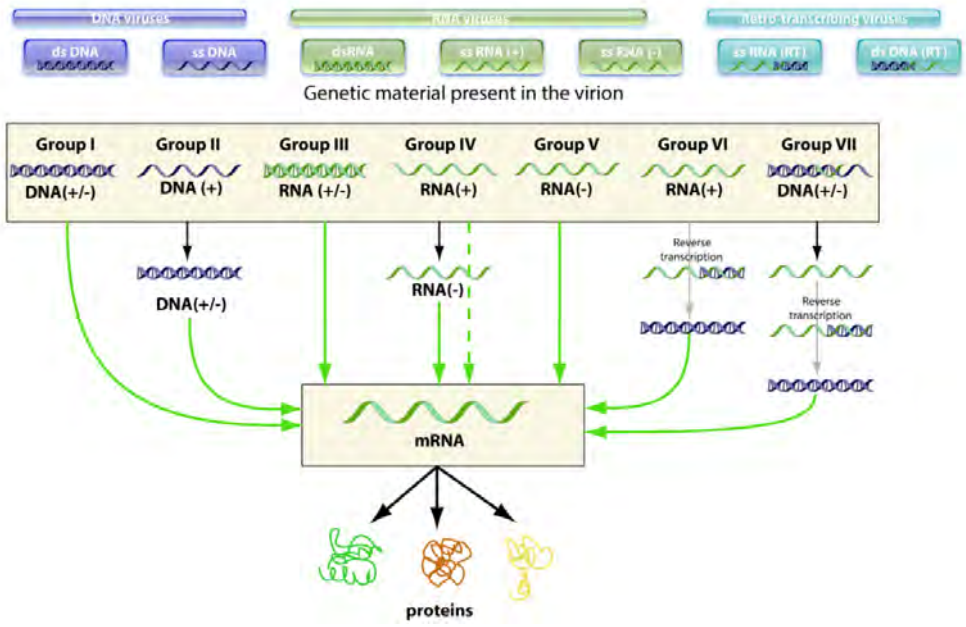


Image 1.1: Strategy of virus replication according to Baltimore Classification (ViralZone)

Despite of the great diversity that viruses have, all virus life cycle can be described using some ambiguous terms which are successively (Hulo, 2017),

- Entry
- Latency
- Transcription
- Replication
- Exit

Zika virus being positive strand RNA virus and follow the rules of class four. Therefore, means that the plus-sense RNA genome is the same sense as mRNA. So, it can be translated using the cellular machinery when it is uncoated in the cell to produce the single polyprotein the polyprotein is then processed into each individual protein. Such one protein is viral polymerase protein, RNA-dependent RNA polymerase, copies the plus-sense genomic RNA into complementary minus-sense RNA.

It is still unclear how the viral translation switches to transcription using the same genomic RNA as the template. New minus-sense strands serve as the template for new plus-sense strand synthesis.

The nascent plus-sense RNA can be the new mRNA for translation, a template for replication, or packaged.

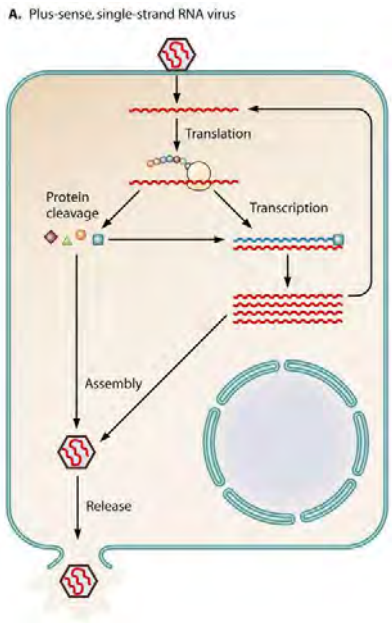


Image 1.2: Positive strand RNA virus replication (American Society of Microbiology)

Viruses can take different mechanisms even if they are in the same stage of their life cycle. For example, there are 8 ways to cross the host membrane, 11 ways to replicate their nucleic acids and more than 4 routes to exit the cell. If Zika virus life cycle is described using precise terms it can be summarized as (Hulo, 2017):

- Attachment
- Apoptotic Mimicry
- Viral endocytosis/ micropinocytosis
- 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

Usually ss (+) RNA virus replication happen within membrane vesicles of the host cell cytoplasm (Hulo, 2017). However, Zika antigens has been found in host cell nuclei (Sharma, 2017) suggesting that Zika virus replication may be different from other flavivirus.

1.3 Characteristics of Zika Virus:

Zika virus is a positive strand RNA virus according to Baltimore Classification. This virus is belonging to the family Flaviviridae under the genome flavivirus. It has been placed under biosafety level 2 (Sharma, 2017)

Significant structural attributes of Zika Virus include,

- It is spherical in shape and approximately 60 nm in diameter (Sharma, 2017)
- It has about 10,800 nucleotide positive strand RNA molecule (Zhou, 2017)
- The RNA is transcribed into a single polyprotein which is later processed into three structural (C protein, M protein, E protein) and seven non- structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5) (Zhou, 2017, Wang, 2017)
- The polyprotein of Zika virus is about 3400 amino acid long (Sharma, 2017)

Here is an image that shows the structural arrangement of ZIKA virus structural proteins and genetic material. The arrangement of the genetic codes is also shown sequentially. Transcription of this mRNA results in the ten structural and non-structural proteins of Zika virus

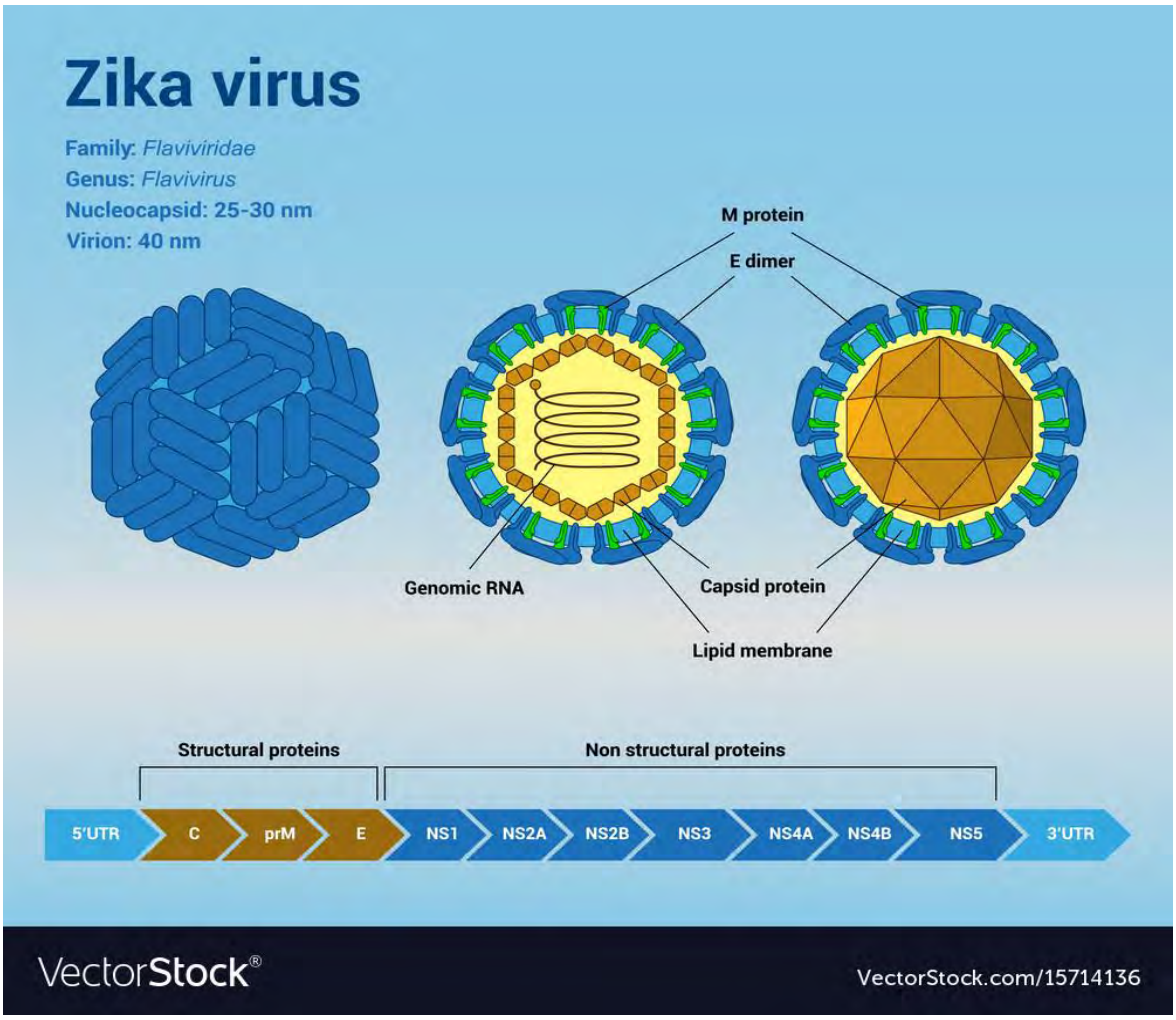


Image 1.3: Structural and Functional components of ZIKV and their locations

The structural and non-structural proteins are key factors about controlling the life cycle and transmission of Zika virus. Their description and functions are described below.

Table 1.1: Functions and Locations of Zika Virus Proteins (Lazear, Diamond 2016)

Protein	Location	Function
Structural Proteins		
Capsid (C)	Structural Core Protein	Nucleocapsid formation by binding to viral RNA
Precursor Membrane (prM/M)	Structural Surface Protein	E protein stabilization, Host cell fusion
Envelope (E)	Structural Surface Protein	Host receptor binding, Host cell fusion, Viral entry
Non-Structural Proteins Mediate viral transcription and replication and mitigate host antiviral responses		
NS1		Virus replication
NS2a		Virus transcription, Virus assembly
NS2b		NS3 cofactor for appropriate serine protease function
NS3	Not a part of virus particle. 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

These proteins work in different way, sometimes in combination to interact with the host and create the complications that Zika Virus cause.

1.4 Vectors and Host Range of Zika Virus

Zika virus is an arbovirus, thus mosquito is its main carrier. Zika virus has been isolated from 17 mosquito species so far. Among them *A. aegypti, A. albopictus, A. hensilli* (responsible for Yap island outbreak), *A. polynesiensis* (responsible for French Polynesia outbreak) are significant (Rawal, 2016). For the recent 2016 outbreak *A. aegypti* was responsible. This species was originated from an ancestral zoophilic tree hole breeding mosquito in the arid environment of North Africa (Shragai, 2017). There are many reasons why this vector became such competent, such as,

- Dense human habitat has created ideal larval habitat for mosquito while providing them good quality blood.
- Migration and Trade between 15th to 19th century particularly slave trade imported mosquitoes from Africa
- They feed almost exclusively on human blood and can exhibit high rates of multiple blood feedings
- They breed in manmade conditions, rest indoors which confirm human exposure
- The eggs of this mosquito can resist desiccation up to 8 months which facilitates its persistence in dry periods.

Because of all these factors, this mosquito has become an ideal carrier for dangerous pathogens like Dengue virus, West Nile virus, Yellow Fever virus, Chikungunya virus, Zika virus etc.

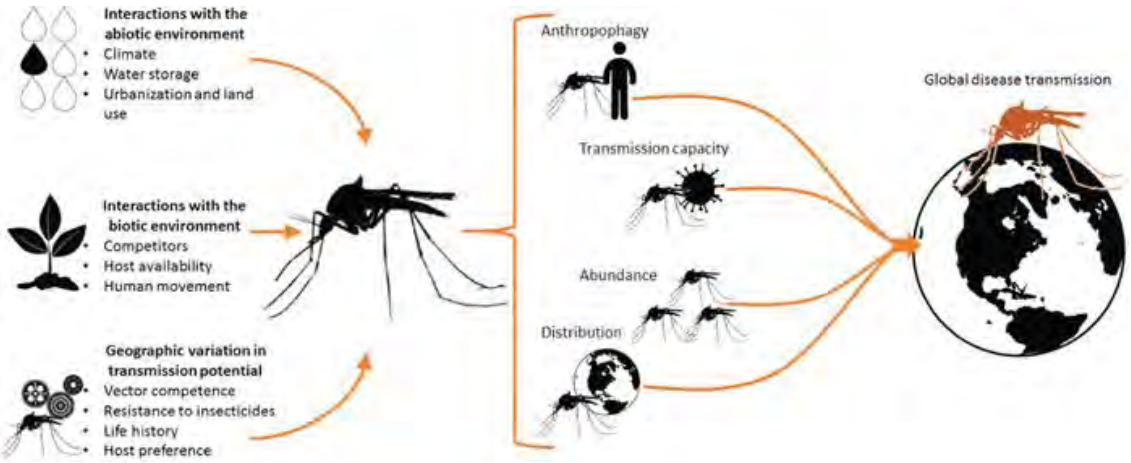


Image 1.4: Mosquito's contribution in disease spreading

To date Zika virus has been found in many organisms. Its main host are human and non-human primates. Different kinds of mosquito also harbour Zika virus. Antibody against Zika has also been found in vertebrates like: rodents, birds, sheep, goat, cattle, reptiles (Rawal, 2016). Zika virus is still believed not to have any zoonotic properties but presence of such antibody in different animals suggest a theory that these animals might have a role in transmission of Zika virus.

1.5 Transmission and Persistence

Zika virus sustains transmission by a human- endemic transmission cycle thus allowing human to serve as its carrier, multiplier and source of more Zika for uninfected mosquitoes.

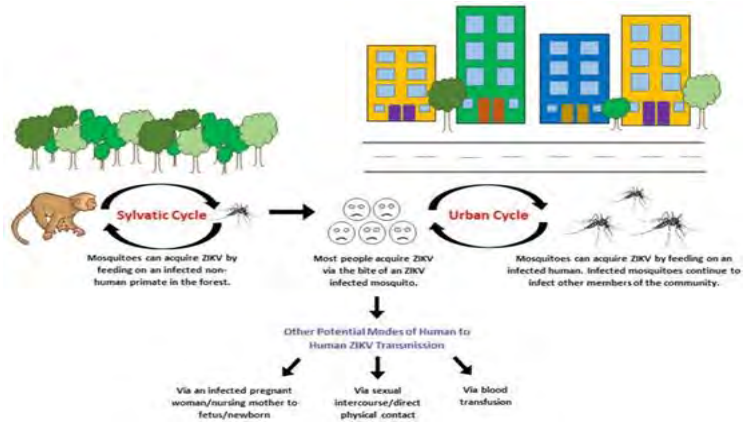


Image 1.5: Zika virus transmission cycle

Primarily Zika virus infection in human occur by biting of vector mosquitos. Transmission can also occur by exchange of body fluids. Zika RNA has been reported in body fluids such as blood, urine, saliva, cerebrospinal fluid, amniotic fluid and breast milk (Shragai, 2017). Some cases have been reported that women who had sexual intercourse with men suffering from Zika illness got the disease (Shragai, 2017). Sometimes victims got infected when blood transfusion (Cavalcanti, 2017). Vertical transmission of Zika virus also occur very frequently. But transmission through breast milk is still a matter of debate. Results of a clinical study done in Brazil suggest that persistence of Zika particles in breast milk may not be enough for efficient transmission of Zika illness (Cavalcanti, 2017).

1.6 Zika Virus Throughout History

Zika virus was discovered first in Ugandan monkeys of Zika forest in 1947. Since then very little research was conducted on this virus as it was known to be quite harmless. The first remarkable outbreak occurred almost sixty years after its discovery.

- In 2007 the first outbreak occurred in Yap Island, Micronesia and almost 70% of its population was infected (Rawal, 2016)
- In 2013, the French Polynesia outbreak occurred (Rawal, 2016)
- In 2014, Zika infected New Caledonia (Rawal, 2016)
- In May 2015, Zika virus in Brazil was confirmed and it was related to a dramatic rise in microcephaly and GBS (Rawal, 2016, Sharma 2017)
- By January 2016 Zika epidemic has spread to many countries of Americas including Bolivia, Brazil, Cape Verde, Colombia, Dominican Republic, Ecuador, El Salvador, French Guinea, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Martinique, Mexico, Panama, Paraguay, Saint Martin, Samoa, Suriname and Venezuela (Rawal, 2016)

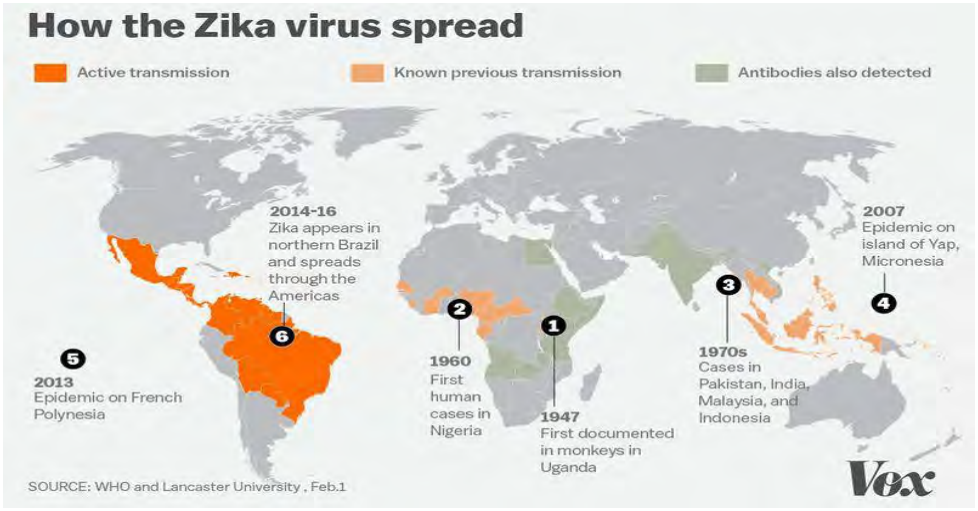


Image 1.6: Transmission of ZIKV around the world and major events according to WHO and Lancaster University

According to WHO and CDC, there are many areas that are at risk of Zika invasion. They have classified countries in different regions under different risk groups. According to a joint reviewed report of WHO, CDC and European CDC as current as March 2018, these countries are,

Asia: Bangladesh, Burma (Myanmar), Cambodia, India, Indonesia, Laos, Malaysia, Maldives, Pakistan, Philippines, Singapore, Thailand, Timor-Leste (East Timor), 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

Countries with a history of flavivirus infection have a higher risk of Zika epidemic

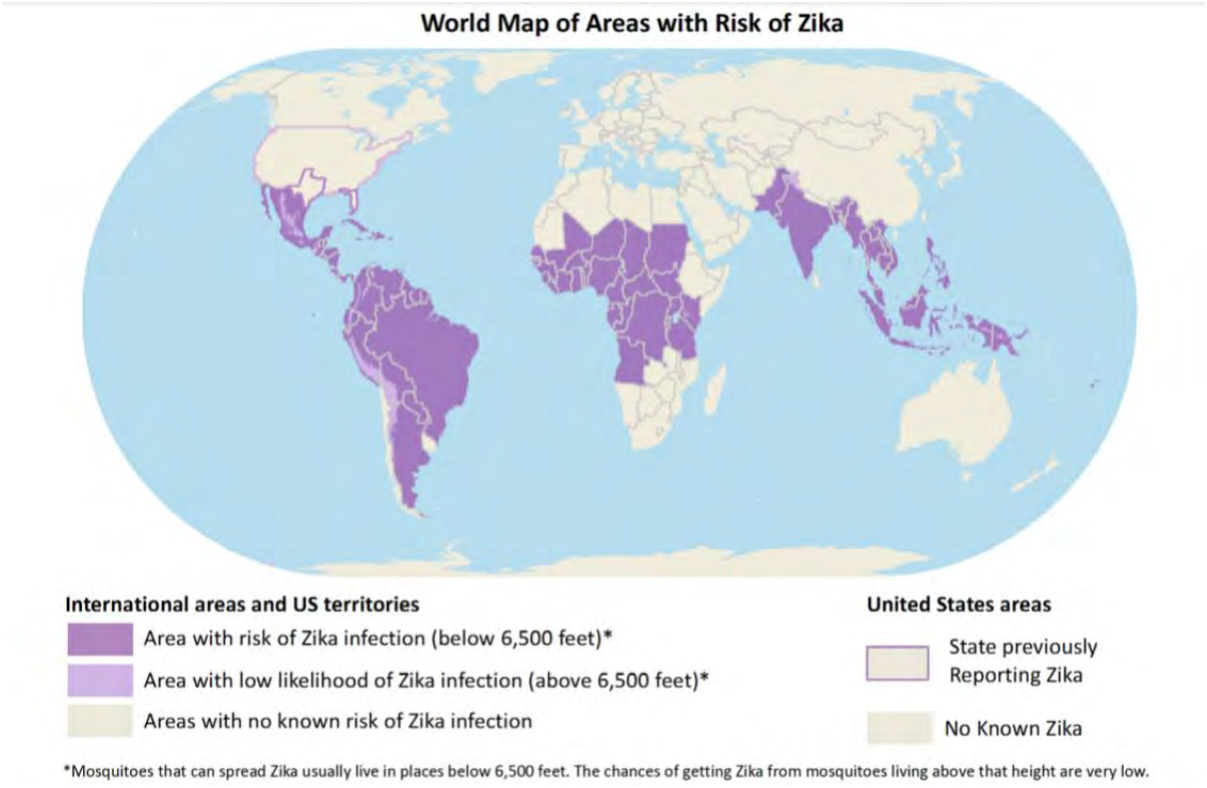


Image 1.7: Countries with Risk of Zika Virus infection (WHO, CDC)

It is postulated that Zika virus originated in East Africa and had spread through the west. Theories also suggest that the Asian lineage emerged because of mutation accumulation in Zika genome. Genomic studies show that Zika virus circulating in America was more than 99% like that of French Polynesia and 89% identical to African strain (Rawal, 2016).

