In Silico Epitope based Peptide Vaccine Designing against Yellow Fever Virus

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

Submitted by: HASANUL KARIM RAFFI

Student ID: 13136001

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

August 2017

DECLARATION

I hereby solemnly declare that the research work embodying the analysis and results reported in the following thesis entitled "*In Silico* Epitope based Peptide Vaccine Designing against Yellow Fever Virus", submitted by the undersigned has been carried out under the supervision of Ms. ShamiraTabrejee, Lecturer, BSc. in Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University, 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.

_____

(Hasanul Karim Raffi)

Candidate

Certified

_____

(Ms. ShamiraTabrejee)

Supervisor

Lecturer

BSc. in Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University, Dhaka

ACKNOWLEDGEMENTS

Above all, I wish to express my most sincere and utmost gratitude to the Almighty, to have blessed me with His mercy, and providing me with the strength, perseverance and understanding, needed for the successful completion of this thesis project.

My deepest appreciation and special thanks to **Professor A.F.M Yusuf Haider,** the 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 exemplary guidance, endless provision and support, during my tenure as a student in BRAC University. I would also thank the authority of BRAC University for providing me with many opportunities and facilities.

I am truly indebted to my supervisor, Lecturer **Ms. ShamiraTabrejee**, Department of Mathematics and Natural Sciences, BRAC University, for believing in me and allowing me to work on what I desired. She has my earnest appreciation. Throughout the journey of this research, Ms. Shamira has constantly encouraged me to explore new ideas and bring out the best from this project. Her helpful advice has refined and enhanced my capability and the success of this project immensely.

I am grateful to all the faculty members of the Department of Mathematics and Natural Sciences for their unwavering support and guidance throughout the entire period of my bachelor's degree.

Finally, I would like thank my parents for their sheer devotion to my education, future and happiness. They offered me their undying support, encouragement along with their patience. They are indeed, the most precious blessings in my life.

I perceive this opportunity as a big milestone in my career development. I will strive to use the gained skills and knowledge in the best way possible, and I will continue to work on their improvement, in order to attain my desired career objectives.

Sincerely,

Hasanul KarimRaffi

Abstract

The most widely used vaccine for treatment of Yellow Fever patients is YF 17D. It is a live attenuated vaccine derived from Asibi strain. It is administered widely to patients throughout the world but it has few problems such as low immune response and it may cause allergic reactions. The aim of this study is todesign an epitope-based peptide vaccine by targeting E protein of Yellow Fever Virus which may induce a stronger immune response.30 sequences of E protein of Yellow Fever Virus strainswere retrieved from NCBI database. E protein was found to be mostly conserved among all the sequences with little variability. Our conserved E region was found to be a probable antigen with a value of 0.4588 in Vaxijen sever.4 epitopes were found to be common in BepiPred and BCPREDS. Three of those epitopes were found to be antigenic. A peptide VKNPTDTGHGT were predicted to have surface accessibility. So the whole epitope VKNPTDTGHGTwas taken for analyzing conservancy and was found to be 96.67% conserved in all sequences. VKNPTDTGHGTepitope also possesses flexibility and accessibility as most of the residues of the peptide were found above the threshold level. Here we suggest in vivo study of our novel peptide antigen in E protein for universal vaccine which may be used to prevent Yellow Fever virus.

TABLE OF CONTENTS

List of Figures and Tables:

List of Abbreviation and Symbols:

YFV	=	Yellow fever virus
APC	=	Antigen presenting cells
NS	=	Non-structural
IL	=	Interleukin
DNA	=	Deoxyribonucleic acid
E	=	Envelope protein
WHO	=	World Health Organization
MHC	=	Major histocompatibility complex
HLA	=	Human leukocyte antigen
HIV	=	Human Immunodeficiency virus
C	=	Capsid protein

Chapter 1: Introduction

Chapter 1: Introduction

1.1 Yellow Fever virus

Yellow fever virus belongs to the family Flaviviridae, genus Flavivirus. It is a arthropod-borne virus. They can be transmitted by vectors such as ticks or mosquitoes. It gets transmitted between humans through Aedesaegypti mosquito. Yellow fever is an acute disease which causes hemorrhagic fever and jaundice. The symptoms of yellow fever includes kidney failure, high fever and internal bleeding. There are 200,000 cases reported annually, which includes 30,000 deaths.(Barros et al., 2011) The YFV genome is single stranded, positive sense RNA, which encodes for a polyprotein. The polyprotein later undergoes post and co translational processing into three structural proteins and seven non-structural proteins. The genome encodes for 370kDa polyprotein precursor, which undergoes further processing in the endoplasmic reticulum and is broken down into three structural proteins (C, M and E) and non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) required mainly for replication.(Volk et al., 2009) The E protein is the most studied one due to its high antigenicity. It is crucial for various events such a binding site for viral attachment, penetration, fusion, host range and cell tropism.(Wu, Bera, Kuhn, Smith, & Smith, 2005)Each E protein monomer has a molecular weight of 50-55 kDA and consists of three distinct domains, I, II, and III. Domain III is a receptor binding domain recognized by neutralizing antibodies. (Volk et al., 2009) E protein consists of a dimer in which each monomer has three β barrel domains. It consists of a cellular binding site, fusion peptide with two transmembrane helices. A fusion peptide of one E monomer is submerged in between domain I and III of the adjacent monomer of a dimer. Capsid protein is 120 amino acids which is involved in the packaging of the viral genome as well as formation of nucleocapsid core.(“mukhopadhyay2005 (1),” n.d.)

1.2 Yellow Fever Vaccines:

Late 1700s YFV was seen and early 1900s it became prevalent. Originally there were two strains that became prevalent. Asibi strain from Ghana was the first one to be discovered. Live attenuated vaccines were formed by the attenuation of Asibi strain for the 146^{th} time. This was known as 17D. There was another live attenuated vaccine from the French strain known as FVV. 17D was less toxic than FVV although it elicit a weaker immune response. 17D vaccine could be used as the backbone for making several recombinant vaccines such as ChimeVax.(Beck & Barrett, 2015) It was a prime discovery and is still grown in chicken embryonic cell culture method due to its uses. The mechanism by which attenuation works is rudimentary and further in silico work needs to be done on 17D to completely understand attenuation. All RNA viruses have a high diversity which makes it difficult to administer a single vaccine for all flavivirus. The low diversity of 17D is considered for its high attenuation and safety. WHO, Global alliance for vaccines and Immunization Vaccine alliance have targeted to immunize 250 million people with YFV in Africa. (Beck & Barrett, 2015)Epitope based vaccine for YF can be used as a better alternative as it has high immunogenicity, hence may elicit a faster immune response. It is safer to use and can be easily stored. In addition, it can be readily produced at a cheaper price than the conventional YF vaccines.

1.3 Epitope based Vaccine:

Active peptide based vaccine will increase patient quality of life by saving time and money. The next generation peptide vaccines include multivalent longer peptides with class I and II MHC epitopes, multi peptide vaccines consisting of class I and II epitopes, a peptide cocktail and hybrid peptides. T-cell vaccines have potential for the development of therapeutic vaccines for chronic viral infection. Epitope vaccines with adjuvants induce a stronger immune response with high immunogenicity. (Kametani, Miyamoto, Tsuda, & Tokuda, 2015)MHC class I epitopes have to be antigen specific to the virus. Analysis of CTL repertoire at the clonal level in human peripheral blood nucleated cells reveal a very broad CTL response. The technical challenge is to

identify a few epitopes among a mixture of >10,000 MHC class I associated epitopes extracted from virus infected cells. Identification of the peptides by cell infected with viral pathogens can generate a library of epitopes. A genome database searching the viral protein origins which can be recognized and processed by the immune system.(Lambert & Fowler, 2005)

1.4 Advantages of epitope based vaccine:

Epitope based vaccine formulated with an adjuvant can induce a strong immune response with high immunogenicity. It will offer a flexible and simple way to synthesize a vaccine. They are considered to be safe, easy to produce and stable. Several peptide vaccines are in development due to their successful stories in animal studies. It is feasible that within the next 5-10 years several therapeutic agents will be on the market. Several therapeutic agent for infectious diseases such as HBV, HCV, HIV and HPV are being clinically tested.(Lambert & Fowler, 2005) Peptides have become ideal candidates for vaccines due to their easier production, chemical stability and lack of pathogenic potential. T-cell epitopes are usually peptide fragments whereas B-cell epitopes can be either protein, lipids or carbohydrates. Production of such immunodominant epitopes can elicit a strong specific immune response with less adverse effects.(Patronov & Doytchinova, 2013)

1.5 Specific aims and objectives of the project:

- To study about the Yellow Fever and its consequences.
- To study about the current prospects of Yellow Fever vaccines
- Search for a conserved region among the Yellow Fever strains.
- Analyze the conserved region to check for antigenicity, protein variability and conservancy.
- Develop an epitope based vaccine against Yellow Fever virus.