Using distance matrix analysis on viral E protein, NS1, NS2a, NS2b and NS3, the Brazilian isolate (BR_ZIKA_AB_ES) was found at least 12.2% divergent from several African strains. Again, the divergence rate of Brazilian isolate with some Asian strains was found maximum 1.9% (Garcez, 2017). This clearly indicates that the Asian lineage has become pathogenic and has caused serious health problems across the globe.

1.7 Complications by Zika Virus:

Zika virus infection in most cases is known to be asymptomatic or mildly symptomatic. In human Zika virus infection is characterized by,

- Mild fever (37 – 38◦C)
- Arthralgia (Mainly small joints of hand and feet)
- Myalgia
- Headache
- Retro- orbital pain
- Conjunctivitis
- Cutaneous maculopapular rash
- Asthenia
- Peripheral edema
- Gastrointestinal disturbance (Yadav, 2016, Rawal, 2016)

Such symptom occurs within 2- 12 days of mosquito bite and resolve within 2- 7 days of onset (Rawal, 2016). However, when Zika virus gets into the system of pregnant women it becomes very notorious. During pregnancy it can cause

- Nervous System Injury
- Placental insufficiency
- In vitro fetal growth insufficiency with or without microcephaly
- Microcephaly (Rawal, 2016)

The main ultrasound finding of Zika infected babies are microcephaly. Along with this, intercranial calcification (cerebellum, intraocular, brain), brain atrophy, ventricular dealation, hydranencephaly, growth retardation also occurs (Rawal, 2016). Altogether these situations are referred as Congenital Zika Syndrome (Zhou, 2017).

In adults Zika virus causes a rare neurological disorder known as GBS or Guillain Barry Syndrome. In 2015 Brazil had a 19% increase in GBS that was linked to Zika outbreak. 90% Of patients who were diagnosed with GBS had symptoms of Zika fever prior to diagnosis (Rawal, 2016)

Apart from that Zika causes some other complication as well. Once Zika virus gets into the nervous system, it invades optic nerve, retina, iris and corona (Zhou, 2017). As a result, Zika causes ophthalmologic infection in adults including optic neuritis, chorio-retinal atrophy, blindness and uveitis.

Moreover, Zika has been linked to ankle edema, axillary or inguinal lymphadenopathy, leukopenia with monocytes and thrombocytopenia from travellers in Italy (Rawal, 2016).

1.8 Microcephaly:

In humans, microcephaly represents a severe congenital defect of the brain of the new-born. Microcephaly is characterized by head circumference more than two standard deviation below the mean for gestational age (Rawal, 2016). Microcephaly can occur by two types of mechanisms (Faizan, 2016).

- The first occurs when the brain fails to grow to its appropriate size during pregnancy at around 32 weeks of the gestation period and is caused by a gradual decrease in the neuron production.
- The second relates to a normal brain size at birth but failure to grow subsequently due to the loss of dendritic connections.

Some researchers have classified 3 types of microcephaly based on Giacomino's classification (Faizan, 2016):

- Microcephalia vera, where the size of the brain remains small without any sign of injury or deformation,
- Microcephalia spuria, which shows some pathological changes and injury to the brain, and

- Microcephalia combinata, which reflects a small brain size with a trace of injury.

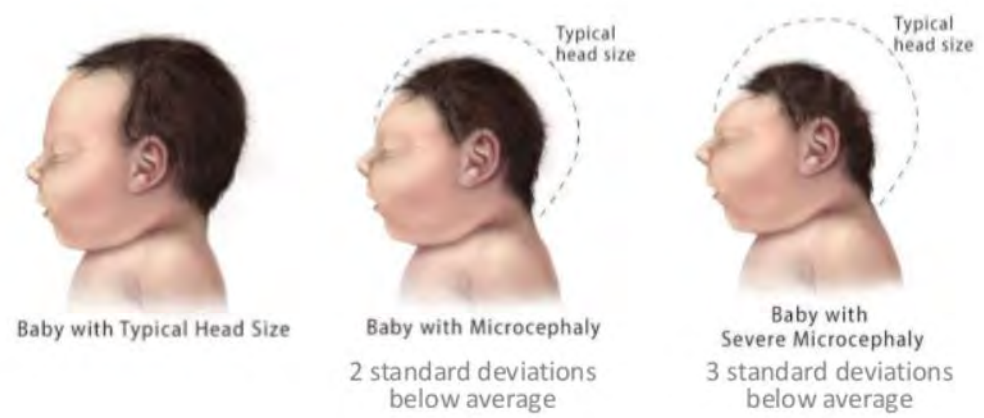


Image 1.8: Types of Microcephaly (CDC, US Department of Health)

The cause of microcephaly can be both genetic and non-genetic. The genetic aspects include ambiguous genitalia and the 14-3-3 epsilon gene that causes lissencephaly. The non-genetic factors include alcohol consumption during pregnancy, abnormal weight gain during the gestation period, poor parental care including malnutrition, incomplete placental development, systemic and metabolic disorders, exposure to teratogens during pregnancy, non-accidental head injury, Rubenstein-Taybi syndrome, and viral infections (Faizan, 2016)

Many experiments have been done to understand the mechanism of Microcephaly caused by Zika virus. And many carry the proof that Zika virus has a significant role here. In a study it was found that Zika Virus disrupts molecular fingerprints in human neurospheres (Garcez, 2017). Here Neural stem cells derived from induced pluripotent stem cells were infected with Zika for 2 hours (MOI 0.025). The result showed that the infected neurospheres were much smaller than control suggesting that growth was impaired.

- After three days of infection, viable neurospheres with ZIKV mediated growth impairments were further characterized. Immunostaining for markers of apoptosis, neural progenitors, and neuronal cell were employed. ZIKV-infected neurospheres showed increased levels of activated caspase 3 and displayed higher amounts of pyknotic nuclei compared to mock. Additionally, the number of progenitor cells was reduced
- After six days in vitro, the number of ZIKV-infected neurospheres was reduced by 50%. Finally, after twelve days of ZIKV infection in vitro, neurospheres were entirely depleted.

- Along with cell death, alterations in cell cycle contributed to a reduction in the number of neural progenitors and newborn neurons. Flow cytometry analyses revealed that cells in ZIKV-infected neurospheres accumulate abnormally in a sub-G1 phase (Garcez, 2017)

There are many theories about how actually Zika virus cause microcephaly in babies. Here one of them is discussed:

- Dermal fibroblasts and epidermal keratinocytes are the primary targets of Zika virus infection.
- This is followed by the infection of dermal dendritic cells (Langerhans cells), which facilitates Zika virus dissemination to different organs through the circulatory system
- The transplacental transmission of Zika virus from mother to foetus may occur by the infection of placental macrophages (Hofbauer cells) and cytotrophoblasts. And macrophages are the main target cells of Zika viral infection in the placenta.
- Viral infection in placental macrophages induces the production of type I interferon and pro-inflammatory cytokines resulting in an antiviral gene expression.
- The entry of Zika virus is mediated by cell surface receptors DC-SIGN, AXL, heat shock proteins, TYRO3, and TIM-1. AXL, a phosphatidylserine protein, belongs to the TAM receptor family of phagocytic receptors. A recent investigation has shown that the AXL protein is overexpressed in developing human brain cells, including radial glia, astrocytes, endothelial, and microglia. Interestingly, these cells with highly expressed AXL protein are particularly vulnerable to Zika virus infection.
- Entry of Zika virus through this receptor stimulates AXL-mediated signalling pathways and suppresses the innate immune response which leads to viral infection in neurons and their associated cells.
- Again, In normal cells, the TLR3 receptor serves as a defender against viral invasion or innate immune response. This study suggested that in Zika virus-infected NPCs, the TLR3-regulated immune network affects the expression of around 41 genes responsible for neurogenesis, the differentiation of NPCs, and apoptosis.
- Zika virus infection leads to an alteration in the regulation of genes associated with the immune response, cell cycle, differentiation, and apoptosis in NPCs, resulting in neurological malformations.

Although the precise molecular mechanism of microcephaly is still unknown, this model has recently been proposed for Zika virus-induced damage to the developing brain. (Faizan, 2016)

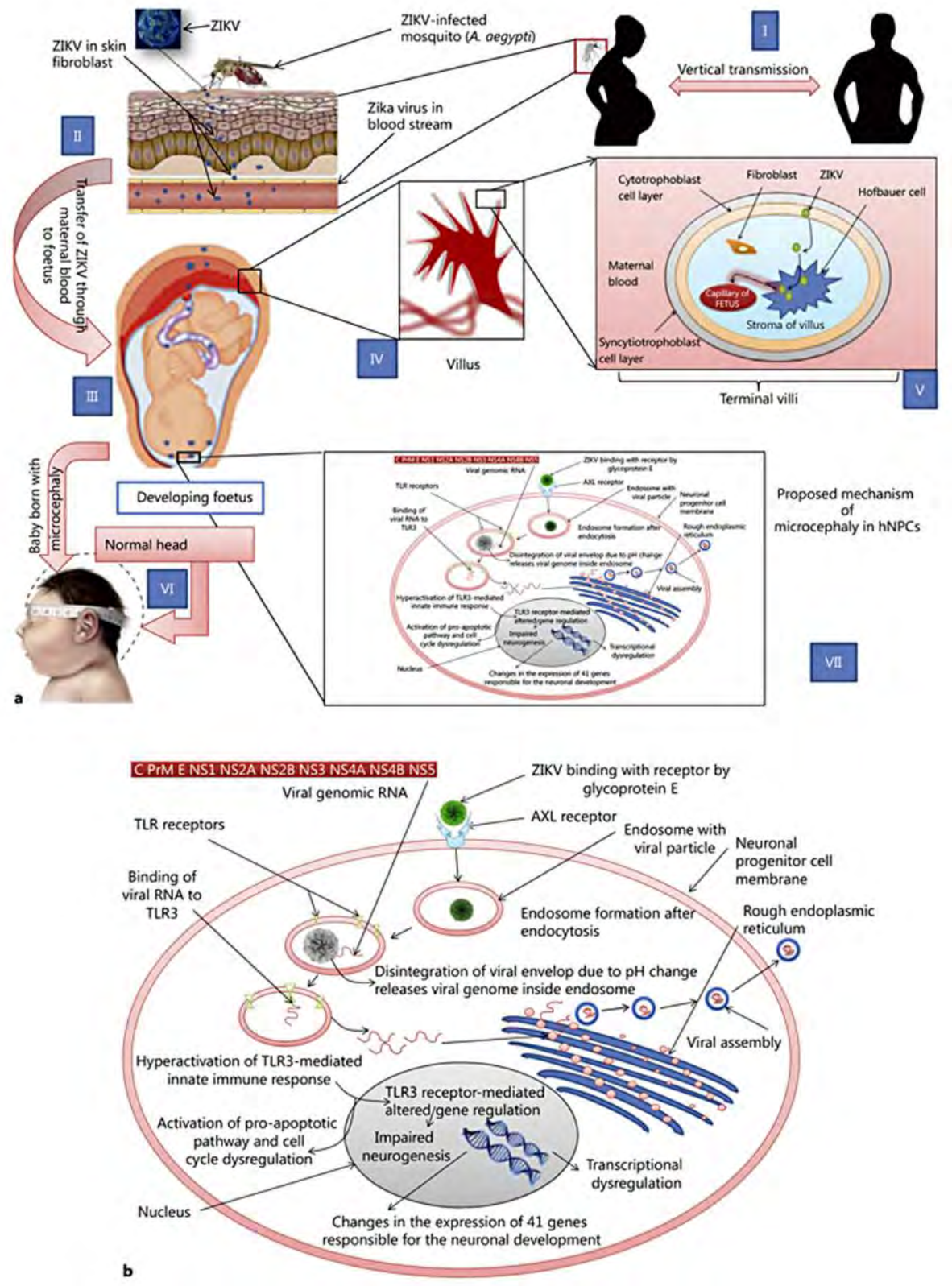


Image 1.9: Proposed Mechanism of Zika Virus Induced Microcephaly

1.9 Bioinformatics:

Bioinformatics is the applications of computer sciences to molecular biology, mostly the study of macromolecules such as proteins and nucleic acids. Bioinformatics is an interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analysing biological, medical, agricultural, environmental etc data.

The scopes of bioinformatics are far and wide. Using bioinformatics techniques tasks like genome sequencing, gene annotation, gene function prediction, gene library establishment, gene to protein identification, function and structure prediction, identifying and understanding the molecular machineries, simulation and inference of metabolic, genetic and protein networks, drug discovery, disease cure etc are being done.

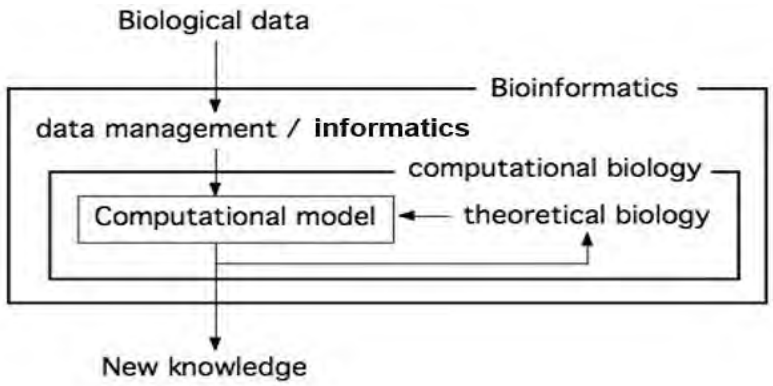


Image 1.10: Role of Bioinformatics

1.10 Multiple Sequence Alignment:

Multiple Sequence Alignment is a type of alignment of 'n' number sequences where of n > 2. This alignment is done by inserting gaps ("-") into sequences so that all the resulting sequences have equal length. This is because later they need to be arranged in a scoring matrix for calculating the alignment.

Multiple sequence alignment (MSA) can be considered as a generalization of Pairwise Sequence Alignment just instead of aligning two sequences, here multiple sequences are used.

MSAs are at the core of comparative genomics studies which is involved with evolutionary histories, functional and structural aspects of sequences, and understanding phenotypic differences between species.

1.11 Protein Annotation:

Protein annotation is a systematic way of describing regions or sites of interest in the any amino acid sequence. Annotation basically aims to assign a certain sequence's characteristics from different perspectives such as, whether its involved in a biological process or what its molecular function can be or if it is a cellular component. There are many databases that are solely dedicated to genome and protein annotation. One of the most popular is GO (Gene Ontology) database. If a query is searched in this database, its structural and functional involvement are derived as results.

1.12 Phylogenetic Tree:

Phylogeny is the study of evolutionary relations among various groups of organisms that can be species or populations of a species. A phylogenetic tree is an expression of phylogeny that indicates evolutionary relatedness and distance among organisms of interest in an arranged topology. Basic components of a phylogenetic tree are, root, branch, nodes, leaves, clades. Branch length is an important indication in a phylogenetic tree that represents the number of changes that have occurred in the branch.

Phylogenetic tree can be constructed on distance based or character-based methods. But the distance-based methods are more popular and accurate.

1.13 Vaccine:

Vaccines is a mixture of weakened or killed form of the microbe, or part of a microbe that doesn't cause disease or reproduce but stimulates the macrophages, which present the antigens to T and B cells

The mock infection is rapidly cleared, and the body is left with memory T cells and B cells ensuring long term immunity. Vaccines are of many types:

- Live, attenuated vaccines
- Inactivated vaccines
- Subunit vaccines
- Toxoid vaccines
- Conjugate vaccines
- DNA vaccines
- Recombinant vector vaccines etc.

1.14 Antigen Prediction:

Antigen is substance that is foreign to the body and when introduced, it stimulates the production of an antibody or any specific immune response. Antigens include toxins, bacteria, foreign blood cells, virus particles etc.

Chemically, antigens have large molecular weight (>10KDa). Antigens can be proteins or polysaccharides. Proteins can include conjugated proteins such as glycoproteins, lipoproteins, and nucleoproteins and polysaccharides usually including lipopolysaccharides. These protein and polysaccharide antigens are found on the surfaces of viruses and cells.

The portion of antigen molecules which can be specifically bound by antibody or antigenic receptor of lymphocytes is called an epitope. The size of an epitope is generally equivalent to 5-15 amino acids or 3-4 sugar residues or 6-8 nucleotide. Every epitope recognizes a different antibody.

Many software can correctly identify epitopes on a protein sequence. Using these software, a probable epitope can be identified prior to vaccine design. This reduces both time and cost behind vaccine development.