Chapter 2: Literature Review

Chapter 2: Literature Review

2.1 Background:

The Yellow fever virus is a member of Flavivirus genus and Flaviviridae family. Its genome consists of a single positive-stranded RNA molecule with an approximate weight of 11kb. It consists of several structural proteins such as capsid, membrane with its precursor, envelope. Additionally it also has several non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. (Volk et al., 2009)Each of them possesses a unique role for viral replication. An YF 17D vaccine was proposed initially and a single dose provide 90% immunity for more than 30 years. It triggers a polyvalent immune response due to its ability to activate different types of dendritic cells, via toll-like receptors (TLRS), which in turn results in the production of a set of cytokines such as interferon alpha, IL-6 etc. Hence this leads to the activation of adaptive immune response. Adaptive immune response is characterized by the clonal expansion of specific activated T $CD8^+$ cells as well as helper T-cells (Th1 & Th2) via cytokine mediated stimulation. (Nogueira et al., 2011)

Attenuated Yellow Fever vaccines 17D/ 17DD are the only vaccines available to prevent the yellow fever infection. Both of them can elicit T-cell responses and neutralizing antibodies. Some toxic side effects have been reported at some places. In order to prevent them, DNA vaccines have been proposed. The tests check whether it can produce similar titer of anti-YF antibodies as the conventional ones.(Keenan & Jaffee, 2012) A study was done in order to check the efficiency of two non-viral DNA antigen based formulations and their immunological properties. The two antigen formulations consisted of DNA encoding the full length envelope protein, another one with full length envelope protein fused with lysosomal associated membrane signal (LAMP-1). The results show that LAMP-1 can elicit the same T-cell responses as the YF-17 DD. It showed 100% protection against YF virus in intracerebrally challenged mice. Hence the results show that DNA vaccine can be an excellent candidate as a replacement of conventional vaccines used for Yellow fever.(Barros et al., 2011)

Among the viral proteins, the E protein is the most targeted one due to its high antigenic potential. It is involved in various purposes such as receptor binding site for viral attachment, fusion, penetration, haemagglutination, host range and cell tropism. It also plays a vital role in

eliciting neutralizing antibodies and immune response. Native E protein presents itself as a homodimer but goes through various conformational during receptor mediated endocytosis. The conformational change occurs in the low pH environment of the endosome during the fusion of viral envelope with the endosomal membrane.

2.2What is a Vaccine?

The term vaccination is derived from the Latin word vacca which means cow. This is because the original procedure involved the inoculation of contents from cowpox lesions into healthy people. This procedure was first tried by Edward Jenner in 1796. A vaccine contains components that is required to elicit an immune response which maybe humoral or cellular immunity. Both of adaptive immunity can protect a patient against a virus infection. (Kametani et al., 2015)The purpose of most viral vaccines is stimulate a long-term immunity against a virus by creating an immunological memory which will be triggered when the virus invades the patient. In order to establish a strong immunological memory the vaccine needs to induce strong immune responses. Effective vaccines are used to protect against diseases such as polio, rubella, rabies and foot and mouth viruses. Whereas effective vaccines are yet to be developed for viruses like HIV-1, hepatitis C, Ebola and herpes simplex virus.(Lambert & Fowler, 2005) Those involved in virus vaccine research face many difficulties such as multiple antigenic variants of target viruses and also high standards of safety and storage. A huge need for new vaccines has produced novel categories of vaccine such as peptide vaccines and DNA vaccines. (Beck & Barrett, 2015)

2.3Types of vaccine:

2.3.1Inactivated vaccine:

A type of vaccine that is grown in cell culture and is inactivated by means of high temperature or pressure. The encapsulated evokes an immune response which can be humoral or cellular immunity. Lack of inactivation can be fatal for the patients. The trick is finding the perfect combination of chemical concentration and reaction time which completely inactivates the virus. The virus must still expose the adequate antigens to stimulate a protective immune response. Jonas Salk created a treatment for poliovirus in which virions were suspended in formalin at 37 degrees for about 10 days. It is vital to understand the kinetics of a virus inactivation to develop an inactivation procedure which guarantees 100% inactivation. The examples of inactivated vaccines are influenza, hepatitis A, poliovirus and foot and mouth disease viruses. (Beck & Barrett, 2015)

2.3.2Live attenuated vaccines:

The attenuated vaccines contain live particles with low levels of virulence. Therefore they can replicate slowly and continuously providing a supply of antigens required to elicit an immune response. They are produced by passing them in animal cell cultures and by selection of less virulent strains or mutagenesis. There are two properties that a vaccine virus must possess. Their antigens must be identical or similar to the wild-type strain so that an immune response against vaccine virus provides protection against the wild type strain. Most of the attenuated virus strains are produced by 'hit and miss' procedures in which they are cultured repeatedly in cells which are unrelated to the normal host. The attenuated vaccines developed by this method are measles, mumps, yellow fever, rubella, canine parnovirus and canine distemper. One of the risks associated with attenuated vaccines is during virus replication nucleotide substitution may take place. This may reverse attenuated virus into virulence form and can be fatal for patients. (Beck & Barrett, 2015)

2.3.3Subunit vaccines:

These are the vaccines that only contain the antigenic part of the virus instead of the whole microorganism. Although it evokes a weaker immune response than conventional vaccines. It causes less allergic reactions and has a size range of 1-20 amino acids. Example of subunit vaccine is Hepatitis B vaccine. The infectivity of a virus is inactivated with formaldehyde and then the virion envelopes are removed using a detergent called Triton X-100. This causes the release of glycoproteins which aggregates to form H 'cartwheels' and N 'rosettes'. These structures are later purified by centrifugation in a sucrose gradient. The subunit vaccines induces poor immune responses hence two doses are necessary to initiate adequate immunity. (Patronov & Doytchinova, 2013)

2.3.4DNA vaccines:

A new approach which uses genetically engineered DNA to induce humoral and cellular immune responses to protein antigens. It is on a test phase for viral, fungal and bacterial models. The cancer DNA vaccines specific immunity against cancer associated antigens. The antigen coding sequence from a DNA virus is inserted into a plasmid which resides in between a strong promoter and poly (A) signal. The plasmid is replicated in bacterial cells and is then extracted, purified to use as a vaccine. The vaccine can be injected into a muscle or by using a gene gun which delivers with DNA-coated gold beads directly into our skin. The examples of experimental DNA vaccines include HIV-1, SARS coronavirus, West Nile virus and foot and mouth viruses. Before any DNA vaccine moves for clinical use we must ensure that injection of DNA will not trigger an anti-DNA autoimmune disease and it will not cause cancer-causing mutations inside the patients. (Beck & Barrett, 2015)

2.3.5Peptide based vaccines:

They have been developed due to the enhanced understanding of the molecular structures of antigens. It can be either T-cell epitope or B-cell epitope. T-cell epitopes are usually peptide fragments whereas B-cell epitope can be lipids, proteins or carbohydrates. There are several reasons why epitope based vaccines are preferred over traditional vaccines such as easy production, chemical stability and lack of infectious materials. Although it requires an adjuvant and has lower immunogenicity. These are areas that needs to be worked on but it is a bright prospect on the evolution of vaccines. Jackob et al was the first person to produce epitope-based vaccine.(Kametani et al., 2015)

2.4.1Classification of Flavivirus:

Flavivirus comprises of about 70 viruses. DEN, JE, SLE and YF are human pathogens. It affects nearly 80 million people all over the world per year. At the beginning flavivirus were serologically classified into eight groups but YF could not be associated into any groups. After the serological classification many viruses have been discovered which have not been classified into any groups yet. Classification is difficult due to geographical distribution, diversity of the arthropod vectors or vertebrate hosts which are involved in the transmission of viruses. Hence a better classification is required.Molecular genetic classification has been performed and one-third of the viruses were used to create phylogenetic trees. For comprehensive phylogeny of the genus Flavivirus 1.0 kb sequence from 3' terminus of NS5 gene was taken for study. Quantitative criteria based on boot strap level and the pairwise nucleotide sequence identity was used to created phylogenetic tree. It includes clade, cluster and species.(Kuno et al., 1998)

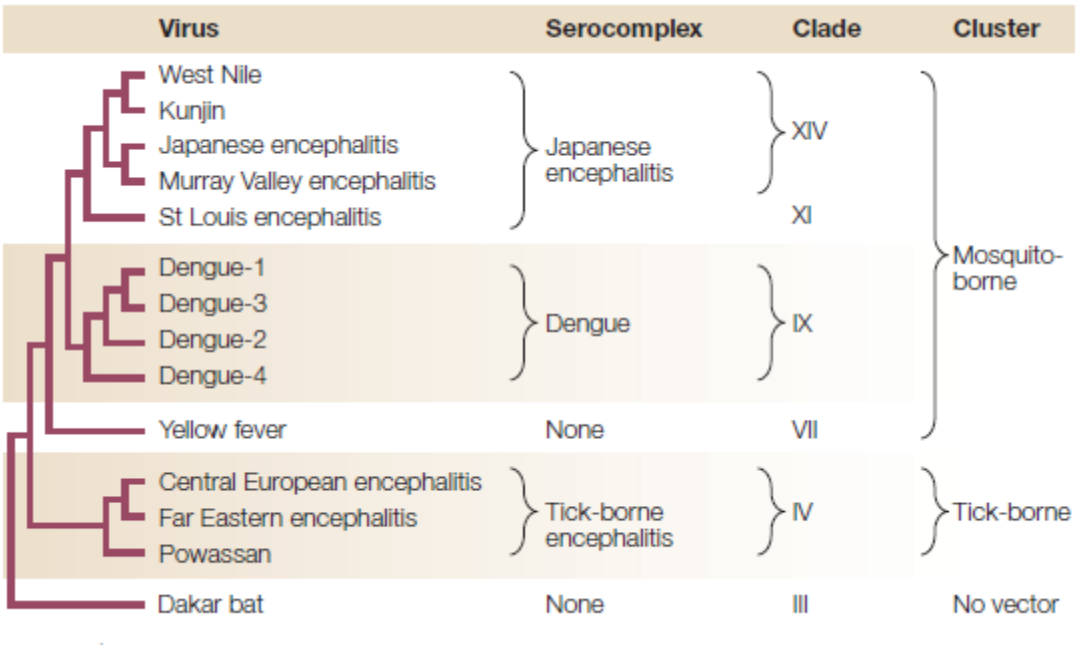

Figure 2.3: Classification of Flavivirus

2.4.2 What is Major Histocompatibility complex(MHC)?

T-cell epitopes are presented by the MHC molecules of antigen presenting cell. There are two types of classes MHC-I and MHC-II. The MHC-I presents peptides of 8-11 amino acids whereas MHC-II present 11-25 amino acids. MHC class II is resented by specialized cell types such as B cells, macrophages and dendritic cells. MHC class I is presented by all nucleated cell bodies. MHC molecules are the most polymorphic proteins which contains more than 6000 classes. It is beyond the capacity with lab experiments to determine all the peptide binding preferences of alleles. Therefore in silico methods are used to reduce the numbers into a feasible amount for lab experiments. T-cell epitope prediction algorithms enables us to predict specific epitopes which could be later produced experimentally. There are various models used to predict them such as

Hidden Markov Model, Motif-based model, Support-vector machines, structure based approaches etc.