1.15 Aims and Objectives:

- To study about Zika virus and its current consequences.
- To identify the potential of Zika virus
- To compare Zika virus with other related viruses.
- To find out the conserved regions of all related viruses and annotate their functions
- To explore the possibilities of drug development against Zika virus
- To identify a competent peptide for developing a vaccine against Zika virus.

Materials and Methods

2.1 Method Summary:

For doing this study, several web-based tools were used. All these tools are authenticated and very popular in such research. The first phase of the work is Multiple Sequence Alignment (MSA) of polyprotein sequences of Zika virus and relative species. For doing this alignment EMBL tool T Coffee (Notredame, 2008) software was used. It gives a colourful visualization along with the alignment. For cross checking the result of T Coffee, the same task was done with NCBI Blast tool COBALT (Papadopoulos, 2007). A phylogenetic tree was also constructed using COBALT primarily. A second phylogenetic tree, using more sophisticated parameters was constructed using MEGA7 (Kumar, 2015). This tree constructed by MEGA7 had 100% data coverage and showed the branch length of all related species to Zika virus.

From the results of MSA, the regions with maximum similarity were marked. The regions of maximum similarity are likely to be conserved. Here such 53 regions (amino acid sequences) were selected. These sequences were then used as inputs in annotation software InterProScan (Jones, 2014) to determine their function. For cross checking the results of InterProScan, the Zika polyprotein sequence was again checked with another such tool called ScanProStite (De Castro, 2006) that also does functional annotation. Along with that ScanProsite also displays the locations and indications of a sequence.

Finally, antigenicity test of the chosen Membrane Glycoprotein (M) was done with Vaxijen 2.0 (Irini, 2007) that calculates the antigenicity of any target organism sequence. Later, Epitope prediction was done respectively by BepiPred (Jespersen, 2017), BCPreds (Chen, 2007), Vaxijen. After that few epitope-compatibility checks was done with IEDB analysis tools (Larsen, 2006).

The entire work is summarized in the flowchart below:

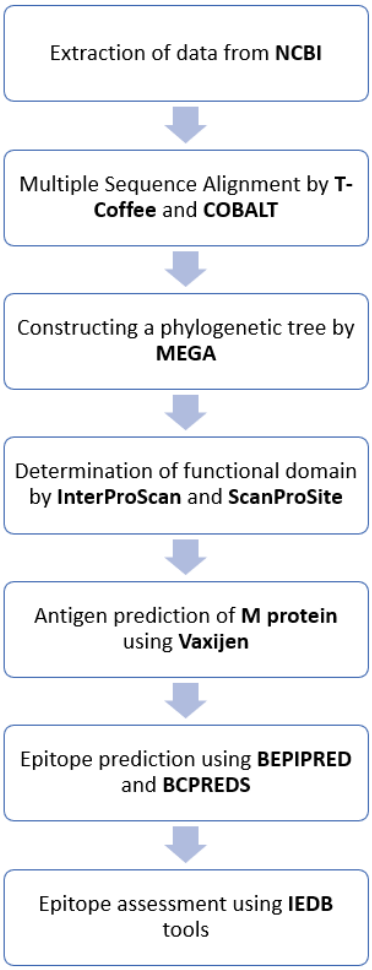


Image 2.1: Summary of the methodology

2.2 Data Collection:

Protein Sequences of Zika virus and its relative species like West Nile virus, Rubella virus, Dengue virus, Ebola virus, HIV virus, Hepatitis C virus, Cytomegalovirus, Chikungunya virus was collected and analysed by different online software.

The first stage of the work is to do multiple sequence alignment. For doing so, protein sequence of Zika virus and other related flavivirus was collected from NCBI Database. Further this data was used for analysis

Table 2.1- Viruses studied with their GenBank ID

	Virus	GenBank ID
1.	Zika Virus Polyprotein	ASK51714.1
2.	West Nile Virus Polyprotein	AAV54504.1
3.	Rubella Virus Polyprotein	AHI17076.1
4.	Dengue Virus 2 Polyprotein	AOE23002.1
5.	Human Immunodeficiency Virus 1 (Partial Polyprotein)	AAB37115.1
6.	Human Immunodeficiency Virus 1 (Partial Polyprotein)	BAK40453.1
7.	Ebola Virus (Structural Polyprotein)	S32585
8.	Hepatitis C virus polyprotein	BAJ07247.1
9.	Chikungunya Virus (Non-Structural Polyprotein)	AMQ76463.1
10.	Cytomegalovirus Polyprotein	OMO70300.1

2.3 Multiple Sequence Alignment Method:

Multiple sequence alignment of protein sequences was performed differently by two software. They are T- Coffee of EMBL and COBALT of NCBI

2.3.1 T- Coffee- EMBL

T-Coffee is a multiple sequence alignment (MSA) software developed by EMBL. Its function is to show the combined results obtained with several alignment methods. Transitive Consistency Score (TCS) is the alignment evaluation score used in T- Coffee. It can identify the most correct positions in an MSA. These positions are the likely to be structurally correct and most informative when estimating phylogenetic trees. The MSA result is shown by colour where different colours indicate different consistency of Alignment.

2.3.2 COBALT- NCBI

COBALT does progressive multiple alignment of protein sequences. The alignment is aided by a collection of pairwise constraints derived from conserved domain database, protein motif

database, and local sequence similarity using RPS-BLAST, BLASTP, and PHI-BLAST, respectively. Computation time is reduced by forming clusters of sequences that share many common words and finding conserved domains and motif matches for only one sequence per cluster.

2.4 Phylogenetic Tree Construction Method:

2.4.1 MEGA

MEGA software has been to provide tools for exploring, discovering, and analysing DNA and protein sequences from an evolutionary perspective. MEGA has functions like constructing phylogenetic trees, computing evolutionary distance, constructing time trees, identifying gene duplications etc. Here a phylogenetic tree using MEGA was constructed. Because If reliable phylogeny is produced, it will shed light on the sequence of evolutionary events that generated the present-day situation.

There are numerous methods for constructing phylogenetic trees. Here NJ/UPGMA method and default settings were used.

2.5 Determination of Functional Domain:

2.5.1 Inter Pro Scan

InterProScan is a software package that allows both protein and nucleic acid sequences to be scanned against InterPro's signatures. Signatures are predictive models, provided by several different databases, that make up the InterPro consortium.

2.5.2 Scan Pro Site

ScanProSite allows to scan proteins for matches against the PROSITE collection of motifs as well as against patterns provided.

PROSITE is a database of protein families and domains. It is based on the observation that, while there is a huge number of different proteins, most of them can be grouped, based on similarities in their sequences, into a limited number of families. Proteins or protein domains belonging to a particular family generally share functional attributes and are derived from a common ancestor. PROSITE currently contains patterns and profiles specific for more than a thousand protein families or domains. Each of these signatures comes with documentation providing background information on the structure and function of these proteins.

2.6 Antigen Prediction Method:

Vaxijen:

Vaxijen is a server for predicting protective antigens of bacterial, viral or tumour origin without alignment. The server contains models derived by auto- and cross-covariance pre-processing of amino acids properties. The predictive ability of the models was tested by internal leave-one-out cross-validation on training sets and by external validation on test sets. The models showed remarkable stability, as tested by combinations of the positive set and five different negative sets. Therefore, it is a reliable and consistent tool for the prediction of antigens. It can be used independently or in combination with other bioinformatics tools used for reverse vaccinology.

2.7 Epitope Prediction Method:

2.7.1 BepiPred 2.0:

BepiPred-2.0, is a web server that allows users to predict B-cell epitopes from antigen sequences. The server is based on a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures. This new method was found to outperform other available tools for sequence-based epitope prediction.

2.7.2 BCPreds:

BCPREDS is a web-based server for predicting B-cell epitopes by choosing among several developed prediction methods. The current version of BCPREDS allows the user to select among three prediction methods: (i) our implementation of AAP method; (ii) BCPred (iii) FBCPred. Users will provide an antigen sequence and optionally can specify desired epitope length and specificity threshold. Results are returned in several user-friendly formats.

2.8 Epitope Assessment Method:

IEDB tools:

IEDB is a collection of tools for predicting and analysing of immune epitopes. It is as a companion site of the Immune Epitope Database (IEDB), a manually curated database of experimentally characterized immune epitopes.

The tools of IEDB are designed to predict regions of proteins that are can potentially be recognized as epitopes in the context of a B cell response. Some of the tools include: BepiPred

Linear Epitope Prediction 2.0, Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Antigenicity, Parker Hydrophilicity Prediction etc.

Results

3.1 Result Summary:

The first experiment done for this analysis was Multiple Sequence Alignment with T Coffee and COBALT respectively. From these alignments 53 regions of amino acid sequences were identified that have maximum similarity among species and are likely to be conserved. Alongside, a phylogenetic tree was constructed with MEGA showing the relation and distance of ZIKA virus and relative species.

Later the 53 regions of similarity were analysed by InterProScan and cross checked with ScanProSite. Thus function, location and other indications of Zika virus proteins were retrieved.

The antigen prediction started with assessing the antigenicity of the selected Membrane Glycoprotein (M). According to the software Vaxijen the threshold value for a virus peptide to become antigen is ≥ 0.4. For M protein the value was determined 0.4358 indicating that this M protein might be antigenic. For further determination of the epitope analysis was done with BepiPred and BCPreds. After testing under many conditions (threshold, epitope length etc) the most compatible epitopes were, a seven amino acid long sequence: GSSTSQK and 12 amino acid long ENWIFRNPGFAL**.** When run in Vaxijen overall Prediction for the antigenicity of for GSSTSQK was 1.1341 and for ENWIFRNPGFAL it was 0.8161, which is above the threshold value (≥ 0.4). Thus, making them probable candidate for vaccine. Later these sequences were checked for conservancy, surface accessibility, hydrophobicity, flexibility and antigenicity by IEDB tools.

3.2 IN SILICO COMPARATIVE STUDY OF ZIKA VIRUS

3.2.1 Multiple Sequence Alignment Results:

The protein sequences of Zika and its relative viruses were analysed by T Coffee and COBALT differently. The aim was to compare the Zika polyprotein sequence to its relatives. Both softwares compare protein sequences given as input and exhibit similar results. Here the target sequence was Zika polyprotein and that is to be compared with other viruses. For making the result more reliable both were used differently.

Zika Polyprotein sequence is the most important material for this study:

```
>ASK51714.1 polyprotein [Zika virus]
MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWG
SVGKKEAMEIIKKFKKDLAAMLRIINARKEKKRRGADTSVGIVGLLLTTVMAAEVTRRGSAYYMYLDRND
AGKAISFPTTLGMNKCYIQIMDLGHMCDATMSYECPMLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEA
WRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQKVIYLV
MILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYC
YEASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFACSKKMTGKSI
QPENLEYRIMLSVHGSQHSGMIVNDTGHETDENRAKVEITPNSPRAEATLGGFGSLGLDCEPRTGLDFSD
LYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKEALVEFKDAHAKRQTVVVLGSQEGAVHTALA
GALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLHGTVTVEVQYAGTDGPCK
VPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKA
FEATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLMWLGLNTK
NGSISLMCLALGGVLIFLSTAVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAI
KQAWEDGICGISSVSRMENIMWRSVEGELNAILEENGVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWK
AWGKSYFVRAAKTNNSFVVDGDTLKECPLNHRAWNSFLVEDHGFGVFHTSVWLKVREDYSLECDPAVIGT
AVKGKEAVHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNT
REGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTRGPSLRSTTASGRVIEEWCCRECTMPPLSFRAK
DGCWYGMEIRPRKEPESNLVRSVVTAGSTDHMDHFSLGVLVILLMVQEGLKKRMTTKIIISTSMAVLIAM
ILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLVSFIFRANWTPRESMLLALASCLL
QTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRAGLATCGGFMLLS
LKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEM
AGPMAAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREII
LKVVLMAICGMNPIAIPFAAGAWYVYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGV
GVMQEGVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGHSEVQLLAVPPGERARN
IQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVEC
FEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAV
NVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTAT
PPGTRDAFPDSNSPIMDTEVEVPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSR
KTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHASAAQRRG
RIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQ
RKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTINNTIMEDSVPAEVWTRHGEKRVLKPRWMDAR
VCSDHAALKSFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPET
LETIMLLGLLGTVSLGIFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIP
EPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKSDLSHLMGRREEGATIGFSMDIDLRPASAWA
IYAALTTFITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFYTWDFGVPLLMIGCYSQLTPLTLIV
AIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAVS
SAILSRTAWGWGEAGALITTATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRR
GGGTGETLGEKWKARLNQMSALEFYSYKKSGITEVCREEARRALKDGVATGGHAVSRGSAKLRWLVERGY
LQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPVLVQSYGWNIVRLKSGVDVFHMAAEPC
DTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGLVRVP
LSRNSTHEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGN
RIERIRSEHAETWFFDENHPYRTWAYHGSYEAPTQGSASSLVNGVVRLLSKPWDVVTGVTGIAMTDTTPY
GQQRVFKEKVDTRVPDPQEGTRQVMSMVSSWLWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEW
KTAVEAVNDPRFWALVDKEREHHLRGECQSCVYNMMGKREKKQGEFGKAKGSRAIWYMWLGARFLEFEAL
GFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEMSRIPGGRMYADDTAGWDTRISRFDLENEALITNQMEK
GHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEV
LEMQDLWLLRRSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWKPST
GWDNWEEVPFCSHHFNKLHLKDGRSIVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFH
RRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDMLVVWNRVWIEENDHMEDKTPVTKWTDIPY
LGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGEEEKYMDYLSTQVRYLGEEGSTPGVL
```

Image: 3.1: Zika Virus Polyprotein Sequence

The alignment indicated the most similar patterns among the different species. These sequences are likely to be more conserved and code for functional proteins.

3.2.1.1 T Coffee Alignment Results:

After the protein sequence data was put in T Coffee software, it showed the alignment according to default settings. T Coffee indicates the most similar regions with colour.

Image 3.2: T Coffee Result (Run 1): Here red indicates good, yellow indicates average and green indicates bad alignments. The line below named 'cons' shows the average consistency of the alignment.

After performing this alignment, the sequences were again aligned where the arrangement of inputs were changed. The species which align well according to this result were grouped together and those who align bad were put in another group. Therefore, the final arrangement:

Table 3.1: Input arrangement after Run 1 in T Coffee Software

Query	Virus	GenBank ID
1.	Zika Virus Polyprotein	ASK51714.1
2.	West Nile Virus Polyprotein	AAV54504.1
3.	Dengue Virus 2 Polyprotein	AOE23002.1
4.	Ebola Virus (Structural Polyprotein)	S32585
5.	Hepatitis C virus polyprotein	BAJ07247.1
6.	Chikungunya Virus (Non-Structural Polyprotein)	AMQ76463.1
7.	Rubella Virus Polyprotein	AHI17076.1
8.	Human Immunodeficiency Virus 1 (Partial Polyprotein)	AAB37115.1
9.	Human Immunodeficiency Virus 1 (Partial Polyprotein)	BAK40453.1
10.	Cytomegalovirus Polyprotein	OMO70300.1

This arrangement was expected to identify the similar regions more easily and they can be noted down for further analysis. This way 53 regions were manually selected. Similar sequences exceeding 10 amino acids were selected because at lease 10 amino acids are needed for coding a protein. These sequences were then scanned by ScanProSite.

The images of all T Coffee results after rearrangement of virus is given below:

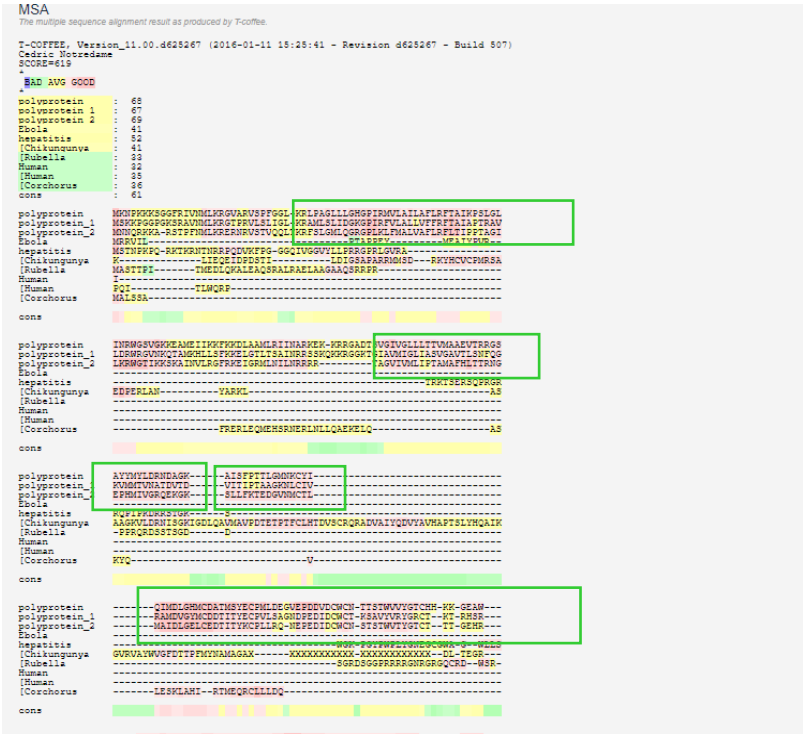


Image: 3.3: T Coffee Results Run 2 (Rearranged)

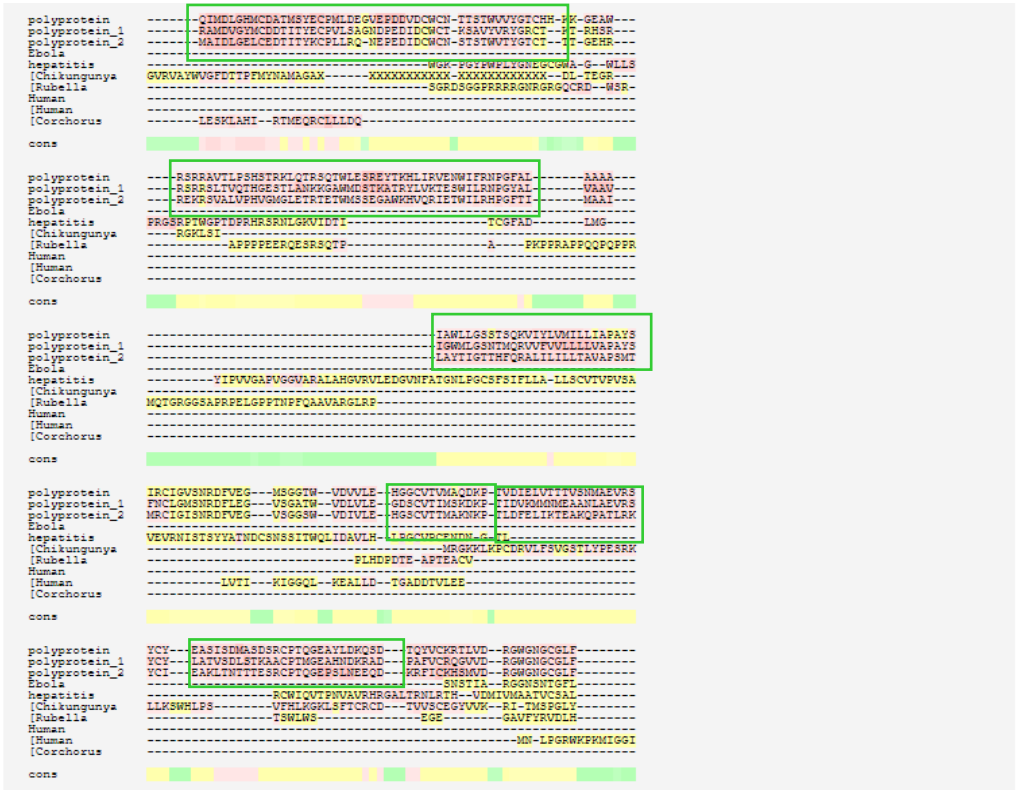


Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

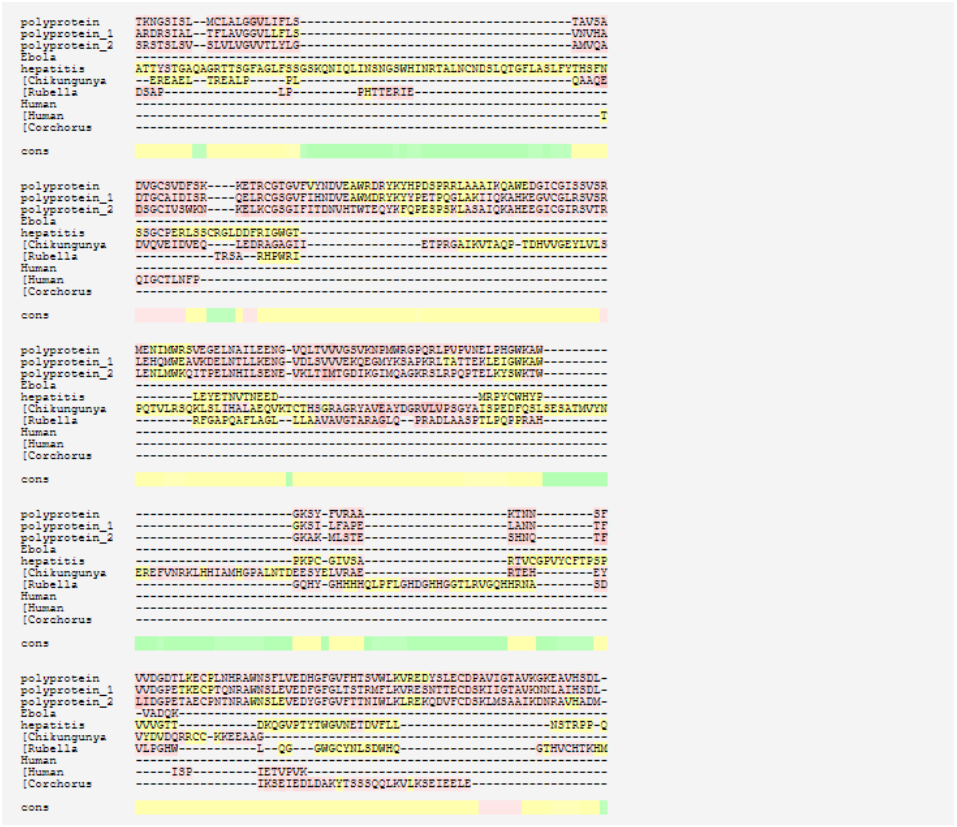


Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3(Continued): T Coffee Results Run 2 (Rearranged)

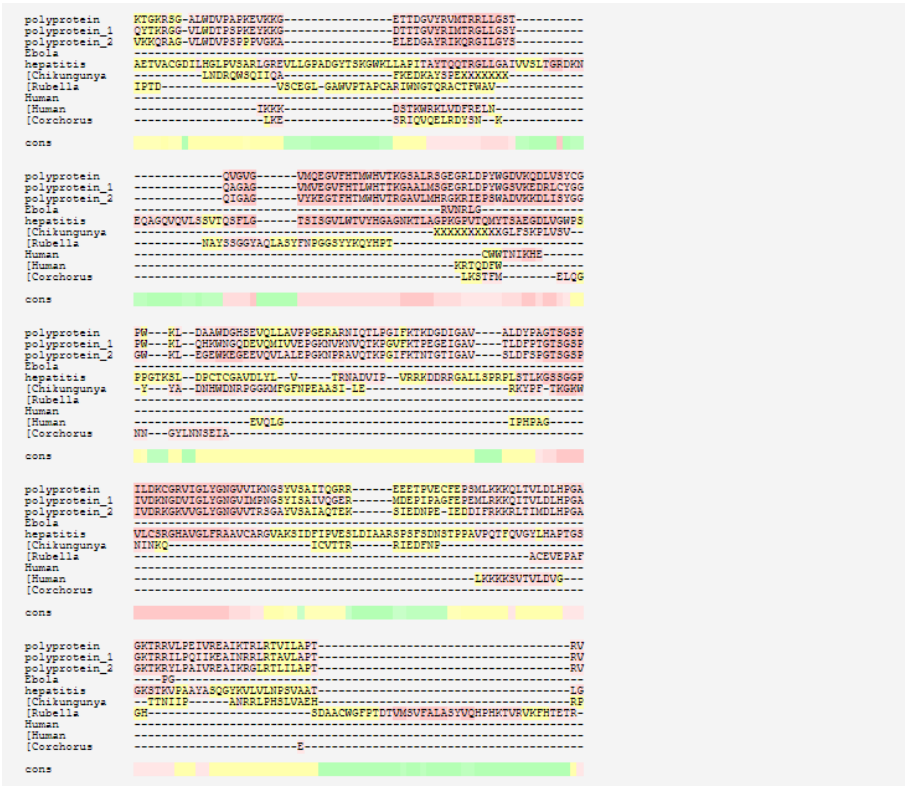


Image: 3.3(Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3(Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3(Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

Image: 3.3 (Continued): T Coffee Results Run 2 (Rearranged)

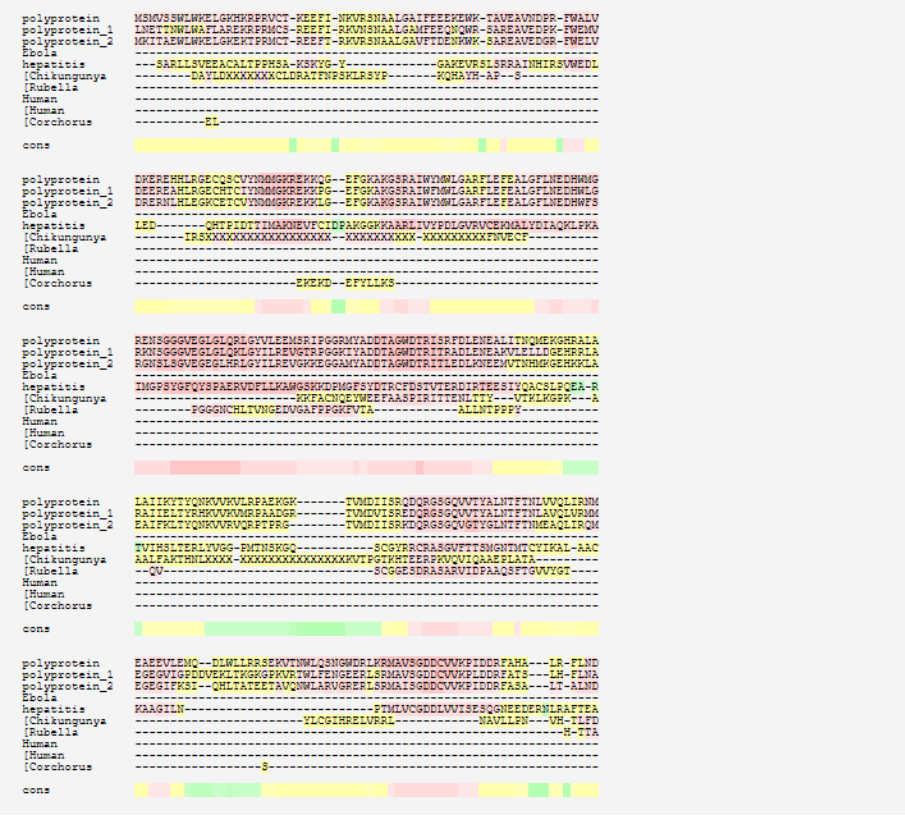


Image: 3.3 (Continued): T Coffee Run 2 (Rearranged)

Image 3.3: T Coffee Result (Rearranged): Here most similar protein sequences were grouped together. So similar sequences have clustered together in the outcome. Here 11 of the 53 regions of similarity are indicated in green boxes. They were all noted down for analysis.

3.2.1.2 COBALT Analysis Results:

MSA with the same data set was done with COBALT to compare the result. COBALT gives a graphical overview and comparative positions of alignment of query sequences. According to COBALT West Nile virus, Dengue virus, Hepatitis C virus and Chikungunya virus was most similar to Zika. Which is corresponding to the result of T Coffee. Cobalt also generates a phylogenetic tree analysing the data provided.

COBALT Results:

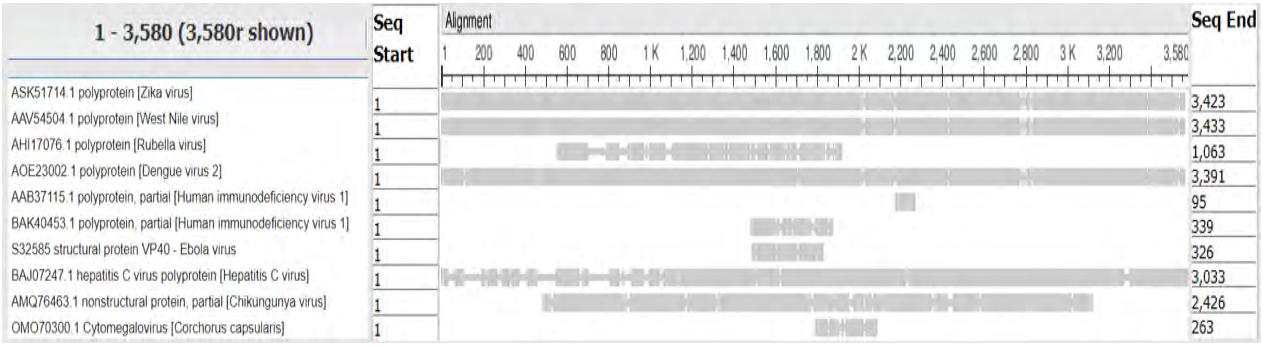


Image 3.4: Graphical Overview of analysis by COBALT. Comparison of Zika and relative species.

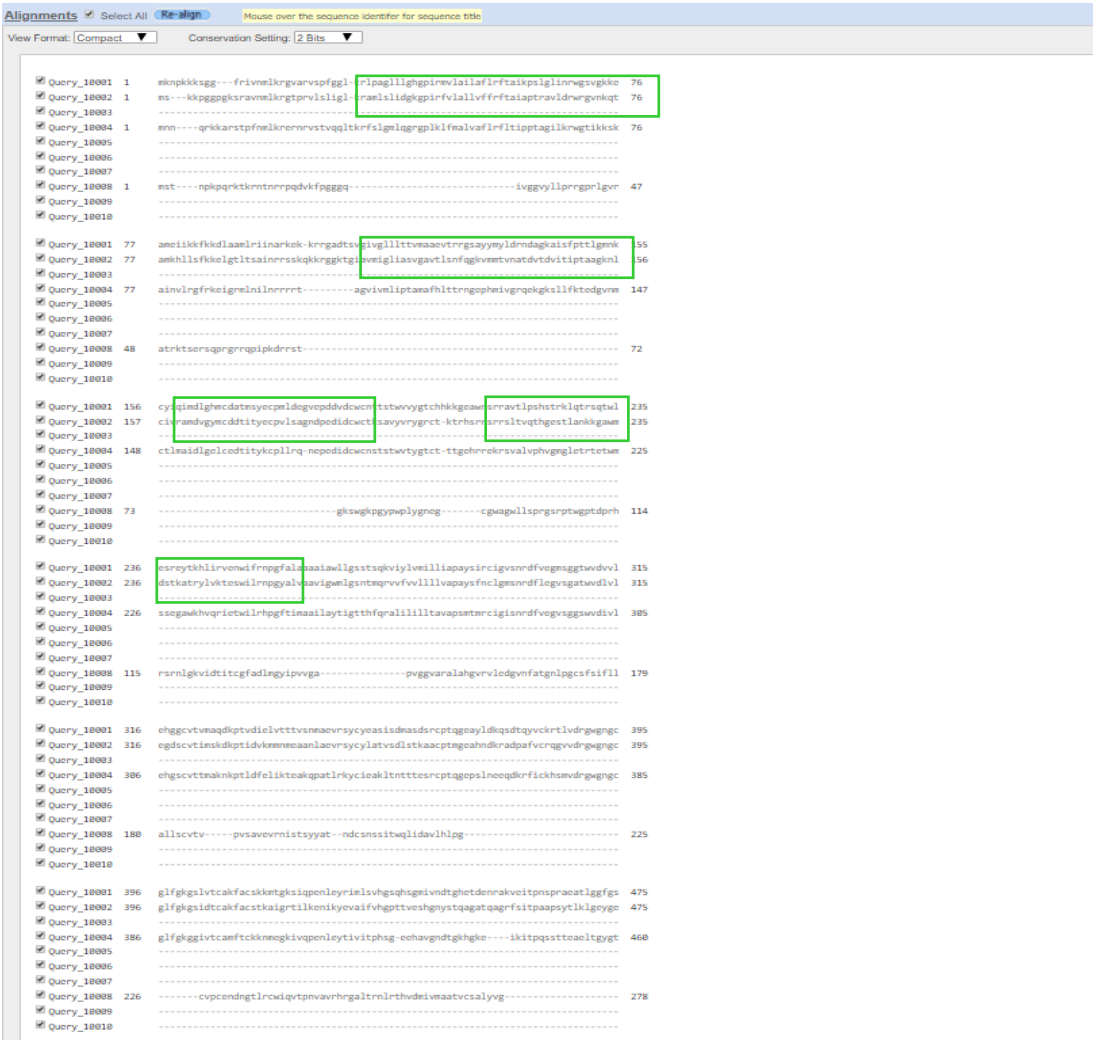


Image 3.5: Alignment by COBALT, Similar regions indicated by green

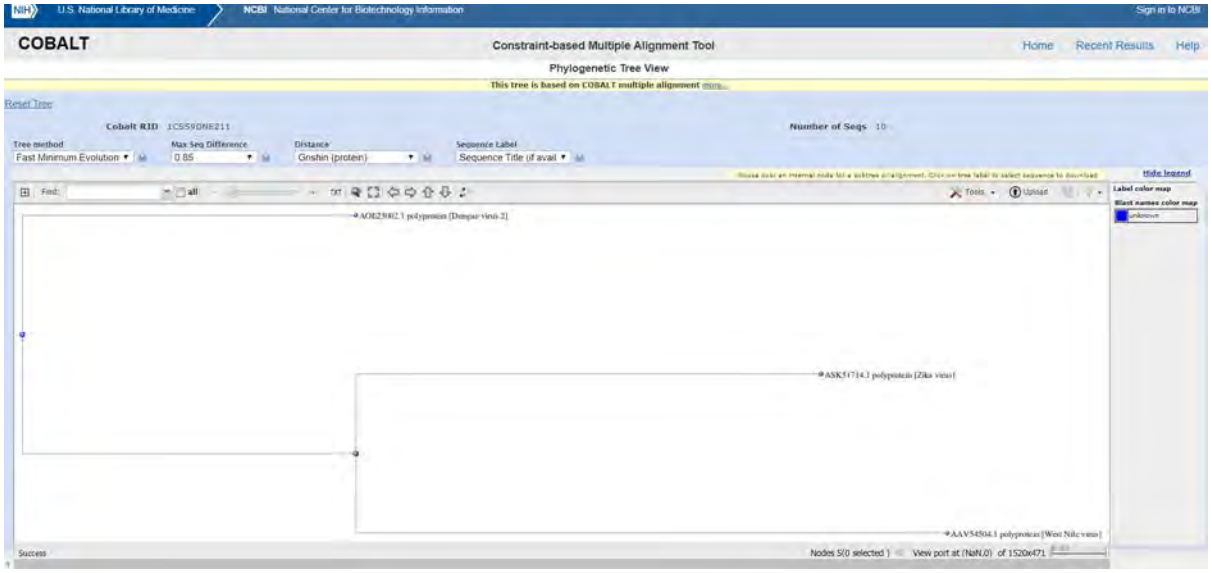


Image 3.6: Phylogenetic Tree by COBALT. Here it shows that Zika virus is genetically closest to West Nile virus. Then DENV 2 virus

3.2.2 Phylogenetic Tree Construction by MEGA:

A phylogenetic tree was constructed using MEGA. For this an alignment was done in mega format using MEGA 7 software. Then the alignment was used for constructing the tree using Maximum Likelihood method. The result is:

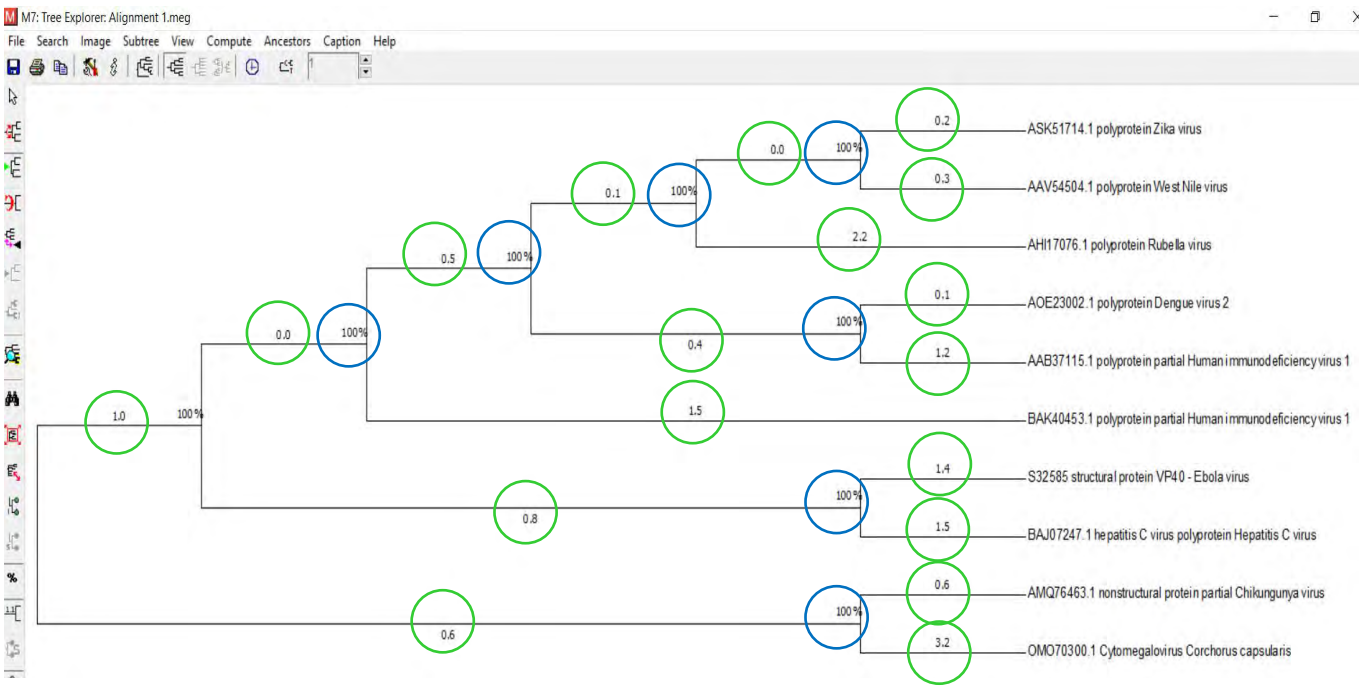


Image 3.7: Phylogenetic Tree by MEGA. Where blue circles show data coverage and green circles show branch lengths.