2.4.3 Structure of E protein:

E protein consists of a dimer in which a monomer contains three β-barrel domains. The central domain (I) constitutes the N-terminal and is flanked by elongated dimerization domain (II) on one side with fusion peptide at its distal end. The other side of domain I contains the domain III. Domain III is an immunoglobulin like domain that contains the receptor binding sites. Domain I and II are connected by four polypeptide linkers whereas domain I and III are connected by a single polypeptide linker. A fusion peptide of one E monomer is submerged in between domain I and III of the adjacent monomer within a dimer.("mukhopadhyay2005 (1)," n.d.)

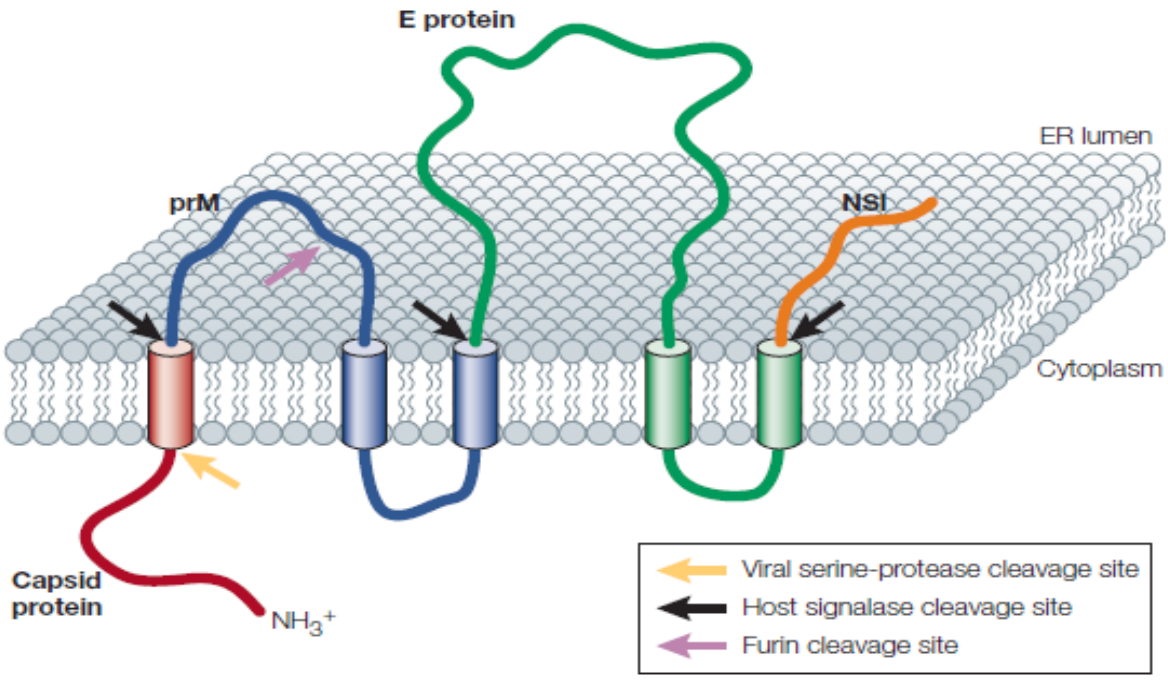

Figure 2.4: Structural proteins of Flavivirus

2.4.4 Entry of E protein:

A 40% amino acid is found among all the flaviviruses for the E protein. The number and positions of the glycosylated residues tend to change among different strains of the same virus. The carbohydrate residues residing on the surface of the virions are responsible for the specific receptor binding. The glycosylated amino acids is spatially on top the E dimer structure near domain III. The intricate structural modifications indicates that surface resides are responsible for specific binding to cellular receptors. Heparin and glycosaminoglycans are low-affinity co-receptors. For YFV domain III contains RGD/RGE sequence which is used as a recognition motif for integrin binding.(“mukhopadhyay2005 (1),” n.d.)

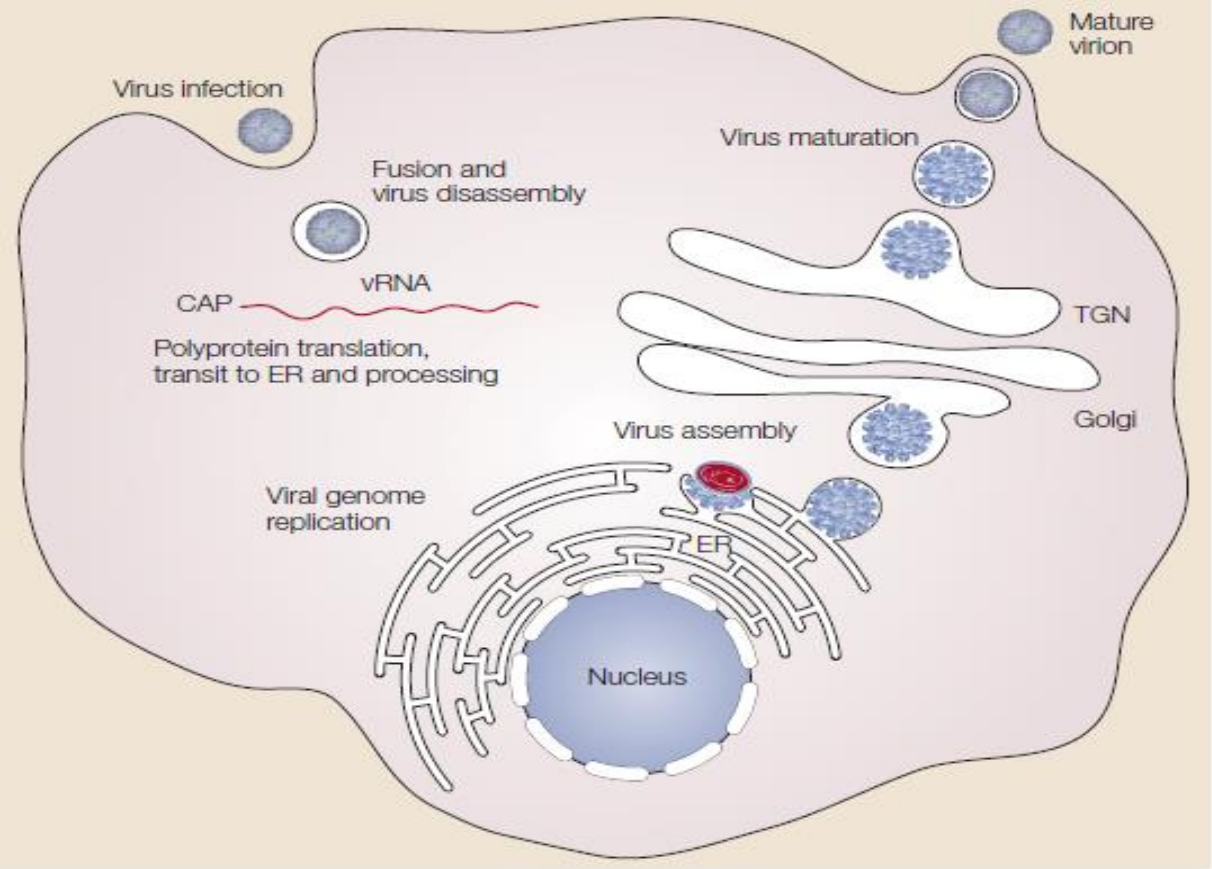

Figure 2.5: Flavivirus infection cycle

2.4.5 Fusion of the E protein:

The E protein usually has two conserved helices in the stem region which resides in between ectodomain and transmembrane region. Acidification causes the dimers to dissociate into monomers that again recombines irreversibly into forming trimers. The X-ray crystallography has shown that acidification causes major conformational changes. The E protein shows trimeric arrangements in which their long dimension stay parallel to each other. Whereas the fusion peptide resides at the tip of the trimer. The domain III folds back 30A and rotates 20 degrees towards the domain II. The C terminus of ectodomain was seen to be position at the end of a hydrophobic groove. The stem helices also move into the groove which brings the transmembrane region in close proximity to host cell membrane.(Wu et al., 2005)

2.4.6 Capsid protein and assembly:

The first events of flavivirus is the formation of NC. It consists of one copy of genomic RNA and multiple copies of capsid protein. NCs are rarely found in the flavivirus which indicates that the particle formation is created in a coordinated process by membrane associated capsid protein and prM-E heterodimers in the ER. Capsid proteins are composed of a conserved hydrophobic region with helices of α2 and α3. It has been observed that deletions in the hydrophobic region of α1 and α2 did not stop the formation of virus particles. This indicates the residues in the hydrophobic region are not compulsory virus assembly. The dimerization is solely maintained by α4 helix.(“mukhopadhyay2005 (1),” n.d.)

2.4.7 Rearrangements during maturation and fusion:

The virus structure undergoes major conformational changes during the viral cycle. This can be inferred by the structural information on mature and immature virions as well as pre-fusion dimer and post-fusion dimer. The maturation occurs in a two-step process. A pH induced conformational change in prM and E occurs in the trans-Golgi network. It is followed by the

cleavage of prM by furin. During the maturation phase 60 trimers of prM0E heterodimers dissociate into forming E homodimers.(“mukhopadhyay2005 (1),” n.d.)

E protein rotates 30 degrees at the hinge between domains of I and II whereas furin cleavage only occurs in acidic conditions. A major rearrangement occurs when the virus fuses with the host cell. The anti-parallel E homodimers dissociates into monomers which re-associates into parallel homotrimers. During this transition, the E protein rotates 35 degrees to move towards the domain I and II hinge region.