3.2.3 T Coffee result analysis by InterProScan:

The 'Regions of Similarity' identified from MSA were put one by one in interProScan to compare it against its repository. InterProScan compares query sequences with Europe PMC. The identified proteins are also annotated by GO terms that indicates its molecular function, biological processes or role as a cellular component. After analysis of the 53 regions by InterProsacn, the result was following:

Table 3.2: Function Determination of Conserved Regions by InterProScan

	Query	Result	Function and Indications
1.	LPAGLLLGHGPIRMVL AILAFLRFTAIKPSLG L	IPR037172 Flavivirus capsid protein C superfamily	The capsid protein C plays a role in virus budding by binding to the cell membrane and gathering the viral RNA into a nucleocapsid that forms the core of a mature virus particle
		IPR001122 Flavivirus capsid protein C	The capsid protein C is a dimeric alpha-helical protein, and its interaction with RNA is critical to produce viable virus particles
2.	VGIVGLLLTTVMAAEV TRRGS	NA	
3.	AYYMYLDRNDAGK-	NA	
4.	AISFPTTLGMNKCYI-	NA	
5.	QIMDLGHMCDATMSYE CPMLDEGVEPDDVDCW CN	IPR002535 Flavivirus polyprotein propeptide	The genome encodes one large ORF, a polyprotein which undergoes proteolytic processing into mature viral peptide chains. This entry consists of a propeptide region of approximately 90 amino acids in length.
6.	RSRRAVTLPSHSTRKL QTRSQTWLESREYTKH LIRVENWIFRNPGFAL	IPR000069 Envelope glycoprotein M, flavivirus	The envelope glycoprotein M is made as a precursor, called prM. The precursor portion of the protein is the signal peptide for the proteins entry into the membrane. prM is cleaved to form M in a late-stage cleavage event. Associated with this cleavage is a change in the infectivity and fusion activity of the virus.
7.	IAWLLGSSTSQKVIYL VMILLIAPAYS	NA	
8.	HGGCVTVMAQDKP	NA	
9.	TVDIELVTTTVSNMAE VRS	NA	
10.	EASISDMASDSRCPTQ GEAYLDKQSD	IPR036253 Flavivirus glycoprotein, central and dimerisation domains superfamily	Glycoprotein E is a class II viral fusion protein that mediates both receptor binding and fusion. Class II viral fusion proteins are found in flaviviruses and alphaviruses and are structurally distinct from class I fusion proteins. Glycoprotein E is comprised of three domains. Domain I (dimerisation domain) is an 8-stranded beta barrel
		IPR013755 Flaviviral glycoprotein, central domain, subdomain 1	Glycoprotein E is comprised of three domains. domain II (central domain) is an elongated domain composed of twelve beta strands and two alpha helices.
11.	TQYVCKRTLVD	NA	
12.	RGWGNGCGLF	Invalid (smaller than 10 aa)	
13.	GKGSLVTCAKFACSKK MTGKSIQPENLEYRIM LSVHG	IPR013755 Flaviviral glycoprotein, central domain, subdomain 1	Glycoprotein E is comprised of three domains. domain II (central domain) is an elongated domain composed of twelve beta strands and two alpha helices.

Table 3.2: Function Determination of Conserved Regions by InterProScan

		IPR036253 Flavivirus glycoprotein, central and dimerisation domains superfamily	Glycoprotein E is a class II viral fusion protein that mediates both receptor binding and fusion. Class II viral fusion proteins are found in flaviviruses and alphaviruses, and are structurally distinct from class I fusion proteins. Glycoprotein E is comprised of three domains. Domain I (dimerisation domain) is an 8-stranded beta barrel
		IPR011998 Flavivirus glycoprotein central and dimerisation domain	Flaviviruses have two envelope glycoproteins (also known as 'spike' glycoproteins), one that undergoes proteolytic cleavage to prime the virus (glycoproteins M and P62 in flaviviruses and alphaviruses, respectively), and the other to mediate receptor binding and fusion (glycoproteins E and E1 in flaviviruses and alphaviruses, respectively). Glycoprotein E/E1 is comprised of three domains: domain I (dimerisation domain) is a beta-barrel, domain II (central domain) is an elongated beta-stranded and alpha-helical domain, and domain III (immunoglobulin-like domain) is an IgC-like beta-sandwich. This entry represents the intertwined central and dimerisation domains found in flaviviral glycoprotein E [PMID: 12759475] and alphaviral glycoprotein E1 [PMID: 11301009
14.	ITPNSPRAEATLGGFG SLGLDCEPRTGLDFSD LYYLTMNNKHWLVHKE WFHDIPLPWHAGADTG TP	IPR036253 Flavivirus glycoprotein, central and dimerisation domains superfamily	Glycoprotein E is a class II viral fusion protein that mediates both receptor binding and fusion. Class II viral fusion proteins are found in flaviviruses and alphaviruses, and are structurally distinct from class I fusion proteins. Glycoprotein E is comprised of three domains. Domain I (dimerisation domain) is an 8-stranded beta barrel
		IPR013756 Glycoprotein E central domain, subdomain 2	Glycoprotein E is comprised of three domains, domain II (central domain) is an elongated domain composed of twelve beta strands and two alpha helices, Domain II can be divided into two structural components, both of which comprise alpha-beta sandwich folds, nominally referred to as regions 1 and 2. This entry represents region 2 of domain II, the central domain. The formation of trimers results in a conformational change in the hinge region of domain II, a key structural element that opens a ligand-binding hydrophobic pocket at the interface between domains I and II. The conformational change results in the exposure of a fusion peptide loop at the tip of domain II, which is required in the fusion step to drive the cellular and viral membranes together by inserting into the membrane [PMID: 12759475].
		IPR011998 Flavivirus glycoprotein central and dimerisation domain	Flaviviruses have two envelope glycoproteins (also known as 'spike' glycoproteins), one that undergoes proteolytic cleavage to prime the virus (glycoproteins M and P62 in flaviviruses and alphaviruses, respectively), and the other to mediate receptor binding and fusion (glycoproteins E and E1 in flaviviruses and alphaviruses, respectively). Glycoprotein E/E1 is comprised of three domains: domain I (dimerisation domain) is a beta-barrel, domain II (central domain) is an elongated beta-stranded and alpha-helical domain, and domain III (immunoglobulin-like domain) is an IgC-like beta-sandwich. This entry represents the intertwined central and dimerisation domains found in flaviviral glycoprotein E [PMID: 12759475] and alphaviral glycoprotein E1 [PMID: 11301009

Table 3.2: Function Determination of Conserved Regions by InterProScan

15.	HWNNKEALVEFKDAHA KRQTVVVLGSQEGAVH TALAGALEAEMDGAKG RLSSGHLKCRLKMDKL RL	IPR036253 Flavivirus glycoprotein, central and dimerisation domains superfamily	Glycoprotein E is a class II viral fusion protein that mediates both receptor binding and fusion. Class II viral fusion proteins are found in flaviviruses and alphaviruses and are structurally distinct from class I fusion proteins. Glycoprotein E is comprised of three domains. Domain I (dimerisation domain) is an 8-stranded beta barrel
		IPR011998 Flavivirus glycoprotein central and dimerisation domain	Flaviviruses have two envelope glycoproteins (also known as 'spike' glycoproteins), one that undergoes proteolytic cleavage to prime the virus (glycoproteins M and P62 in flaviviruses and alphaviruses, respectively), and the other to mediate receptor binding and fusion (glycoproteins E and E1 in flaviviruses and alphaviruses, respectively). Glycoprotein E/E1 is comprised of three domains: domain I (dimerisation domain) is a beta-barrel, domain II (central domain) is an elongated beta-stranded and alpha-helical domain, and domain III (immunoglobulin-like domain) is an IgC-like beta-sandwich. This entry represents the intertwined central and dimerisation domains found in flaviviral glycoprotein E [PMID: 12759475] and alphaviral glycoprotein E1 [PMID: 11301009
16.	LCTAAFTFTKIPAETL HGT	NA	
17.	MQTLTPVGRLITANPV ITESTENSKMMLELDP PFGDSYI	IPR000336 Flavivirus/Alphavirus glycoprotein, immunoglobulin-like domain superfamily	Glycoprotein E found in Alphavirus and Flavivirus is comprised of three domains. Domain III (immunoglobulin-like domain) is an IgC-like module with ten beta strands. This entry represents the Ig-like domain III, which contains a putative receptor-binding loop
		IPR014756 Immunoglobulin E-set	The immunoglobulin (Ig) like fold, which consists of a beta-sandwich of seven or more strands in two sheets with a greek-key topology, is one of the most common protein modules found in animals. Many different unrelated proteins share an Ig-like fold, which is often involved in interactions, commonly with other Ig-like domains via their beta-sheets [PMID: 7932691]. Of these, the "early" set (E set) domains are possibly related to the immunoglobulin (IPR007110) and/or fibronectin type III (IPR003961) Ig-like protein superfamilies.
		IPR027287 Flavivirus glycoprotein E, immunoglobulin-like domain	Glycoprotein E is comprised of three domains. Domain III (immunoglobulin-like domain) is an IgC-like module with ten beta strands. This entry represents the Ig-like domain III, which contains a putative receptor-binding loop
18.	VVDGDTLKECPLNHRA WNSFLVEDHGFGVFHT SVWLKVREDYSLECDP AVIGTAVKGKEAVHSD L	IPR001157 Flavivirus non-structural protein NS1	The flavivirus nonstructural glycoprotein NS1 is an enigmatic protein whose structure and mechanisticfunction have remained somewhat elusive ever since it was first reported in 1970. NS1 is a major player in diagnosis, viral replication, protection and pathogenesis. It is found in cell-associated membrane forms as well as a secreted lipoparticle. NS1 engages with a range of complement factors to subvert the immune response.
19.	TWRLKRAHLIEMKTCE WPKSHTLWTDGIEESD LIIPKS	IPR001157 Flavivirus non-structural protein NS1	The flavivirus nonstructural glycoprotein NS1 is an enigmatic protein whose structure and mechanisticfunction have remained somewhat elusive ever since it was first reported in 1970. NS1 is a major player in diagnosis, viral replication, protection and pathogenesis. It is found in cell-associated membrane forms as well as a secreted lipoparticle. NS1 engages with a range of complement factors to subvert the immune response.
20.	TREGYRTQMKGPWHSE ELEIRFEE	NA	

Table 3.2: Function Determination of Conserved Regions by InterProScan

21.	ALAWLAIRAMVVPR	NA	
22.	AWRAGLATCGGFM	NA	
23.	MAGPMAAVGLLIVSYVVS	NA	
24.	ALWDVPAPKEVKKG	NA	
25.	ALDYPAGTSGSP	NA	
26.	PSMLKKKQLTVLDLHPGA	NA	
27.	GKTRRVLPEIVREAIKTRLR	NA	
28.	TKAGKRVIQLSRKTFETEFQKT	NA	
29.	FVVTTDISEMGANFKADRVID	NA	
30.	RHGEKRVLKPRWMDARVCSDHAALKSFKEF	IPR027417 P-loop containing nucleoside triphosphate hydrolase	The P-loop NTPase fold is the most prevalent domain of the several distinct nucleotide-binding protein folds. The most common reaction catalysed by enzymes of the P-loop NTPase fold is the hydrolysis of the beta-gamma phosphate bond of a bound nucleoside triphosphate (NTP). P-loop NTPases are characterised by two conserved sequence signatures, the Walker A motif (the P-loop proper) and Walker B motifs which bind, respectively, the beta and gamma phosphate moieties of the bound NTP, and a Mg2+ cation.
31.	AAGKRGAAFGVMEALGTLPGHM	NA	
33.	ETGSRPYKAAAAQLPETLETIMLLGLLGTVS	IPR000404 Flavivirus non-structural protein NS4A	The NS4A protein is small and poorly conserved among the Flaviviruses. NS4A contains multiple hydrophobic potential membrane spanning regions [PMID: 2174669]. NS4A has only been found in cells infected by Kunjin virus
34.	LGIFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLI	IPR000404 Flavivirus non-structural protein NS4A	The NS4A protein is small and poorly conserved among the Flaviviruses. NS4A contains multiple hydrophobic potential membrane spanning regions [PMID: 2174669]. NS4A has only been found in cells infected by Kunjin virus
35.	PQDNQMAIIIMVAVGLLGLITANE	NA	
36.	LGWLERTKSDLSHLMGR	NA	
37.	IDLRPASAWAIYAALTTFITPAVQHAVTT	IPR001528 Flavivirus non-structural protein NS4B	The NS4A protein is small and poorly conserved among the Flaviviruses. NS4A contains multiple hydrophobic potential membrane spanning regions [PMID: 2174669]. NS4A has only been found in cells infected by Kunjin virus
38.	ATQAGVLFGMGKGMPFYTWDFGVPLLMIGCYSQ	IPR001528 Flavivirus non-structural protein NS4B	The NS4A protein is small and poorly conserved among the Flaviviruses. NS4A contains multiple hydrophobic potential membrane spanning regions [PMID: 2174669]. NS4A has only been found in cells infected by Kunjin virus
39.	AIILLVAHYMYLIPGLQAA	NA	
40.	AARAAQKRTAAG	NA	
41.	SGAKSNTIKSV	NA	
42.	STTSQLLLGRMDGPR	NA	
43.	YEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEH	NA	
44.	AETWFFDENHPYRTWAYHGSYE	IPR000208 RNA-directed RNA polymerase, flavivirus	RNA-directed RNA polymerase (RdRp) (EC:2.7.7.48) is an essential protein encoded in the genomes of all RNA containing viruses with no DNA stage [PMID: 2759231, PMID: 8709232]. It catalyses synthesis of the RNA strand complementary to a given RNA template, but the precise molecular mechanism remains unclear.

Table 3.2: Function Determination of Conserved Regions by InterProScan

45.	SASSLVNGVVRLLSKPWD VVTGVTGIAMTDTTPYG	IPR000208 RNA-directed RNA polymerase, flavivirus	do
46.	QQRVFKEKVDTRVPDP QEGTRQV	NA	
47.	VYNMMGKREKKQG	NA	
48.	EFGKAKGSRAIWYMWL GARFLEFEALGFLNED HWMG	IPR000208 RNA-directed RNA polymerase, flavivirus	do
49.	QDQRGSGQVVTYALNT FTNL	IPR000208 RNA-directed RNA polymerase, flavivirus IPR007094	do
		RNA-directed RNA polymerase, catalytic domain	do
50.	EKVTNWLQSNGWDRLK RMAVSGDDCVVKPIDD RFAHA	IPR000208 RNA-directed RNA polymerase, flavivirus	do
51.	MGKVRKDTQEWKPSTG	NA	
52.	PCRHQDELIGRA	NA	
53.	RVSPGAGWSIRETACL AKSYAQM	IPR000208 RNA-directed RNA polymerase, flavivirus	do

3.2.4 Virus Polyprotein analysis by ScanProsite:

All the protein sequences were analysed by ScanProSite for further determining their details. Besides functions ScanProsite identifies its location and other characteristics like if it's an active site, domain, binding site etc. The results of Virus Polyprotein analysis by ScanProsite are following:

Table 3.3: Analysis of Virus Polyprotein by ScanProstie. Hits for all PROSITE (release 2017_11) motifs on sequences AAB37115-1, AAV54504-1, AHI17076-1, AMQ76463-1, AOE23002-1, ASK51714-1, BAJ07247-1, BAK40453-1, OMO70300-1, S32585

Here, Upper case represents match positions, lower case inserts positions, and the '-' symbol represents deletions relative to the matching profile.