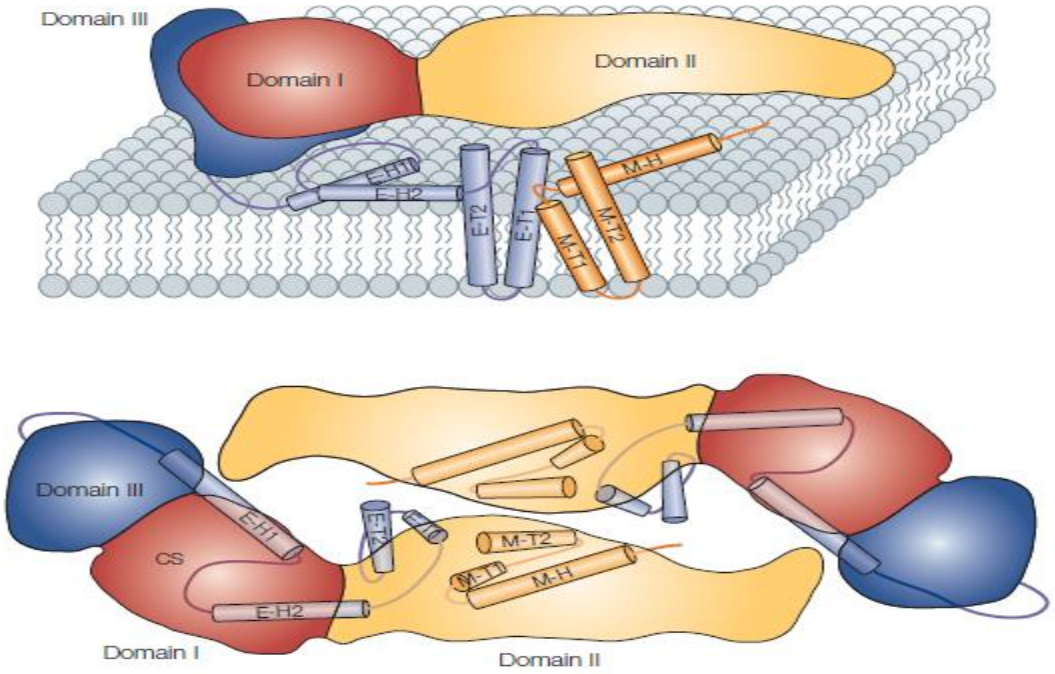

Figure 2.6: E protein ectodomain and transmembrane domain, side and top views

The class I fusion process is initiated by the acidification or virus binding to a receptor. The fusion peptide goes through a conformational change and inserts himself into the host cell membrane. Thus anchoring itself in both viral and cellular membranes. The fusion protein folds back onto itself which brings the transmembrane domain and fusion peptide close to each other. As a result bringing the viral and host membrane towards each other.On the other hand during class II process, a trimer of E protein is formed. Domain II of the E protein extends towards the host cell and the fusion peptide gets inserted into the membrane by bending at hinge region

between domain I and II. Hence it causes the fusion peptide to be anchored to both viral and host cell membranes.

When two lipid membranes are fused, the stem region is bound into the channel along with the side of domain II. As a result transmembrane domains and fusion of E protein are in close proximity which is the same for class I process.(“mukhopadhyay2005 (1),” n.d.)

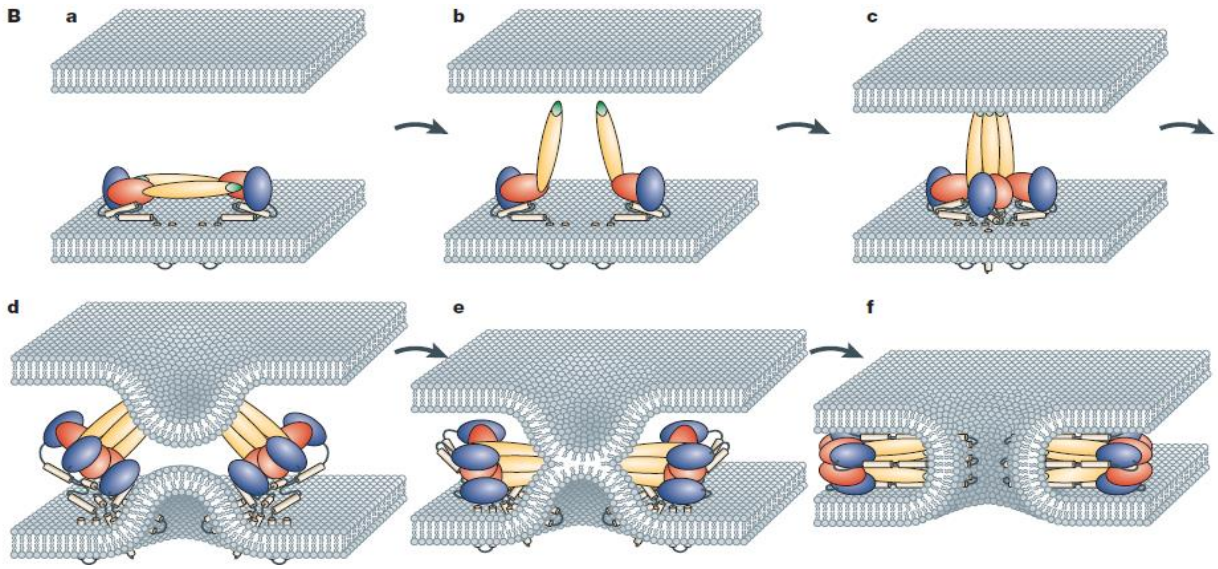

Figure 2.7: Class II fusion process

2.5 Sequence based methods:

Motif-search based approach:

The combinations of preferred amino acids at a peptide anchor position is known as a motif. It is the most outdated method but still widely used for epitope prediction. The peptide sequence is searched for a motif from a motif library. The MHC binding motif for a peptide can be compared with binders and non-binders. EPIPREDICT is a tool that is used to predict the MHC class II binding epitopes. SYFPEITHI is used to score and evaluate peptide sequences. Whereas EPIMER is another tool that is used to predict HIV related epitopes. The accuracy is around 60-

70%, the correlation between predicted and experimentally determined affinity is weak. (Patronov & Doytchinova, 2013)

Artificial neural network:

A desirable method which is used for finding relationship and describing non-linear data. ANN methods are used by bioinformatics researchers to study on drug solubility, cardiac disease, epitope prediction etc. During epitope prediction the peptide length can be variable. The sequences in the set are aligned by keeping a reference as an anchor position. For MHC class I constructing models are simpler as the peptide length does not vary much whereas MHC class II prediction becomes complex due to the variability of fragment size. NETCTL server uses a method to integrate Class I binding, proteasomal C-terminal cleavage and Transported associated antigen processing (TAP) transporter efficiency. NETMHC is a server that is created based on weight matrices and ANN.(Patronov & Doytchinova, 2013)

Support vector machine:

Support vector machine is a computer science concept that is made to comprehend data analysis and pattern recognition. It is created by Vapnix and was initially created for image classification and regression analysis. It is classified as a non-probabilistic binary linear classifier. Hence it is represented as a two sets of points in space which falls into two sub categories that are separated by a definite gap. (Patronov & Doytchinova, 2013)

Hidden Markov Model:

Hidden Markov Model were first applied for speech recognition in the1970. In the 1980s they found their application in the biological sequences. It has been widely used in bioinformatics, proteomics, predicting protein sequences, protein homology analysis and transmembrane region. It is also used for alignment of sequences, protein family identification by Pfam and SMART. Additionally it is used for studying gene splicing and phylogenetic tress. PREDTAI was used for the prediction of peptide binding to hTAP which requires a three layer propagation network with the sigmoid activation function.(Patronov & Doytchinova, 2013)

Chapter 3: Materials and Methods

Chapter 3: Materials and Methods

3.1 Software tools used to analyse:

3.1.1 Clustal Omega-Embl:

This is a multiple sequence alignment program at allows us to align multiple sequence alignments of divergent sequences. Evolutionary relationships can also be viewed as Cladograms or Phylograms. Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences.(Sievers et al)

3.1.2 MEGA 6:

This is an integrated tool for conducting alignments both manually and automatically, it also allows us to create phylogenetic trees. The alignments can be done using various matrices and parameters can be adjusted according to the user needs.Molecular Evolutionary Genetics Analysis (**MEGA**) is computer software for conducting statistical analysis of molecular evolution and for constructing phylogenetic trees. It includes many sophisticated methods and tools for phylogenomics and phylomedicine. It is licensed as proprietaryfreeware.(Tamura,Stecher,Peterson,Filipski,and Kumar 2013).

3.1.3 Jalview:

It is a program that is used for visualisation, editing and analysis of multiple sequence alignments. It is a free program that is accessible to all users and is ideal to visualise the conserved regions after multiple sequence alignment. Jalview is a free, open source program developed for the interactive editing, analysis and visualization of multiple sequence alignments. It can also work with sequence annotation, secondary structure information, phylogenetic trees and 3D molecular structures.(Waterhouse et.al)

3.1.4 Protein Variability Server:

A web based tool that enables to check protein sequences using variability metrics for computing the absolute site variability in multiple sequence alignments. PVS can perform several output task such as: Plot variability , Mask variability in sequence. Return Conserved fragments or Map structural variability. By default, PVS will plot the sequence variability. When Mask variability in sequence or Return Conserved fragments are selected, a variability threshold must be provided.This parameter has to be set within the range of 0 to 4.3 (default is 1.0) when Shannon is the selected variability method, and within the range of 0 to 1 (default is 0.46) when Simpson is the selected variability method. Those positions with a variability value above the selected threshold are filtered out.Positions with a variability value under the selected threshold are considered of low variability (highly conserved). If both Shannon and Simpson methods are selected, PVS will proceed considering the variability threshold as for Shannon.**(**Garcia-Boronat et.al)

3.1.5 VaxiJen v2.0:

It is a first server that predicts protective antigens and subunit vaccines. A designated threshold is given based on the type of organism. The algorithm allows the server to check antigenicity of the given sequence. The length of the sequence has to be greater than six residues.VaxiJen is the first server for alignment-independent prediction of protective antigens. It was developed to allow antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment. The server can be used on its own or in combination with alignment-based prediction methods. (Irini et al)

3.1.6 Bepipred:

It is a web server that allows us to predict B-cell epitopes from antigen sequences. It is based on a random forest algorithm and displays results in a user-friendly and informative way.For each input sequence, Bepipred outputs a prediction score for every residue. The positions of the linear B-cell epitopes are predicted to be located at the residues with the highest scores. The expected

sensitivity/specificity of the prediction is a function of the score cutoff or threshold the user chooses. (Larsen et al)

3.1.7 BCPREDS:

It is another web server that helps us to predict B-cell epitopes from sequences. It currently uses three prediction methods. It lets the user change various parameters such as sequence length and specify threshold as well.Because it is often valuable to compare predictions of multiple methods, and consensus predictions are more reliable than individual predictions, the BCPREDS server allows users to choose the method for predicting B-cell epitopes among several developed prediction methods.

The current implementation of BCPREDS allows the user to select among three prediction methods: (i) our implementation of AAP method [Chen et al., 2007]; (ii) BCPred [EL-Manzalawy et al., 2008]; (iii) FBCPred [EL-Manzalawy et al., 2008b]. Users provide an antigen sequence and optionally can specify desired epitope length and specificity threshold. Results are returned in several user-friendly formats. (El-Manzalawy et al)

3.1.8 Kolaskar and Tangaonkar antigenicity prediction tool:

It is the simplest method for the prediction of antigenic determinants. The method is created by the presence of amino acids residues in experimentally determined epitopes.Kolaskar and Tangaonkar's method predicts antigenic epitopes of given sequence, based on physicochemical properties of amino acid residues that frequently occur in experimentally determined antigenic epitopes. Previously reported data appreciated this method as it gives 75% experimental accuracy. **(**Kolaskar et. al)

3.1.9 Emni surface accessibility prediction tool:

The calculation is based on surface accessibility scale on a product instead of an addition within the window. It is one of the parameters required to be an ideal epitope. The calculation was based on surface accessibility scale on a product instead of an addition within the window. The accessibility profile was obtained using the formulae $Sn = (n+4+i)\ (0.37)^{-6}$ where Sn is the surface probability, dn is the fractional surface probability value, and i vary from 1 to 6. A

hexapeptidesequence with Sn greater than 1.0 indicates an increased probability for being found on the surface.(Vita et. al)

3.1.10 Karplus and Schulz Flexibility prediction tool:

In this method, flexibility scale based on mobility of protein segments on the basis of the known temperature B factors of the a-carbons of 31 proteins of known structure was constructed. The calculation based on a flexibility scale is similar to classical calculation, except that the center is the first amino acid of the six amino acids window length, and there are three scales for describing flexibility instead of a single one.**(**Karplus et. al)

3.1.11 Parker Hydrophilicity prediction tool:

In this method, hydrophilic scale based on peptide retention times during high-performance liquid chromatography (HPLC) on a reversed-phase column was constructed. A window of seven residues was used for analyzing epitope region. The corresponding value of the scale was introduced for each of the seven residues and the arithmetical mean of the seven residue value was assigned to the fourth, (i+3), residue in the segment. **(**Parker et. al)

Chapter 4: Results

Chapter 4: Results

4.1 Multiple sequence alignment by CLUSTAL OMEGA:

30 sequences of E protein was retrieved from NCBI database.Multiple sequence alignment of 30 Yellow fever virus strains were performed.The dashes indicate the mismatched sequences among them whereas red marked region indicates the conserved region of E protein. Multiple Sequence alignment (MSA) was performed using CLUSTAL OMEGA as shown below:

Figure 4.1: CLUSTAL O(1.2.4) multiple sequence alignment:

```
AAA92696.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETAAIDGPAEARKVC
AAA92705.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92706.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAG24925.1      ------------------------------------------------------------
AAD45534.1      ------------------------------------------------------------
AAD45533.1      ------------------------------------------------------------
AAD45531.1      ------------------------------------------------------------
AAD45532.1      ------------------------------------------------------------
AAA92691.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92692.1      AHCIGITDRDFIEGVHGGTWLSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92694.1      AHCIGVADRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPVEARKVC
AAX47569.1      AHCIGIADRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPVEARKVC
AAA92693.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92698.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92700.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAR86693.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLETVAIDGPAEARKVC
AAA92702.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAY68350.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAY68346.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAY68347.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAY68348.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAY68349.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92697.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAX47570.1      AHCIGITDRDFIEGVHGGTWVSATLEQGKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92699.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92703.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAX47568.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92704.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92695.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC
AAA92701.1      AHCIGITDRDFIEGVHGGTWVSATLEQDKCVTVMAPDKPSLDISLQTVAIDGPAEARKVC


AAA92696.1      YSAVLTNVKINDKCPSTGEAHLEEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92705.1      YSAVLTNVKINDKCPSTGEAHLEEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92706.1      YSAVLTNVKINDKCPSTGEAHLEEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAG24925.1      ------------------------------------------------FGKGSIVACAKFT
AAD45534.1      ------------------------------------------------FGKGSIVACAKFT
AAD45533.1      ------------------------------------------------FGKGSIVACAKFT
```

```
AAD45531.1      ------------------------------------------------FGKGSIVACAKFT
AAD45532.1      ------------------------------------------------FGKGSIVACAKFT
AAA92691.1      YSAVLTHVKINDKCPSTGEAHLAEENEGDHACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92692.1      YSAVLTHVKINDKCPSTGEAHLAEENEGDHACKRTYSDRVWGNGCGLSGKGSIVACAKFT
AAA92694.1      YNAVLTHVKIDDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAX47569.1      YNAVLTHVKINDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92693.1      YNAVLTHVKINDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92698.1      YNAVLTHVKINDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92700.1      YNAVLTHVKINDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAR86693.1      YNAVLTHVKINDKCPSTGEAHLAEENEGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92702.1      YSAVLTHVKINDKCPSTGEAHLAEESDGDNACNRTYSDRGWGNGCGLFGKGSIVACAKFT
AAY68350.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAY68346.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAY68347.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAY68348.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAY68349.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92697.1      YSAVLTHVKINDKCPSTGEAHLAEENDGGNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAX47570.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92699.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92703.1      YSAVLTHVKINDKCPSTGEAHLAEENDGGNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAX47568.1      YSAVLTHVKINDKCPSTGEAHLAEENDGGNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92704.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92695.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
AAA92701.1      YSAVLTHVKINDKCPSTGEAHLAEENDGDNACKRTYSDRGWGNGCGLFGKGSIVACAKFT
                                                             ************

AAA92696.1       CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92705.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGRATL
AAA92706.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAG24925.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAD45534.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAD45533.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAD45531.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAD45532.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAA92691.1      CAKSMSLFEVDQTKIQYVIRAQLHVGANQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAA92692.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNADIKTLKFDALSGSQEAEFTGYGKATL
AAA92694.1      CAKSMSLFEVDQTKIQYVIKAQLHVGAKQEDWKTDIKTLKFDVLSGSQEAEFTGYGKVTL
AAX47569.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWKTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92693.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92698.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92700.1      CAKSMSLFEVDQTKIQYVIRAQLHVGGKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAR86693.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92702.1      CAKSMSLFEVDQTNIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTDYGKATL
AAY68350.1       CAKSMSLFEVDRTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSDSQEAEFSRYGKATL
AAY68346.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSDSQEAEFSRYGKATL
AAY68347.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSDSQEAEFSRYGKATL
AAY68348.1      CAKSMSPFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSDSQEAEFSRYGKATL
AAY68349.1      CAKSMSPFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSDSQEAEFSRYGKATL
AAA92697.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAX47570.1      CAKSMSLFEVDRTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92699.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92703.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAX47568.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92704.1      CAKSMSLFEVDQTKIQYVIRAQLHVGAKQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92695.1      CAKSMSLFEVDQTKIQYVIRAQLHVGANQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
AAA92701.1      CAKSMSLFEVDQAKIQYVIRAQLHVGANQENWNTDIKTLKFDALSGSQEAEFTGYGKATL
                ****** ****:::*****:******.:**:*::********.**.******: **:.**
```

```
AAA92696.1      ECQVQTAVDFSNSYIAEMEKESWIVDKQWAQDLTLPWQSGSHGVWREMHHLVEFEPPHAA
AAA92705.1      ECQVQTAVDFSNSYIAEMEKESWIVDKQWAQDLTLPWQSGSRGVWREMHHLVEFEPPHAA
AAA92706.1      ECQVQTALDFSNSYIAEMEKESWIVDKQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAG24925.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAD45534.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAD45533.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMYHLVEFEPPHAA
AAD45531.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAD45532.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92691.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92692.1      ECQVQTAVDFSNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92694.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAX47569.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92693.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92698.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92700.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAR86693.1      ECQVQTAVDFGNSYIAEMEKESWIVDRQWAQDLTLPWQSGSGGVWREMHHLVEFEPPHAA
AAA92702.1      ECQVPAAVDFGNSYIAEMEKDSWIVVRQWALDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAY68350.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAY68346.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAY68347.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAY68348.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAY68349.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAA92697.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAX47570.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGTWREMHHLVEFEPPHAA
AAA92699.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAA92703.1      ECQVQTAVDFGNSYIAEMEKDSWIVVRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAX47568.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAA92704.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAA92695.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
AAA92701.1      ECQVQTAVDFGNSYIAEMEKDSWIVDRQWAQDLTLPWQSGSGGIWREMHHLVEFEPPHAA
                **** :*:**.*********:**** :*** ********** * ****:***********

AAA92696.1      TIKVLALGNQEGSLKTALTGAMRVTKDTSNSKLYKLHGGHVACRVKLSALTLKGTSYKMC
AAA92705.1      TIKVLALGNQEGSLKTALTGAMRVTKDTNNSKLYKLHGGHVACRVKLSALTLKGTSYKMC
AAA92706.1      TIKVLALGNQEGSLKTALTGAMRVTKDTNNSKLYKLHGGHVACRVKLSALTLKGTSYKMC
AAG24925.1      TIKVLALGNQDGSLTTALTGAMRVTKDTNGSNLYKLHGGHVSCRVKFSALTLKGTSYKMC
AAD45534.1      TIKVLALGNQEGSLKTALTGAMRVTKATNGSNLSRLHGGHVSCRVKLSALTLKGTSYKMC
AAD45533.1      TIKVLALGNQEGSLKTALTGAMRATKDTNGSNLYKLHGGHVSCRVKLSVLTLKGTSYKMC
AAD45531.1      TIKVLALGNQEGSLKTALTGPMRATKDTNGSNLYKLHGGHVSCRVKLSVLTLKGTSYKMC
AAD45532.1      TIKVLALGNQEGSLKTALTGPMRATKDTNGSNLSKLHGGPVSCRVKLSVLTLKGTSYKMC
AAA92691.1      TIKVLALGNQEGSLKTALTGAMRVTKDTNGSNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92692.1      TIKVLALGNQEGSLKTALTGAMRVTKDTNGSNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92694.1      TIRVLALGDQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAX47569.1      TIRVLALGDQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92693.1      TIRVLALGNQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92698.1      TIRVLALGNQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92700.1      TIRVLALGNQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAR86693.1      TIRVLALGNQEGSLKTALTGAMRVTKDTNDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92702.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAY68350.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAY68346.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAY68347.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAY68348.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAY68349.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92697.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNIYNLHGGHVSCRVKLSALTLKGTSYKMC
AAX47570.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
```

```
AAA92699.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92703.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAX47568.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92704.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92695.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
AAA92701.1      TIRVLALGNQEGSLKTALTGAMRVTKDENDNNLYKLHGGHVSCRVKLSALTLKGTSYKMC
                **:*****:*:***.***** **.**  ...:: .**** *:****:*.***********

AAA92696.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTAAVNKGILVTVNPIASTND
AAA92705.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTAAVNKGILVTVNPIASTND
AAA92706.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTAAVNKGILVTVNPIASTND
AAG24925.1      TDKMSLVKNPTDTGHGTAVMEVKVPKGAPCRIPVMVADDLTASVNKGILVTVNPIASTNE
AAD45534.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCGIPVMVADDLTASVNKGILVTVNPIASTNE
AAD45533.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCGIPVMVADDLTASVNKGILVTVNPIASTNE
AAD45531.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTASVNKGILVTVNPIASTNE
AAD45532.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTASVNKGILVTVNPIASTNE
AAA92691.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTASVNKGILVTVNPIASTNE
AAA92692.1      TDKMSFVKNPTDTGHGTAVMQVKVPKGAPCRIPVMVADDLTASVNKGILVTVNPIASTNE
AAA92694.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCRIPVIVADDLTAAINKGILVTVNPIASTND
AAX47569.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCRIPVIVADDLTAAINKGILVTVNPIASTND
AAA92693.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAINKGILVTVNPIASTND
AAA92698.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAINKGILVTVNPIASTND
AAA92700.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAINKGILVTVNPIASTND
AAR86693.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAINKGILVTVNPIASTND
AAA92702.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAY68350.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAY68346.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAY68347.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAY68348.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAY68349.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAA92697.1      TDKMSFVKNPTDSGHGTVVMQVKVSKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAX47570.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAA92699.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVTPIASTND
AAA92703.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAX47568.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAA92704.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAA92695.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
AAA92701.1      TDKMSFVKNPTDTGHGTVVMQVKVPKGAPCKIPVIVADDLTAAVNKGILVTVNPIASTND
                *****:******:****.**:*** ***** ***:*******::********.******:

AAA92696.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92705.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92706.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAG24925.1      DDVLIEVNPPFGDSPFIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAD45534.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAD45533.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAD45531.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAD45532.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92691.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92692.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWHF
AAA92694.1      DEVLIEVNPPFGDSYIIIGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAX47569.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92693.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92698.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92700.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAR86693.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92702.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
```

```
AAY68350.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAY68346.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAY68347.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAY68348.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAY68349.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92697.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAX47570.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92699.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92703.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAX47568.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92704.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92695.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
AAA92701.1      DEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLAVMGDAAWDF
                *:************ :*:**************************************.*

AAA92696.1      GSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGIVLIWVGINTRNMTMSMSMILV
AAA92705.1      GSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGVVLIWVGINTRNMTMSMSMILV
AAA92706.1      GSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGVVLIWVGINTRNMTMSMSMILV
AAG24925.1      SSAGGSFTSVGKGIHMVFGSAFQG------------------------------------
AAD45534.1      SSAGGFFTSVGKGIHMVFGSAFQG------------------------------------
AAD45533.1      SSAGGFFTSVGKGIHMVFGSAFQG------------------------------------
AAD45531.1      SSAGGFFTSVGKGIHMVFGSAFQG------------------------------------
AAD45532.1      SSAGGFFTSVGKGIHMVFGSAFQG------------------------------------
AAA92691.1      SSAGGFFTSVGKGIHMVFGSAFQGLFGGLSWITKVIMGAVLIWVGINMRNMTMSMSMILV
AAA92692.1      SSAGGFFTSVGKGIHMVFGSAFQGLFGGLSWITKVIMGAVLIWVGINMRNMTMSMSMILV
AAA92694.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLNWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAX47569.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLNWITKVIIGAVLIWVGINTRNMTMSMSMILV
AAA92693.1      SSAGGLFTSIGKGIHTVFGSAFQGLFGGLNWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92698.1      SSAGGLFTSIGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92700.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAR86693.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92702.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAY68350.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAY68346.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAY68347.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAY68348.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAY68349.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92697.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAX47570.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92699.1      SSAGGFFTSVGKGIHPVFGSAFQGLFGGLSWITKVIMGAVHIWVGINTRNMTMSMSMILV
AAA92703.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAX47568.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92704.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92695.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
AAA92701.1      SSAGGFFTSVGKGIHTVFGSAFQGLFGGLSWITKVIMGAVLIWVGINTRNMTMSMSMILV
                .**** ***:***** ********

AAA92696.1      GVIMMFLSLGVGA
AAA92705.1      GVIMMFLSLGVGA
AAA92706.1      GVIMMFLSLGVGA
AAG24925.1      -------------
AAD45534.1      -------------
AAD45533.1      -------------
AAD45531.1      -------------
AAD45532.1      -------------
AAA92691.1      GVIMMFLSLGVGA
AAA92692.1      GVIMMFLSLGVGA
```

```
AAA92694.1      GVIMMFLSLGVGA
AAX47569.1      GVIMMFLSLGVGA
AAA92693.1      GVIMMFLSLGVGA
AAA92698.1      GVIMMFLSLGVGA
AAA92700.1      GVIMMFLSLGVGA
AAR86693.1      GVIMMFLSLGVGA
AAA92702.1      GVIMMFLSLGVGA
AAY68350.1      GVIMMFLSLGVGA
AAY68346.1      GVIMMFLSLGVGA
AAY68347.1      GVIMMFLSLGVGA
AAY68348.1      GVIMMFLSLGVGA
AAY68349.1      GVIMMFLSLGVGA
AAA92697.1      GVIMMFLSLGVGA
AAX47570.1      GVIMMFLSLGVGA
AAA92699.1      GVIMMFLSLGVGA
AAA92703.1      GVIMMFLSLGVGA
AAX47568.1      GVIMMFLSLGVGA
AAA92704.1      GVIMMFLSLGVGA
AAA92695.1      GVIMMFLSLGVGA
AAA92701.1      GVIMMFLSLGVGA
```

Figure 4.1: Multiple Sequence Alignment done by Clustal Omega

4.2 Multiple Sequence Alignment to visualize by the software Jalview as shown below:

Jalview was used to visualize our Multiple Sequence Alignment (MSA) of 30 strains of Yellow Fever viruses. The yellow stretch indicates the conserved region of E protein.The brown region indicates the mismatched sequences among them.Our conserved region starts from the position 108 and ends at 445 residue.