Num	Name	Sequence	Predicted Features		
			Indication	AA	Function
1.	PS51527 **FLAVIVIRUS_NS2B**	SWPPSEVLTAVGLICALAGGFAKADI-EMAGPMAAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMAICGMNPIAIPFAAGAWYVYVKTGKR	Region	1425-1464	Interacts with and activates NS3 Protease
2.	PS51528 **FLAVIVIRUS_NS3PRO**	SGALWDVPAPKEVKKGE--TTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMW**H**VTKGSALRSGEGRLDPYWGDVKQ**D**LVSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGT**S**GSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVEC	Domain	1503-1680	Peptidase S7
			Act Site	1553	Charge relay system for serine protease NS3 activity
			Act Site	1577	do
			Act Site	1637	do
3.	PS51192 **HELICASE_ATP_BIND_1**	PSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAI-KTRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAV----------NVTHSGTEIVDLMCHATFTSRLLQPIR-VPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTE	NA	NA	NA
4.	PS51194 **HELICASE_CTER**	PIMDTEVEVPERawssgfdwVTDHSGKTVWFVPSVRNGNEIAACLTK--------AgKRVIQLSRKTFETEFqktKHQEWDFVVTTDISEMGANFKADRVIDsrrclkpvILDGERvilaGPmpvthASAAQRRGRIGRNPNkPGDEYLYGGGcaeTDEDHAHWLEaRMLLDNIylqDGLIASLYRPE	Domain	1834-2013	Helicase C Terminal
5.	PS51591 **RNA_CAP01_NS5_MT**	GGGTGETLGEKW**K**AR**LN**Q**M**SALE**F**YSY**K**KSGITEVCREEARRALKDGVATGGHAV**S**RGSA**K**LRWLVERGYLQPYGKVIDLGCGRG**GW**SYYAATIRKVQEVKGY**TK**GGPGHEEPVLVQSYGWNIVRLKSGV**DV**FHMAAEPCDTLLC**DI**GE**S**SSSPEVEEARTLRVLSMVGDWLEKRPGAFCI**K**VLCPYTSTMMETLERLQRRYGGGLVRVPLS**R**N**S**TH**E**M**Y**WVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRA	Domain, Binding, Binding Site	2521-2740	mRNA cap, NS5 type MT, mRNA cap by carbonyl oxygen, mRNA cap binding, S adenosyl L Methionine via carbonyl oxygen, S adenosyl L Methionine, Essential for 2- O- methyltransferase activity,
6.	PS50507 **RDRP_SSRNA_POS**	GRMYADDTAGWDTRISRFDLENEALITNQMEKGHRALALAIIKYtyqnkvvKVLRPAEKGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVqLIRNMEAeEVLEmqdlwllrrsekvtnwlqsngwdRLKRMAVSGDDCVVKPIDDRFAHA	NA	NA	NA

Table 3.3: Analysis of Virus Polyprotein by ScanProstie. Hits for all PROSITE (release 2017_11) motifs on sequences AAB37115-1, AAV54504-1, AHI17076-1, AMQ76463-1, AOE23002-1, ASK51714-1, BAJ07247-1, BAK40453-1, OMO70300-1, S32585

7.	PS51693 **HCV_NS2_PRO**	RASLLRVPY-FVRAHALLRVCTLVRHLAgARYIQMLLITIGRWTGTYIYD**H**LSPLSTWAA QGLRDLAVAV**E**PVVFSPMEKKVIVWGAETVA**C**GDILHGLPVSARLGREVLLGPADGYTSK GWKLL	Domain	907- 1030	Peptidase C 18
			Act Site	956	For protease NS2- 3 activity shared with dimeric partner
			Act Site	976	do
			Act Site	997	do
8.	PS51822 **HV_PV_NS3_PRO**	APITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQSFLGTSISGVLWTVY**H**GAG NKTLAGPKGPVTQMYTSAEG**D**LVGWPSPPGTKSLDP**C**T**C**GAVDLYLVTRNADVIPVRRKD DRRGALLSPRPLSTLKGS**S**GGPVL**C**SRG**H**AVGLFRAAVCARGVAKSIDFIPVESLDIAAR SP	Domain	1031- 1212	Peptidase S29
			Act Site	1087	Charge relay system for NS3 activity
			Act Site	1111	do
			Metal	1127	Zinc
			Metal	1129	Zinc
			Act Site	1169	Charge relay system for NS3 activity
			Metal	1175	Zinc
			Metal	1179	Zinc
9.	PS51743 **ALPHAVIRUS_MT**	------------------------ KLIEQEidpDSTILDIGSAPARRMMSDRK--YHCV CPMRSAED---- PERLANyarklasaagkvldRNISGkigdlqavmaVPdtetptfCLHTDVSCRQRADVAIYQDvyAVHAPTSLYHQAI------ KGVRVaYWVGFDTTPFMYNamaGa xxxxxxxxxxxxxxxxxxxxxxxdltegRRGKLSIMRGKKLKPCDRVLFSVG---STLY PESRKLLKSWH	Domain	1- 211	Alphavirus like MT
10.	PS51520 **NSP2PRO**	DTFQNKANV**C**WAKSLVPILETAGIKLNDRQWSQIIQAFKEDKAYSPEXXXXXXXXXXXXX XXXXGLFSKPLVSVYYADN**H**WDNRPGGKMF GFNPEAASILERKYPFTKGKWNINKQICVT TRRIEDFNPTTNIIPANRRLPHSLVAEHRP VKGERMEWLVNKINGHHVLLVSGYNLVLPT KRVTWVAPLGVRGADYTYNLELGLPATLGR YDLVVINIHTPFRIHHYQQCVDHAMKLQML GGDSLRLLKPGGSLLIRAYGYADRTSERVI CVLGRKFRSSRALKPPCVTSNTEMFFLFSN FDNGRRNFTTHVMNNQLNAAFVGQ	Domain	956- 1279	Peptidase C9
			Act Site	965	For cystine protease nsP2 activity
			Act Site	1035	do
11.	PS50994 **INTEGRASE**	------------------------------ ----------IPAETGQETAYFLLKLAAR -- WPVKVIHTDNGPNFTSATMKAACWWTNIKH EFGIPYNPQRQGVVEAMNKELKSIIQQV RDQA--EQLKTAVQMAVFVH---------- ----------	Domain	1- 95	Integrase catalytic
12.	PS50175 **ASP_PROT_RETROV**	KEALL**D**TGADDTVLEEMNLPGRWKPK-- MIGGIGGFIKVRQYDQVXIEICGHKAMGTV LV GPTPVNIIGRNL	Domain	20- 89	Peptidase A2
			Act Site	25	Target activity visibility for protease activity
13.	PS50878 **RT_POL**	EGKISKIGPENPYNTPVFAIKKKDSTkWRK LVDFRELNKRTQDFWEVQLGIPHPAGLKKK KSVTVLDVGDAYFSVPLDKDFRKYTAFTIP svnnetpGIRYQYNVLPQGWKGSPAIFQCS MTKILEPFRKQNpdivIYQYMDDLYVGSDL EIGqHRTKIEELRQHLLRWGFTTPDKKHQK EPPFLWMGYEL	Domain	143- 333	Reverse Transcriptase
			Metal	209	Magnesium, Catalytic
			Metal	284	Do
			Metal	285	Do

3.2.5 GO Term Prediction:

A task of GO term prediction with InterProScan with only the polyprotein sequence of Zika Virus was done to summarize the overall functions that occur in Zika virus particle.

The result was:

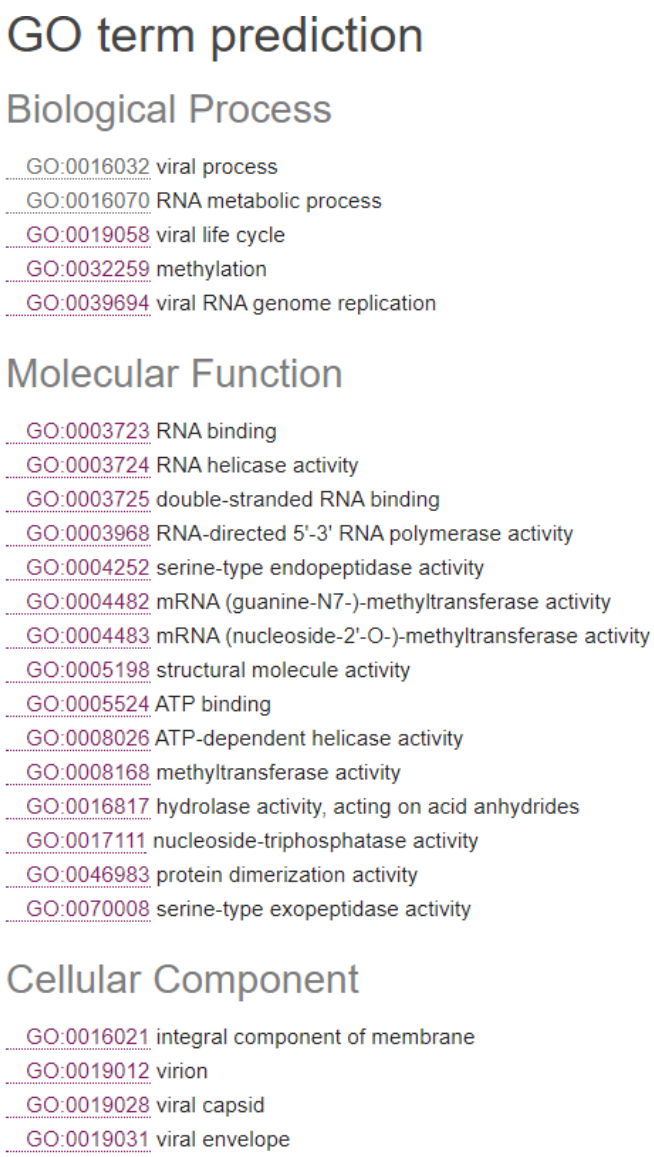


Image 3.8: GO Term Prediction

Later more details about all the biological processes that occur in Zika Virus were retrieved using the GO IDs and searching the GO Database.

Biological Process:

GO:0016032 **Viral process**

A multi-organism process in which a virus is a participant. The other participant is the host. Includes infection of a host cell, replication of the viral genome, and assembly of progeny virus particles. In some cases, the viral genetic material may integrate into the host genome and only subsequently, under particular circumstances, 'complete' its life cycle.

GO:0016070 **RNA metabolic process**

The cellular chemical reactions and pathways involving RNA, ribonucleic acid, one of the two main type of nucleic acid, consisting of a long, unbranched macromolecule formed from ribonucleotides joined in 3',5'-phosphodiester linkage.

GO:0019058 **Viral life cycle**

A set of processes which all viruses follow to ensure survival; includes attachment and entry of the virus particle, decoding of genome information, translation of viral mRNA by host ribosomes, genome replication, and assembly and release of viral particles containing the genome.

GO:0032259 **Methylation**

The process in which a methyl group is covalently attached to a molecule.

GO:0039694 **Viral RNA genome replication**

The replication of a viral RNA genome.

Molecular Function:

GO:0003723 **RNA binding**

Interacting selectively and non-covalently with an RNA molecule or a portion thereof.

GO:0003724 **RNA helicase activity**

Catalysis of the reaction: NTP + H2O = NDP + phosphate, to drive the unwinding of a RNA helix.

GO:0003725 **double-stranded RNA binding**

Interacting selectively and non-covalently with double-stranded RNA.

GO:0003968 **RNA-directed 5'-3' RNA polymerase activity**

Catalysis of the reaction: nucleoside triphosphate + RNA(n) = diphosphate + RNA(n+1); uses an RNA template, i.e. the catalysis of RNA-template-directed extension of the 3'-end of an RNA strand by one nucleotide at a time.

GO:0004252 **serine-type endopeptidase activity**

Catalysis of the hydrolysis of internal, alpha-peptide bonds in a polypeptide chain by a catalytic mechanism that involves a catalytic triad consisting of a serine nucleophile that is activated by a proton relay involving an acidic residue (e.g. aspartate or glutamate) and a basic residue (usually histidine).

GO:0004482 **mRNA (guanine-N7-)-methyltransferase activity**

Catalysis of the reaction: S-adenosyl-L-methionine + G(5')pppR-RNA = S-adenosyl-L-homocysteine + m7G(5')pppR-RNA. m7G(5')pppR-RNA is mRNA containing an N7-methylguanine cap; R may be guanosine or adenosine.

GO:0004483 **mRNA (nucleoside-2'-O-)-methyltransferase activity**

Catalysis of the reaction: S-adenosyl-L-methionine + m7G(5')pppR-RNA = S-adenosyl-L-homocysteine + m7G(5')pppRm-RNA. R may be guanosine or adenosine.

GO:0005198 **structural molecule activity**

The action of a molecule that contributes to the structural integrity of a complex or its assembly within or outside a cell.

GO:0005524 **ATP binding**

Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.

GO:0008026 **ATP-dependent helicase activity**

Catalysis of the reaction: ATP + H2O = ADP + phosphate, to drive the unwinding of a DNA or RNA helix.

GO:0008168 **methyltransferase activity**

Catalysis of the transfer of a methyl group to an acceptor molecule.

GO:0016817 **hydrolase activity, acting on acid anhydrides**

Catalysis of the hydrolysis of any acid anhydride.

GO:0017111 **nucleoside-triphosphatase activity**

Catalysis of the reaction: a nucleoside triphosphate + H2O = nucleoside diphosphate + phosphate.

GO:0046983 **protein dimerization activity**

The formation of a protein dimer, a macromolecular structure consists of two noncovalently associated identical or nonidentical subunits.

GO:0070008 **serine-type exopeptidase activity**

Catalysis of the hydrolysis of a peptide bond not more than three residues from the N- or C-terminus of a polypeptide chain by a catalytic mechanism that involves a catalytic triad consisting of a serine nucleophile that is activated by a proton relay involving an acidic residue (e.g. aspartate or glutamate) and a basic residue (usually histidine).

Cellular Component

GO:0016021 **integral component of membrane**

The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane.

GO:0019012 **virion**

The complete fully infectious extracellular virus particle.

GO:0019028 **viral capsid**

The protein coat that surrounds the infective nucleic acid in some virus particles. It comprises numerous regularly arranged subunits, or capsomeres.

GO:0019031 **viral envelope**

The lipid bilayer of a virion that surrounds the protein capsid. May also contain glycoproteins.

3.3 EPITOPE PREDICTION FOR VACCINE DESIGN:

The softwares used for epitope prediction are Vaxijen, BepiPred, BCPreds and IEDB Tools.

Zika virus membrane glycoprotein M was chosen as the target sequence because of maximum sequence data availability. The whole sequence is of 75 amino acid. In FASTA format the sequence looks like:

>YP_009430299.1 membrane glycoprotein M [Zika virus]

AVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTS
QKVIYLVMILLIAPAYS

This sequence was later analysed by all the different Softwares:

3.3.1 Vaxigen result for M Protein: Primary result of antigenicity:

Vaxijen is a software developed by Edward Jenner Institute and it calculates the antigenicity of a sequence using ACC (Auto Cross Covariance) function. When M protein sequence was run in Vaxijen, the result of antigenicity was 0.4358, which is above the threshold value of ≥0.4. This indicates that M protein is has good antigenic property.

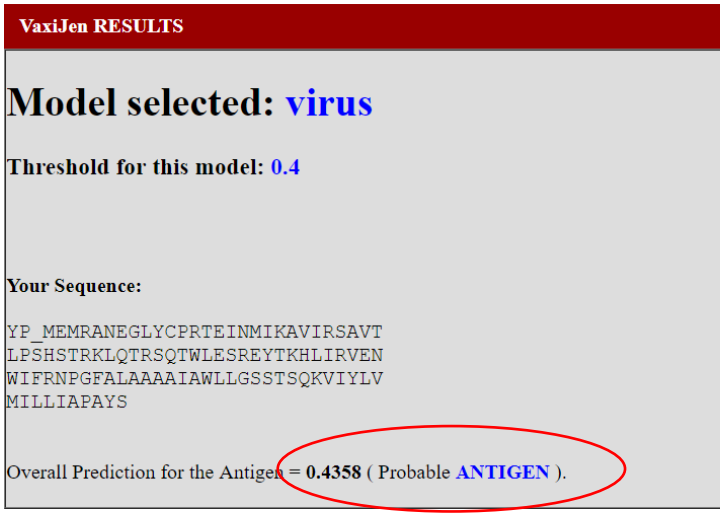


Image 3.9: Analysis of M Protein by Vaxijen. Preliminary analysis shows that the sequence is probably antigenic

3.3.2 BepiPred: Prediction of B Cell epitope

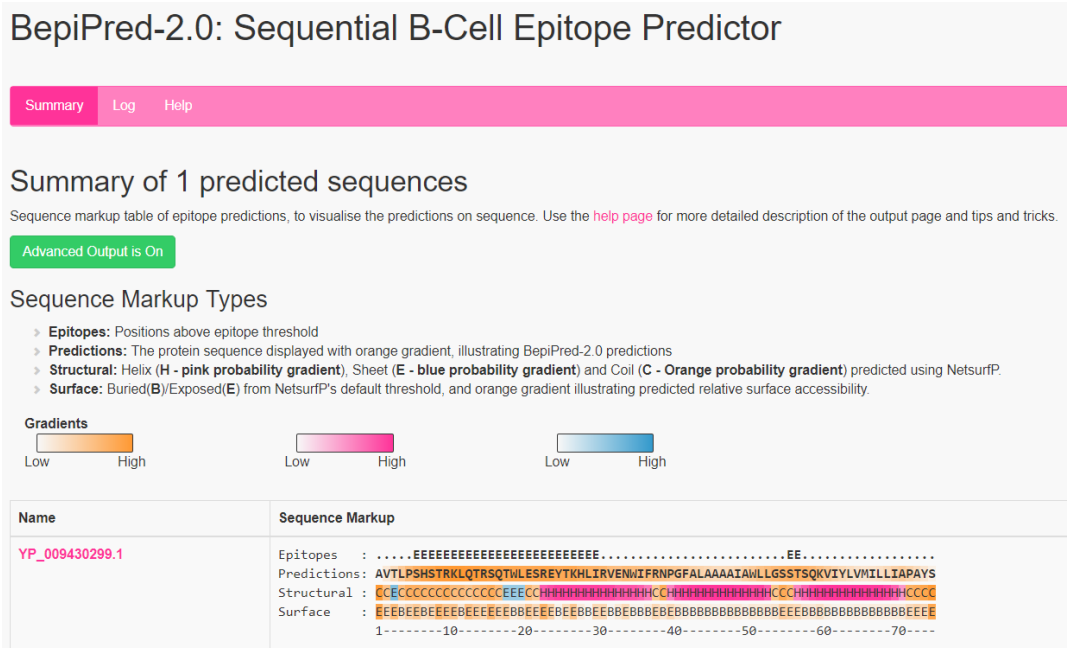
After the antigenicity of M protein was confirmed, the sequence was put in BepiPred B Cell epitope Predicter to identify probable B Cell epitopes in a sequence. This software also predicts the proteins properties like structure (Helix/ Coil/ Sheet), surface accessibility (Exposed/

Buried) taking each residue of amino acid into account. Different threshold values can be set to identify desired epitopes. This threshold value is a correlation of specificity and sensitivity. The more the value of specificity, the less the number of epitopes.

Here, four different threshold values were taken for getting maximum numbers of epitope.

BepiPred Epitope Prediction:

(A) Threshold: 0.50



(B) Threshold: 0.48

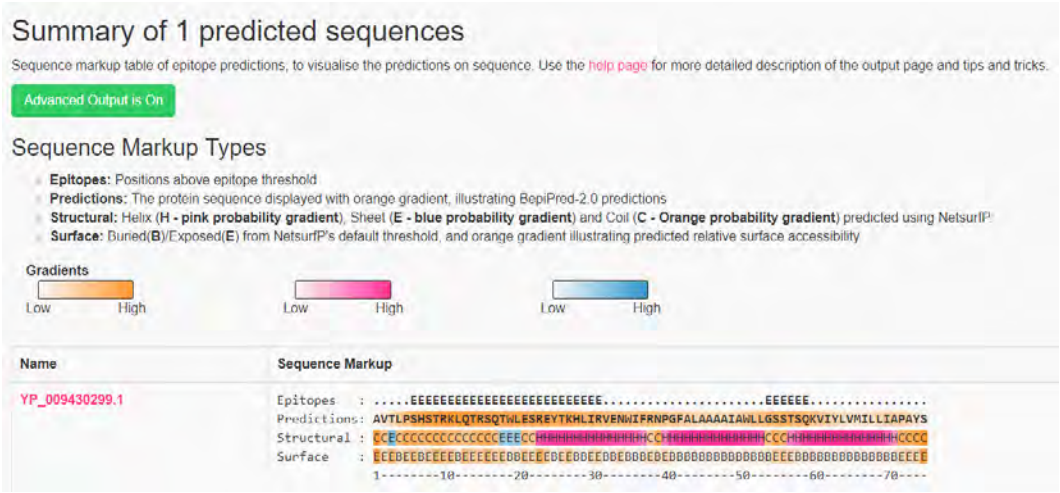
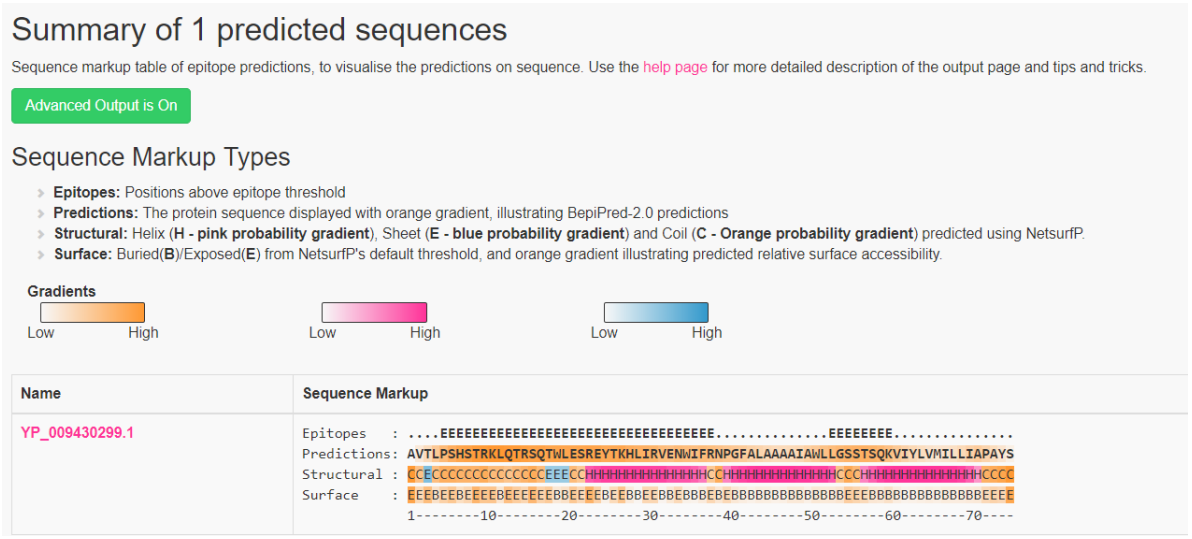


Image 3.10: Epitope prediction by BepiPred with Different Threshold value. (A) Threshold value 0.50, (B) Threshold Value 0.48, (C) Threshold value (D) Threshold value 0.40

(C) Threshold: 0.45



(D) Threshold: 0.40

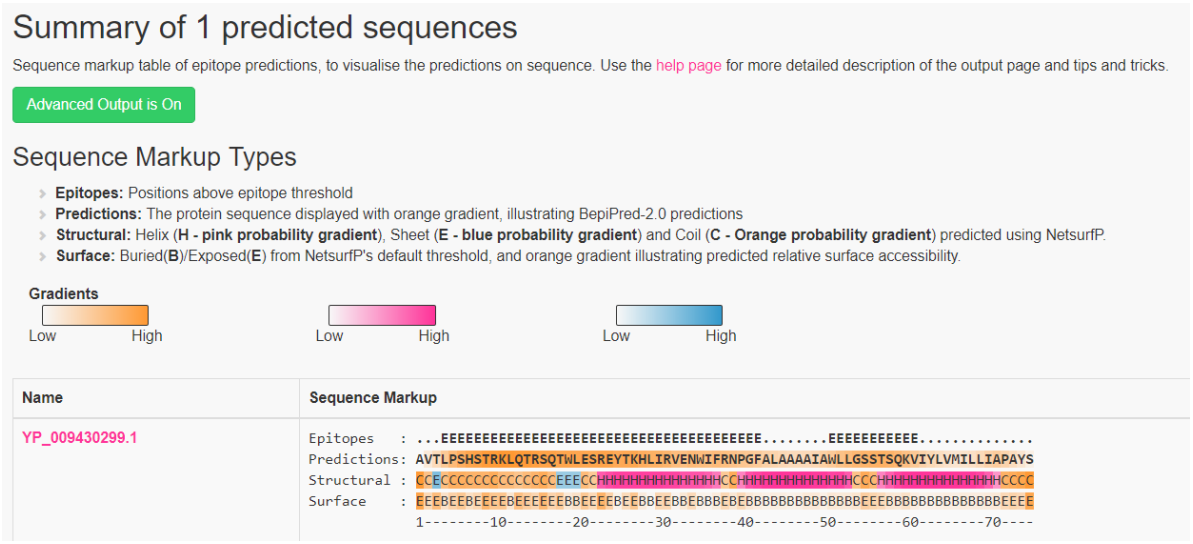


Image 3.10: Epitope prediction by BepiPred with different threshold value(A) Threshold value 0.50, (B) Threshold Value 0.48, (C) Threshold value (D) Threshold value 0.40

Image 3.11: Epitope Threshold Guidance for BepiPred

3.3.3 BCPREDS: B Cell epitope prediction

Selection of B cell epitopes was also done using BCPred software. Here for the selection, epitope length was considered. Epitopes of 12 and 20 amino acid in length were desired.

BCPreds Epitope Prediction

(A) Epitope length: 12

BCPREDS Server 1.0

Submitted sequence: 75 amino acids
Epitope length: 12 amino acids
Classifier Specificty: 75%
Prediction method: bcpred
Use overlap filter: yes

BCPred Predictions

Position	Epitope	Score
5	PSHSTRKLQTRS	0.66
54	GSSTSQKVIYLV	0.624
33	ENWIFRNPGFAL	0.536
19	WLESREYTKHLI	0.289

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
AVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQK 60
....EEEEEEEEEEE..EEEEEEEEEEE..EEEEEEEEEEE.........EEEEEEE
VIYLVMILLIAPAYS 75
EEEEE..........
```

(B) Epitope length: 20

Submitted sequence: 75 amino acids
Epitope length: 20 amino acids
Classifier Specificty: 75%
Prediction method: bcpred
Use overlap filter: yes

BCPred Predictions

Position	Epitope	Score
25	YTKHLIRVENWIFRNPGFAL	0.756

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
AVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQK 60
.........................EEEEEEEEEEEEEEEEEEE.................
VIYLVMILLIAPAYS 75
...............
```

Image 3.12: Epitope Prediction by BCpred by taking Different Lengths, (A) 12, (B) 20

3.3.4 Predicting the Epitopes:

In Vaxigen server, there is a different threshold value for every organism such as bacteria, virus, fungi, tumour etc. Any input that has a score above that threshold value is antigenic. The more the score, the more the probability of being an antigen. For viruses this value is: ≥0.4, Above this value, sequence is considered antigenic. All the epitopes that were derived from

different softwares under different conditions were all examined by Vaxijen differently for checking antigenicity.

Table 3.4: All predicted epitopes after being checked by Vaxijen. Antigens having a threshold value or above marked blue:

Tool	Sequence	Vaxigen Score	Antigenic?
BepiPred (T 0.5)	SHSTRKLQTRSQTWLESREYTKHLI	0.3429	No
BepiPred (T 0.48)	SHSTRKLQTRSQTWLESREYTKHLIR	-0.0190	No
	GSSTSQK	1.1341	Yes
BepiPred (T 0.45)	PSHSTRKLQTRSQTWLESREYTKHLIRVENWIFR	0.1514	No
	LGSSTSQK	1.1401	Yes
BepiPred (T 0.40)	LPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGF	0.2445	No
	WLLGSSTSQKV	0.5346	Yes
BCPreds (12)	PSHSTRKLQTRS	0.1220	No
	GSSTSQKVIYLV	0.5978	Yes
	ENWIFRNPGFAL	0.8161	Yes
	WLESREYTKHLI	0.3746	No
BCPreds (20)	YTKHLIRVENWIFRNPGFAL	0.3429	No

Here 5 Sequences appear to have a value of antigenicity above 0.4. These epitopes are:

Table 3.5: Epitopes having above the Threshold Value of Vaxijen

Sl	Sequence	Value
1.	GSSTSQK	1.1341
2.	LGSSTSQK	1.1401
3.	WLLGSSTSQKV	0.5346
4.	GSSTSQKVIYLV	0.5978
5.	ENWIFRNPGFAL	0.8161

From this table it can be seen that the first four sequences have one sequences in common. That is **GSSTSQK.** So, this common sequence was chosen as one of the ideal epitopes. And the fifth one is **ENWIFRNPGFAL** which has a good antigenicity value of itself.

So, altogether, the predicted B Cell Epitopes are,

1. **GSSTSQK,**

Overall Prediction for the Antigen = 1.1341 (Probable ANTIGEN by Vaxijen).

2. **ENWIFRNPGFAL**

Overall Prediction for the Antigen = 0.8161 (Probable ANTIGEN by Vaxijen).

3.4 CHECKING THE EPITOPES AS IDEAL VACCINE CANDIDATE:

For being an ideal vaccine, an epitope should have properties like

- conservancy,
- hydrophobicity,
- surface accessibility,
- flexibility,
- antigenicity etc.

These properties of both the candidate epitopes were examined by IEDB conservancy tools and B cell tools.

3.4.1 Epitope Conservancy Test:

Epitope conservancy is important for designing a sensitive vaccine. If an epitope is conserved a more precise vaccine can be prepared with it. Here both the epitopes were found to be 100% conserved in terms of both M protein and the entire ZIKV polyprotein. Which means both the sequences are present at only one location on the polyprotein sequence.

Result of Epitope Conservancy analysis:

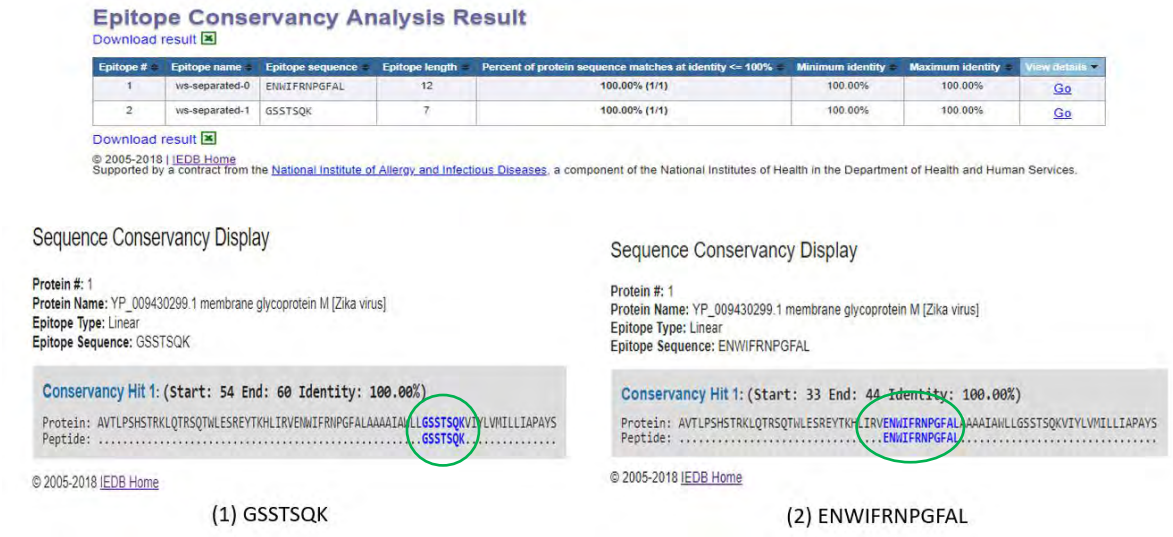


Image 3.13: Epitope Conservancy in terms of ZIKV M protein (Both 100% conserved)

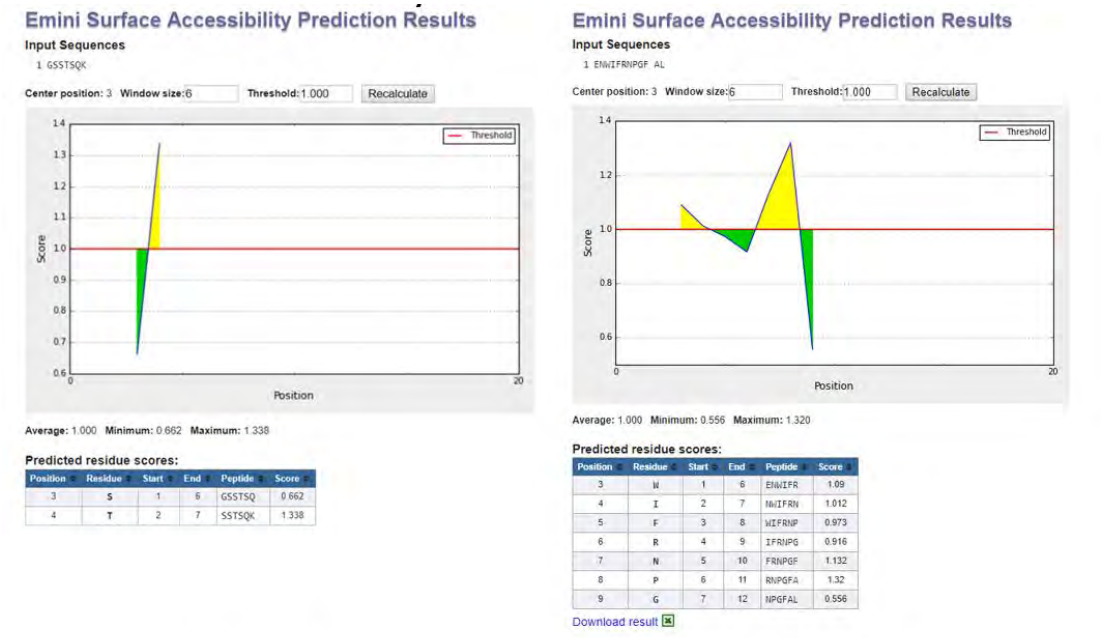
(1) GSSTSQK

(2) ENWIFRNPGFAL

Image 3.14: Epitope Conservancy in terms of ZIKA Polyprotein (Both 100% conserved)

3.4.2 Surface Accessibility Test:

Surface accessibly is very important for a vaccine candidate. Because antibody should bind to the exposed residues of a peptide. Thus, for being a good vaccine, a peptide should have surface accessibility. Here the test was done with IEDB Emnini Surface Accessibility Prediction tool where score above a threshold value is considered to be surface accessible. Where, for residue 1, two amino acids out of seven were found to be surface accessible and for residue 2, seven out of twelve amino acids were surface accessible.



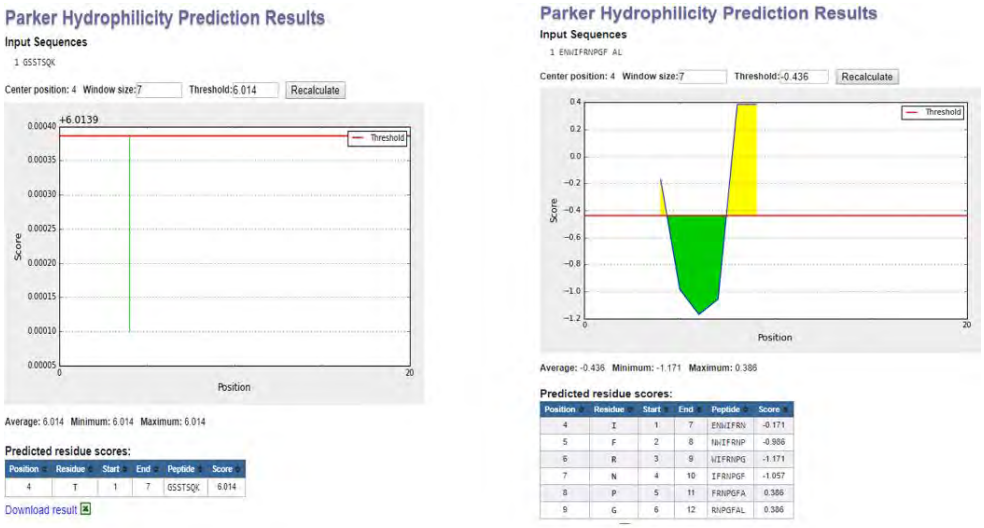
(1) GSSTSQK (2) ENWIFRNPGFAL

Image 3. 15: Surface accessibility of epitopes

3.5.3 Hydrophobicity Test:

For being a good vaccine, a peptide should have hydrophobic properties. Because hydrophobicity helps to maintain the purity of a vaccine by preventing its interaction with soluble substances. Here the test for hydrophobicity was done with IEDB Parker Hydrophobicity Prediction tool where hydrophobicity is indicated by a threshold score. Here, in epitope 1, one residue was found to be hydrophobic and had the marginal threshold score. But in epitope 2, six out of 12 residues were found to be hydrophobic indicating a good hydrophobicity.

Parker Hydrophobicity Prediction Result:



(1) GSSTSQK (2) ENWIFRNPGFAL

Image 3.16: Hydrophobicity of Epitopes

3.5.4 Flexibility Test:

A good vaccine should be flexible because antibody needs to get access to it. Therefore, Karplus and Schulz flexibility test was done by IEDB tool. The tool showed error result for epitope 1 but for epitope 2, it showed a good result putting five out of twelve residues above the threshold value of flexibility.

Flexibility Prediction Result:

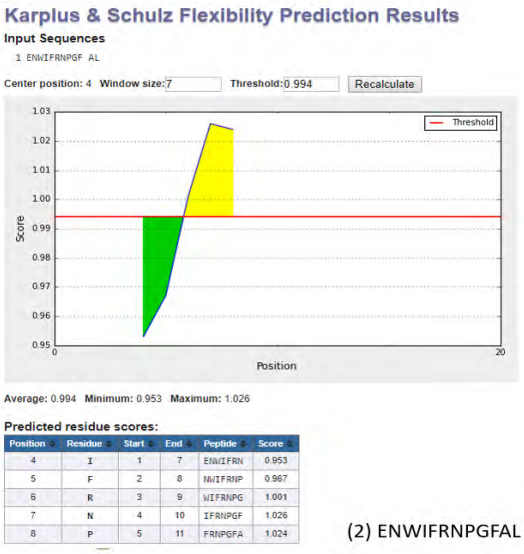
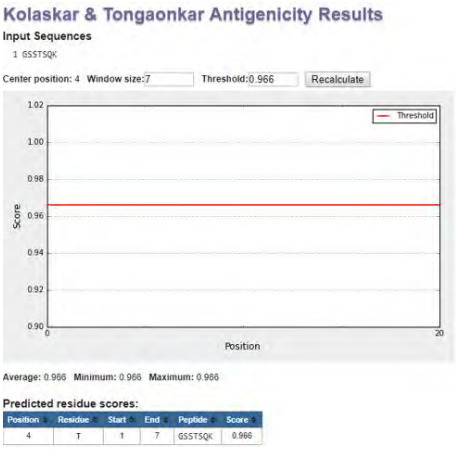


Image 3.17: Flexibility of Epitope 2

3.5.5 Antigenicity Prediction Test:

Finally, a last test for antigenicity test was done using Kolaskar and Tongaonkar Antigenicity Prediction tool to assess the antigenicity of the epitopes. Here, in epitope 1, one residue was found to be antigenic. And in epitope 2, six out of twelve residues were found to be antigenic. Thus, the second epitope shows greater antigenicity.

Antigenicity Prediction Result:



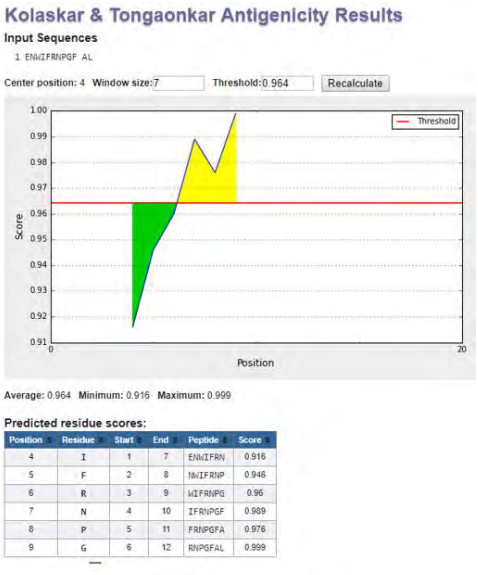


Image 3.18: Antigenicity of the Epitopes

Final Remarks on IEDB tests:

Considering all these facts, the second epitope appears to be more compatible than the first one for vaccine design.

Discussion & Conclusion:

Discussion

In the conducted study the aim was to provide a complete insight about Zika virus because this virus needs to be studied more since there is no vaccine against Zika virus so far. To ensure the accuracy of this investigation similar tests using many softwares, different annotation tools, various vaccine designing tools were done. The comparative analysis covers almost all possible functions and involvements of Zika Virus with its host. However, more studies need to be done to identify the precise activities and mechanism of infection by Zika Virus. Many studies have been conducted to pin down the actual cellular interactions of the virus.

In case of identifying ideal epitopes for vaccine design, the results suggest that epitope ENWIFRNPGFAL is more compatible than epitope GSSTSQK because it gave better results in almost every test. The epitope ENWIFRNPGFAL has shown a desirable amount of antigenicity, surface accessibility, hydrophobicity, flexibility and conservancy. However, more studies are required to materialize a peptide vaccine against Zika Virus. The type, the route of delivery, adjuvants etc must be taken into account.

Here, a comparison of bioinformatic studies on Zika virus and some studies directing to Zika virus cellular interactions are discussed. Vector control and awareness were the practical ways of getting the world risk free from Zika. While studying the prevention of Zika virus vector control is very important. Many drug approaches are being tested against Zika virus such as antiviral therapy, useful mutations, vaccine studies etc. To develop a successful antidote for Zika virus all these things need to be shed light on.

4.1 Comparison of Bioinformatic Study on ZIKA virus

Bioinformatics studies on ZIKA virus has become very popular in recent years. This study was inspired by some of such studies. In this study B Cell epitopes for Zika virus vaccine were identified. B Cell studies on Zika virus was found to be limited in number. But prediction of T cell epitopes in ZIKV polyprotein was done in many studies. In one study they found 23 epitopes of MHC I and 48 epitopes for MHC II molecules (Hamza, 2016). Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus was done in another different study. Here, the immune-informatics approach was used for the screening of potential major histocompatibility complex class I restricted epitopes, which may be competent to generate a cell-mediated immune response in humans. A total of 63 epitopes were identified, which revealed a comprehensive binding affinity to the 42 different human leukocyte antigens class I supertypes (Manas, 2016). Another study of identifying anti ZIKA phytochemicals was

done using molecular docking. Here virus protease, methyltransferase, RNA dependent RNA pol was homology modelled and docked against an in-house library of phytochemicals. Finally, 43 chemicals had drug like effect on the targets (Kendall, 2016). Again, RNA data pipeline for ZIKA virus research (Wang, 2016), complete databases with information of ZIKA genomics, proteomics, therapeutic information have been constructed (Gupta, 2016) to assist such research. One such database in ZikaVR. That has a repository of relevant data of Zika virus (Gupta, 2016)

4.2 Cellular Interactions' Studies on Zika Virus:

Zika virus has a complete set of interactions with its host cells. It is important to shed light on these interactions because they provide important information for drug design. There are many experimental proofs that Zika virus enters cells by interacting with many cell surface receptors, upregulates or downregulates many genes and finally cause a disease by a complex network of interactions. The GO Prediction, InterProScan and ScanProSite analysis in this study already summarizes a lot of these information in a generalized way. Still there are certain interactions that are more specific and provide insight for a preventive measure. Such few interactions are described below:

- In a study where the disruption of molecular fingerprints in human neutrospheres were being tested, it was seen combining proteomics and mRNA transcriptional profiling that, over 500 proteins and genes associated with the Brazilian ZIKV infection were differentially expressed in these cells (Garcez, 2017). These genes and proteins provide an interactome map, which indicates that ZIKV controls the expression of RNA processing bodies, miRNA biogenesis and splicing factors required for self-replication (Garcez, 2017). This experiment also showed ZIKV alters cell cycle and triggers caspase-mediated cell death in iPS-derived neural progenitors.
- Again, RNA data from microcephalic embryonic mice, human fibroblasts and neural progenitors showed deregulation of many individual genes related to viral response (Garcez, 2017)
- Neurotropic ZIKV could enter the CNS and cause disease through many surface proteins, such as the candidate AXL receptor, which facilitates ZIKV to invade the developing CNS (Zhou, 2017)

- ZIKV infection elicits the deregulation of genes related to neurogenesis like Nestin and Sox2, differentiation, and immune response-associated factors including TLR3, which has been validated in several independent in vivo and in vitro studies (Zhou, 2017)
- ZIKV also dultiple key cellular signaling cascades, including the PI3K-Akt-mTOR pathway. The nonstructural proteins NS4A and NS4B of ZIKV does this. This pathway is critical for neurogenesis from NPCs, as well as for subsequent migration and maturation, and autophagy regulation in brain development (Zhou, 2017).
- In a different study it was shown that an American strain from an infected fetal brain (FB-GWUH-2016) and a closely-related Asian strain (H/PF/2013), productively infect human iPSC-derived brain organoids. Both strains readily target to and replicate in proliferating ventricular zone (VZ) apical progenitors. The main phenotypic effect was premature differentiation of neural progenitors associated with centrosome perturbation, even during early stages of infection, leading to progenitor depletion, disruption of the VZ, impaired neurogenesis, and cortical thinning (Gabrial, 2017)
- It was found in an experiment that Zika virus antagonizes type 1 interferon response during infection of human dendritic cells (21, 26). Human DCs supported productive infection by a contemporary Puerto Rican isolate of ZIKA with considerable variability in viral replication. Infection of DCs with both contemporary and historic ZIKV isolates led to minimal up-regulation of T cell co-stimulatory and MHC molecules, along with limited secretion of inflammatory cytokines. Inhibition of type I interferon (IFN) protein translation was observed during ZIKV infection, despite strong induction at the RNA transcript level and up-regulation of other host antiviral proteins (Bowen, 2017)
- Several studies have identified microglial nodules, gliosis, neuronal and glial cells degeneration and necrosis in the brain of ZIKV infected infants, suggesting that ZIKV could play a role in these neurological disorders through neuroinflammation and microglial activation (Tricarico, 2017). A more specific study was done that showed that U87-MG cells are susceptible to ZIKV infection. ZIKV can successfully replicate in infected cells causing oxidative stress, inflammasome activation and subsequent release of mature IL-1β; this process culminates in cell death.
- In another study it was shown that ZIKA virus elicits P53 activation and genotoxic stress in human neural progenitors similar to mutations involved in severe forms of genetic microcephaly and p53 (Ghouzzi, 2016). ZIKV infection increases total P53

levels and nuclear accumulation, as well as P53 Ser15 phosphorylation, correlated with genotoxic stress and apoptosis induction.

4.3 Vector Control to Prevent Zika Virus:

As mosquito is the carrier of Zika virus, and primarily it infects human with this disease and many others, vector control is considered. Vector control strategies and vector-pathogen interaction of all possible mosquito species are advised to be notices owing to ZIKV's ability to evolve (Zhou, 2017)

Because of the lack of vaccines and antiviral therapeutics against ZIKV, initially larvicides are greatly emphasized. To prevent human to human transmission, it is recommended that public health authorities in ZIKV endemic regions to provide access to contraceptives, prenatal care, and safe abortion services.

Entomological surveillance allows for early detection of a potential virus outbreak, vector distribution and density, and evaluation of vector control strategies (Zhou, 2017). A more technical approach would be through the introduction of genetically modified male mosquitoes carrying a dominant lethal gene expressed at the larval stage which causes death in offspring upon mating with wild female mosquitoes This novel technology in known as Gene Drive.

Another potential strategy, which has shown positive potential for DENV control, is using the endosymbiotic relationship between Aedes mosquitoes and the *Wolbachia* bacteria (Zhou, 2017). Such approach could bring good result for Zika virus as well. The endosymbiotic relationship interferes with virus replication in the mosquitoes. The concept is using *Wolbachia* to immunize the mosquitoes against DENV and then setting them loose in the wild, where those mosquitoes will pass the bacterium to their offspring. This may cause some amount of harm to the mosquito but the potential benefits for humans cannot be denied also. If *Wolbachia* infected mosquitoes become predominant in the wild, it is expected dengue disease among people to drop. However, the potential of the virus to evolve and overcome the inhibitory effect of the endosymbiotic relationship must be taken in to consideration also.

4.4 Drug Approaches to Resist Zika Virus:

Because of relentless research all over the world drugs against Zika viruses are being developed. So far, the only FDA approved drug for Zika virus is Sofosbuvir, that was used as a treatment for Hepatitis C. This acts as an inhibitor of HCV NS5. NS5 (viral polymerase) is a

key enzyme for virus to survive. Disruption of this enzyme inhibits viral activity (Bullard-Feibelman, 2017).

NS5 is considered a popular target of drug design. NS5 is a multidomain protein which has N terminal methyltransferase domain and C terminal RNA dependent RNA polymerase domain. The methyltransferase domain (MTase) of ZIKV non-structural protein 5 (NS5) can sequentially methylate guanine N-7 and ribose 2′-O to form m7NGpppA2′Om cap structure in the new RNA transcripts. This methylation step is crucial for ZIKV replication cycle and evading the host immune system. Defect in the MTase activity can be lethal for Flavivirus making it a target for drug design.

Again, drug targets using useful mutations are also being tested. In experiments it has been shown that certain mutations can disrupt Zika virus proliferation. Such as, Thr233, a unique residue found in ZIKV but not in other flaviviruses, organizes a central hydrogen bonding network at NS1 dimer interface. Mutation of Thr233 to Ala disrupts this elaborated interaction network and destabilizes the NS1 dimeric assembly in vitro (Wang 2017). Again, it is shown in an experiment that a single mutation in the envelope protein modulates flavivirus antigenicity, stability, and pathogenesis (Goo, 2017).

To date, no vaccine against ZIKV has entered the clinical stage. Recently, an Indian biotech company claimed that it has two ZIKV vaccine candidates awaiting pre-clinical trials (Sharma, 2017). SynCon Pharmaceuticals (USA) has also developed a DNA-based vaccine against ZIKV, which is expected to enter clinical trials (Sharma, 2017). Since low genetic variation is observed in different ZIKV strains, it is likely that a single vaccine may be effective against all circulating ZIKV strains. However, the effect of pre-existing immunity against other Flaviviruses on the immunity against ZIKV must be further investigated (Sharma, 2017)

As a consideration, drug that inhibit a certain stage of virus life cycle is often targeted. A few drugs that have already shown in vitro inhibition in case of DENV include mefenamic acid, tetracyclines, amodiaquine and chloroquine (Sharma, 2017). These drugs can be tested for Zika as well because Zika virus and DENV are so closely related.

4.5 Conclusion

All these information from all different studies should provide a versatile idea about Zika virus and its interactions. Moreover, it must be noted that Flavivirus vaccine development is limited by the nature of outbreaks, being sporadic and unpredictable making rapid vaccine production a challenge. Therefore, vast studies should be done to develop a vaccine against Zika Virus to counteract the next epidemic.

References:

References:

- Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 6.3, Viruses: Structure, Function, and Uses.
- Yadav, S., Rawal, G., & Baxi, M. (2016). Zika virus: A pandemic in progress. Journal of Translational Internal Medicine, 4(1), 42–45. http://doi.org/10.1515/jtim-2016-0009
- Sharma, A., & Lal, S. K. (2017). Zika Virus: Transmission, Detection, Control, and Prevention. Frontiers in Microbiology, 8, 110. http://doi.org/10.3389/fmicb.2017.00110
- Rawal, G., Yadav, S., & Kumar, R. (2016). Zika virus: An overview. Journal of Family Medicine and Primary Care, 5(3), 523–527. http://doi.org/10.4103/2249-4863.197256
- 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, 137, 134–140. http://doi.org/10.1016/j.antiviral.2016.11.023
- Catherine, S. (2018, Feb 13). *The Baltimore Classification System.* Retrieved From https://www.news-medical.net/life-sciences/The-Baltimore-Classification-System.aspx
- Hulo, C., Masson, P., de Castro, E., Auchincloss, A. H., Foulger, R., Poux, S., … Le Mercier, P. (2017). The ins and outs of eukaryotic viruses: Knowledge base and ontology of a viral infection. PLoS ONE, 12(2), e0171746. http://doi.org/10.1371/journal.pone.0171746
- Zhou, K., Wang, L., Yu, D., Huang, H., Ji, H., & Mo, X. (2017). Molecular and cellular insights into Zika virus-related neuropathies. Journal of Neurovirology, 23(3), 341–346. http://doi.org/10.1007/s13365-017-0514-3
- Wang, D., Chen, C., Liu, S., Zhou, H., Yang, K., Zhao, Q., … Yang, H. (2017). A Mutation Identified in Neonatal Microcephaly Destabilizes Zika Virus NS1 Assembly in Vitro. Scientific Reports, 7, 42580. http://doi.org/10.1038/srep42580
- Lazear, H. M., & Diamond, M. S. (2016). Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. Journal of Virology, 90(10), 4864–4875. http://doi.org/10.1128/JVI.00252-16
- Shragai, T. , Tesla, B. , Murdock, C. and Harrington, L. C. (2017), Zika and chikungunya: mosquito-borne viruses in a changing world. Ann. N.Y. Acad. Sci., 1399: 61-77. doi:10.1111/nyas.13306

- Cavalcanti M.G., Cabral-Castro M.J., Goncalves J.L.S., Santana L.S., Pimenta E.S., Peralta J.M. (2017), Zika virus shedding in human milk during lactation: an unlikely source of infection? International Journal of Infectious Diseases, 57 , pp. 70-72.
- Garcez, P. P., Nascimento, J. M., de Vasconcelos, J. M., Madeiro da Costa, R., Delvecchio, R., Trindade, P., … Rehen, S. K. (2017). Zika virus disrupts molecular fingerprinting of human neurospheres. Scientific Reports, 7, 40780. http://doi.org/10.1038/srep40780
- Faizan M, I, Abdullah M, Ali S, Naqvi I, H, Ahmed A, Parveen S, Zika Virus-Induced Microcephaly and Its Possible Molecular Mechanism. Intervirology 2016;59:152-158
- Hamza Dar, Tahreem Zaheer, Muhammad Talha Rehman, Amjad Ali, Aneela Javed, Gohar Ayub Khan, Mustafeez Mujtaba Babar, Yasir Waheed,Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach, Asian Pacific Journal of Tropical Medicine, Volume 9, Issue 9, 2016, Pages 844-850, ISSN 1995-7645, https://doi.org/10.1016/j.apjtm.2016.07.004.
- Manas R. Dikhit, Md. Yousuf Ansari, Vijaymahantesh, Kalyani, Rani Mansuri, Bikash R. Sahoo, Budheswar Dehury, Ajay Amit, Roshan K. Topno, Ganesh C. Sahoo, Vahab Ali, Sanjiva Bimal, Pradeep Das, Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus: A first step towards vaccine development, Infection, Genetics and Evolution, Volume 45, 2016, Pages 187-197, ISSN 1567-1348, https://doi.org/10.1016/j.meegid.2016.08.037.
- Kendall G. Byler, Ifedayo Victor Ogungbe, William N. Setzer, In-silico screening for anti-Zika virus phytochemicals, Journal of Molecular Graphics and Modelling, Volume 69, 2016, Pages 78-91, ISSN 1093-3263, https://doi.org/10.1016/j.jmgm.2016.08.011.
- 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, 5, 1574. http://doi.org/10.12688/f1000research.9110.1
- Gupta, A. K., Kaur, K., Rajput, A., Dhanda, S. K., Sehgal, M., Khan, M. S., … Kumar, M. (2016). ZikaVR: An Integrated Zika Virus Resource for Genomics, Proteomics, Phylogenetic and Therapeutic Analysis. Scientific Reports, 6, 32713. http://doi.org/10.1038/srep32713

- Goo, L., VanBlargan, L. A., Dowd, K. A., Diamond, M. S., & Pierson, T. C. (2017). A single mutation in the envelope protein modulates flavivirus antigenicity, stability, and pathogenesis. PLoS Pathogens, 13(2), e1006178. http://doi.org/10.1371/journal.ppat.1006178
- Zhou, H., Wang, F., Wang, H., Chen, C., Zhang, T., Han, X., … Yang, H. (2017). The conformational changes of Zika virus methyltransferase upon converting SAM to SAH. Oncotarget, 8(9), 14830–14834. http://doi.org/10.18632/oncotarget.14780
- Gabriel, Elke; Ramani, Anand; Karow, Ulrike; Gottardo, Marco; Natarajan, Karthick; Gooi, Li Ming; Goranci-Buzhala, Gladiola; Krut, Oleg; Peters, Franziska; Nikolic, Milos; Kuivanen, Suvi; Korhonen, Essi; Smura, Teemu; Vapalahti, Olli; Papantonis, Argyris; Schmidt-Chanasit, Jonas; Riparbelli, Maria; Callaini, Giuliano; Krönke, Martin; Utermöhlen, Olaf; Gopalakrishnan, Jay. (2017). Recent Zika Virus Isolates Induce Premature Differentiation of Neural Progenitors in Human Brain Organoids, Cell Stem Cell, 20(3), PMID: 28132835, DOI: 10.1016/j.stem.2016.12.005.
- Bowen, J. R., Quicke, K. M., Maddur, M. S., O'Neal, J. T., McDonald, C. E., Fedorova, N. B., … Suthar, M. S. (2017). Zika Virus Antagonizes Type I Interferon Responses during Infection of Human Dendritic Cells. PLoS Pathogens, 13(2), e1006164. http://doi.org/10.1371/journal.ppat.1006164
- Paola Maura Tricarico, Ilaria Caracciolo, Sergio Crovella, Pierlanfranco D'Agaro, Zika virus induces inflammasome activation in the glial cell line U87-MG, Biochemical and Biophysical Research Communications, Volume 492, Issue 4, 2017, Pages 597-602, ISSN 0006-291X, https://doi.org/10.1016/j.bbrc.2017.01.158.
- Ghouzzi, V. E., Bianchi, F. T., Molineris, I., Mounce, B. C., Berto, G. E., Rak, M., … Di Cunto, F. (2016). ZIKA virus elicits P53 activation and genotoxic stress in human neural progenitors similar to mutations involved in severe forms of genetic microcephaly and p53. Cell Death & Disease, 7(10), e2440–. http://doi.org/10.1038/cddis.2016.266
- Notredame C, Higgins DG, Heringa J (2000-09-08). "T-Coffee: A novel method for fast and accurate multiple sequence alignment". J Mol Biol. 302 (1): 205–217. doi:10.1006/jmbi.2000.4042. PMID 10964570.
- Papadopoulos JS and Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics 23:1073-79.

- Sudhir Kumar, Glen Stecher, and Koichiro Tamura (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. Molecular Biology and Evolution
- Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F. Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics, Jan 2014; doi:10.1093/bioinformatics/btu031
- De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W362-5
- Irini A Doytchinova and Darren R Flower. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine. 2007 25:856-866.
- Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 2017 (Web Server issue). doi: 10.1093/nar/gkx352
- Chen J, Liu H, Yang J, Chou K (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33: 423-428.
- Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. Immunome Res 2:2.