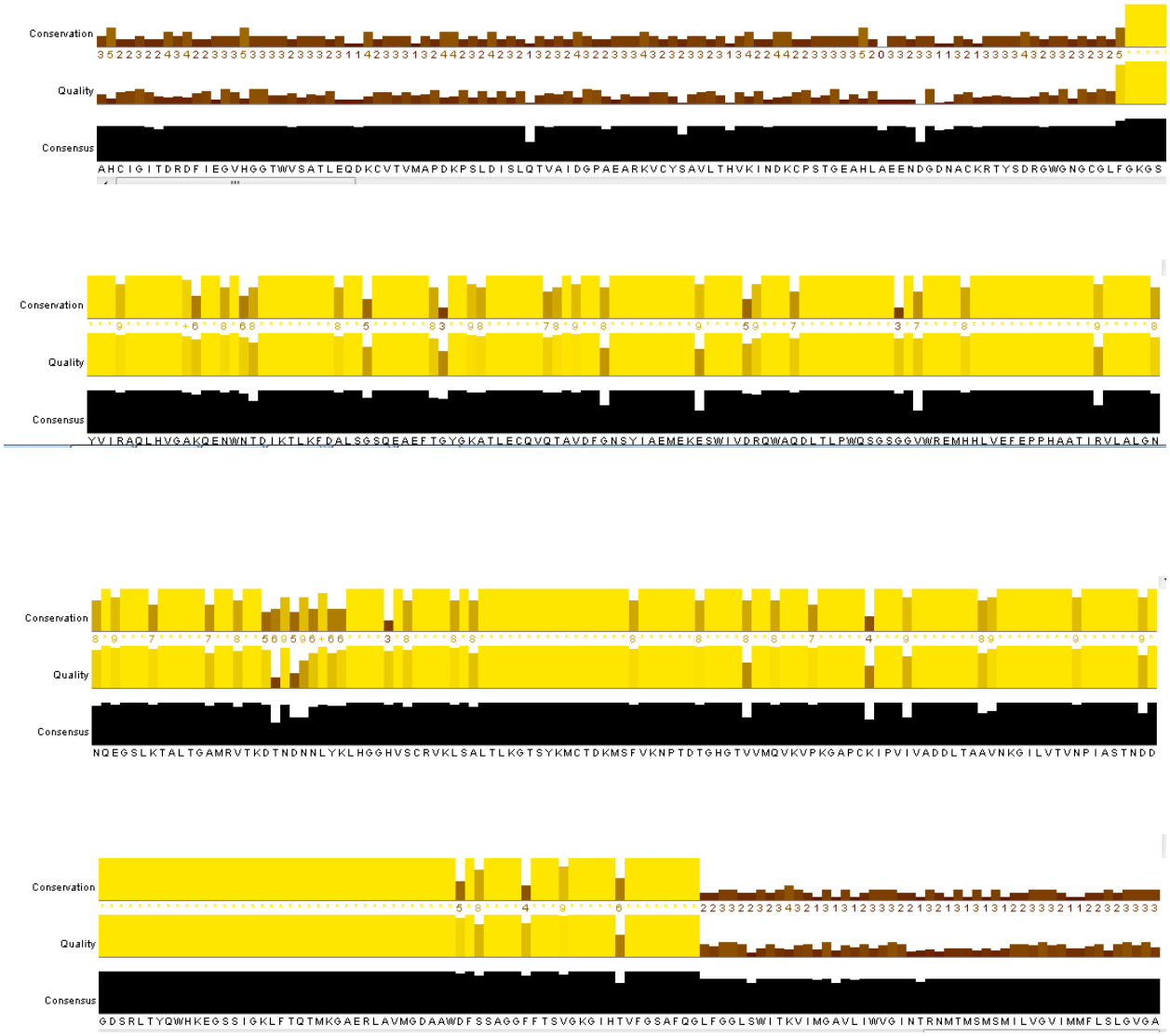

Figure 4.2: MSA Image produced by Jalview

Conserved Sequence of E protein was found as given below:

PPHAATIRVLALGNQEGSLKTALTGAMRVTKDENDNNIYNLHGGHVSCRVKLSALTLK
GTSYKMCTDKMSFVKNPTDSGHGTVVMQVKVSKGAPCKIPVIVADDLTAAVNKGILVT
VNPIASTNDDEVLIEVNPPFGDSYIIVGTGDSRLTYQWHKEGSSIGKLFTQTMKGAERLA
VMGDAAWDFSSAGGFFTSVGKGIHTVFGSAFQ

The number of residues for conserved region is 337 amino acids. It starts from the position of 108 and ends at 445.

4.3Antigenicity results by Vaxijen:

Antigenicity of the conserved E protein was checked using Vaxijen v2.0 server which showed it to be antigenic. The result was 0.4588 which was above the threshold of 0.4 for virus.The size of our conserved E region is 337 residues.Vaxijen was used to check for antigenicity and the threshold for virus was 0.4

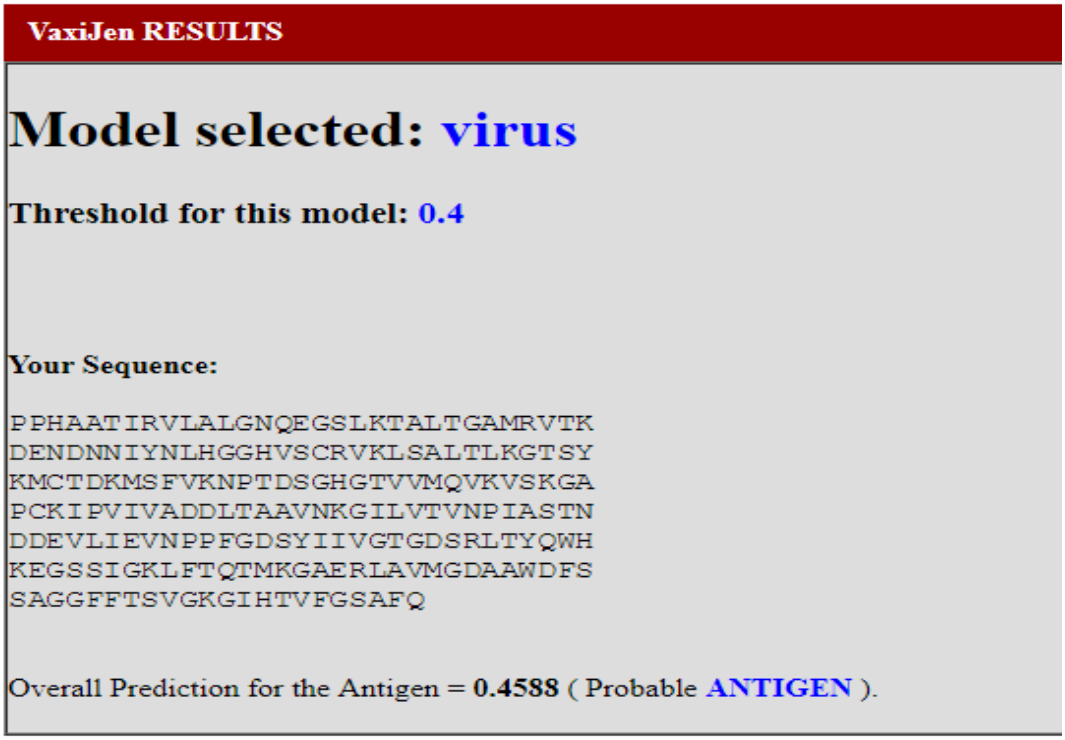

Figure 4.3: Vaxijen results showing antigenicity of conserved E protein of YFV

4.4The antigenicity results byKolaskar and Tongaonkar:

The antigenicity was checked by Kolaskar and Tangaonkar. Our window size was taken 7 with a center position of 4. The threshold was taken as 1.027. The yellow peaks indicate the residues to be above the threshold.The green peaks indicate the residues to be below the threshold.103 out of 209 residues is above the threshold.

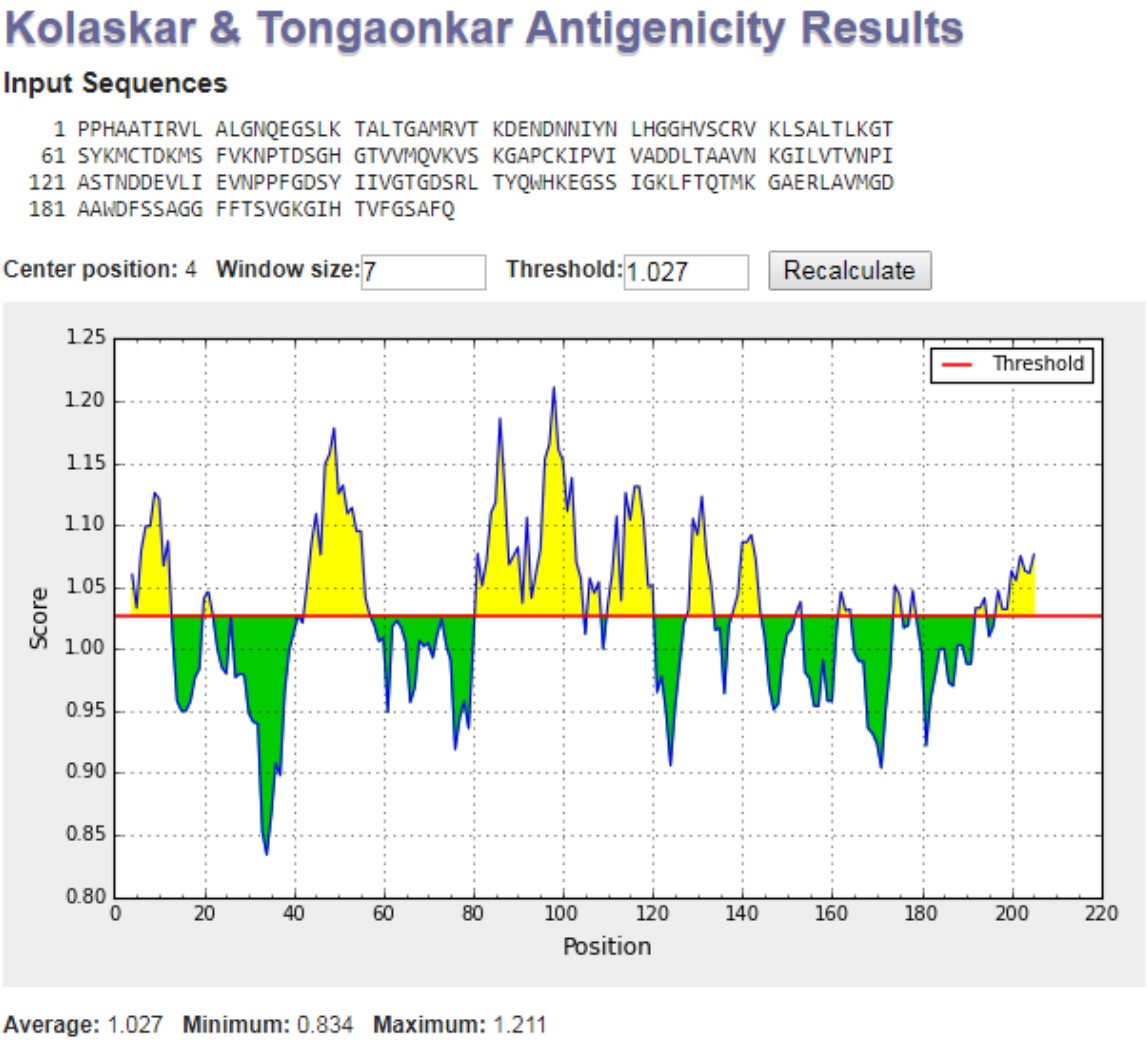

Figure 4.4: Kolaskar and Tangoankar results showing antigenicity of conserved E protein

4.5Protein variability:

For checking the variability of the protein sequences we have used Protein Variability server. The red bars indicate the variability in the protein sequence. Our conserved E region starts from 108 and ends at 445.There is little variability indicating with few red bars shown in our conserved region. The results show little variability in the sequences as shown below:

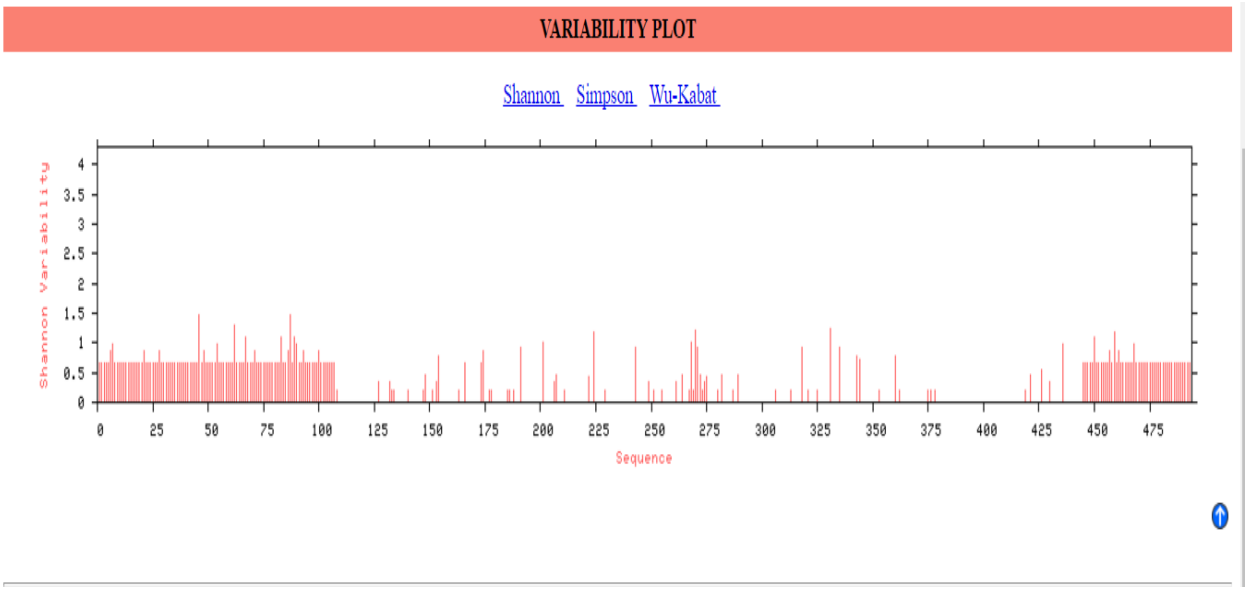

Figure 4.5: Results showing little variability by Protein Variability Server

.

4.6Epitope prediction:

Bepipred and BCPREDS epitope prediction tools were used to find out epitopes from our conserved E protein region. The size range was taken to be six or above amino acids. Bepipred gave us 10 epitopes whereas BCPREDS gave us 14 epitopes from our conserved region as shown below:

Epitopes found using BCPREDS :

1. TKDTNGSNLYKL
2. VPKGAPCRIPVM
3. EMEKESWIVDRQ
4. NPPFGDSYIIVG
5. HVGANQENWNAD
6. FVKNPTDTGHGT
7. FTGYGKATLECQ
8. GNQEGSLKTALT
9. TLKGTSYKMCTD
10. YQWHKEGSSIGK
11. GGVWREMHHLVE
12. NPIASTNEDEVL
13. ACAKFTCAKSMS
14. AGGFFTSVGKGI

Epitopes found using Bepipred:

1.GANQENWNA (146-154)
2.LSGSQEAEFTGY (164-175)
3.LPWQSGSGGV (215- 224)
4.EPPHAA (235-240)
5.NQEGSLK (249-255)
6.RVTKDTNGS (263-271)
7.VKNPTDTGHGT (307-317)
8.VPKGAPC (324-330)
9.IASTNE (355-360)
10.HKEGSS (390-395)

Common peptides between both servers are:

The red marked ones indicate the common epitopes. There are four common epitopes found between the two servers.The size range is taken to be above six residues.

Epitopes identified by BepiPred 1.0 server	Epitopes identified by BCPREDS server
1.GANQENWNA (146-154)	1. HVGANQENWNAD (144-155)
2.LSGSQEAEFTGY (164-175)	2. FTGYGKATLECQ (172-183)
3.LPWQSGSGGV (215- 224)	3. GGVWREMHHLVE (222-233)
4.EPPHAA (235-240)	4. TLKGTSYKMCTD (291-302)
5.NQEGSLK (249-255)	5. GNQEGSLKTALT (248-259)
6.RVTKDTNGS (263-271)	6. TKDTNGSNLYKL (265-276)
7.VKNPTDTGHGT (307-317)	7. FVKNPTDTGHGT (306-317)
8.VPKGAPC (324-330)	8. VPKGAPCRIPVM (324-335)
9.IASTNE (355-360)	9. NPIASTNEDEVL (353-364)
10.HKEGSS (390-395)	10. YQWHKEGSSIGK (387-398)

Among the four epitopes, three were antigenic as shown below but only one peptide was only conserved in majority of the sequences. VKNPTDTGHGT with a peptide length of 11 amino acids had the conservancy of 96.67%. It was conserved in 29 out of 30 sequences.

4.6.1Antigenicity of common epitopes was checked using VaxiJen v2.0 server as shown below:

No.	Epitope	Antigenic	Value
1	GANQENWNA	Yes	0.7463
2	NQEGSLK	No	-0.1216
3	TKDTNGS	Yes	0.7006
4	VKNPTDTGHGT	Yes	1.0508

4.6.2 Conservancy of epitopes was checked using IEDB conservancy analysis tool as shown below:

No.	Epitope sequence	Length	Conservancy
1	GANQENWNA	9	No
2	NQEGSLK	7	No
3	TKDTNGS	7	No
4	VKNPTDTGHGT	11	Yes

4.7 Parameters of ideal epitope:

The results of the epitope VKNPTDTGHGT using the parameters are shown below:

Emni surface accessibility:

Our epitope was checked for surface accessibility by Emni surface accessibility.Our window size was taken 6 with a center position of 3. The threshold was taken as 1.000.The yellow peak indicates the residues that are above the threshold.3 out of 6 peptide fragments were above the threshold. Therefore we can conclude that VKNPTDTGHGT has the surface accessibility of an ideal epitope.

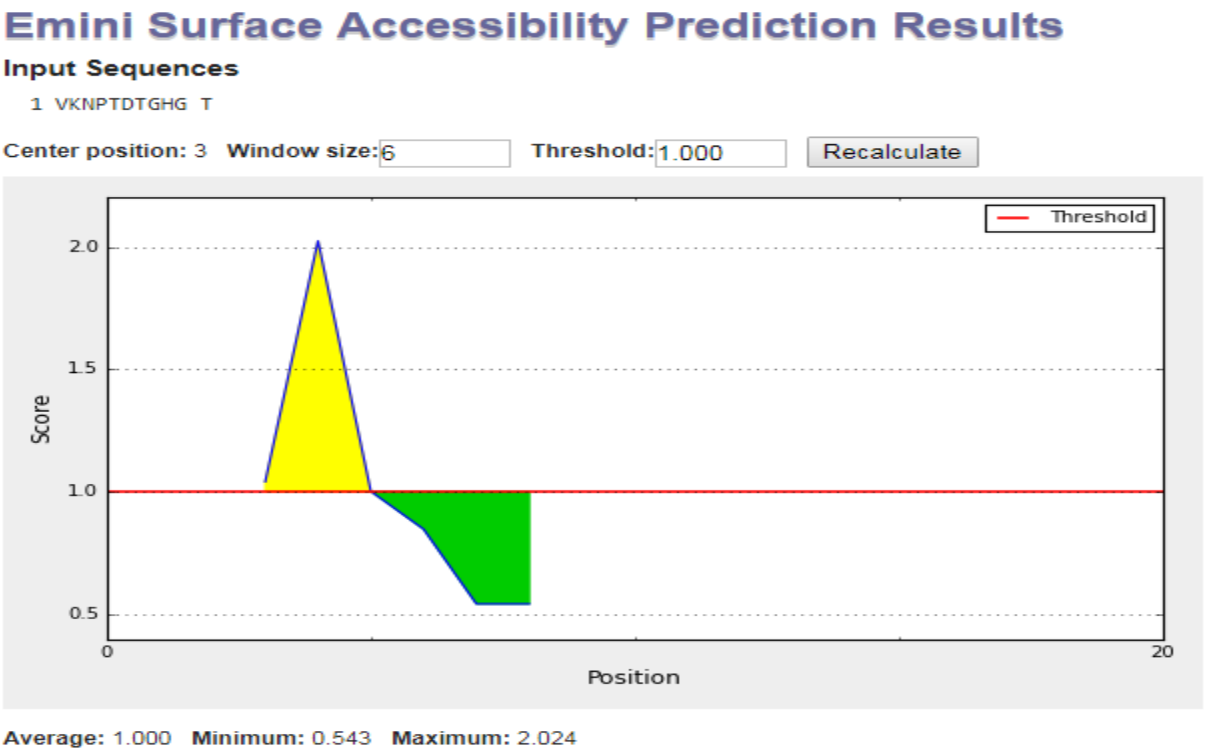

Figure 4.7.1: Results showing Emni surface accessibility of VKNPTDTGHT

Table 4.7.1: Peptide fragments found by Emni surface accessibility

Predicted residue scores:

Position	Residue	Start	End	Peptide	Score
3	N	1	6	VKNPTD	1.041
4	P	2	7	KNPTDT	2.024
5	T	3	8	NPTDTG	1.002
6	D	4	9	PTDTGH	0.848
7	T	5	10	TDTGHG	0.543
8	G	6	11	DTGHGT	0.543

Karplus and Schulz flexibility:

Our epitope was checked for its flexibility by Karplus and Schulz. Our window size was taken 7 with a center position of 4. The threshold was taken as 1.059. The yellow peak indicates the residues that are above the threshold. 2 out of 4 peptide fragments were above the threshold. Therefore we can conclude that VKNPTDTGHGT has the flexibility of becoming an ideal epitope.

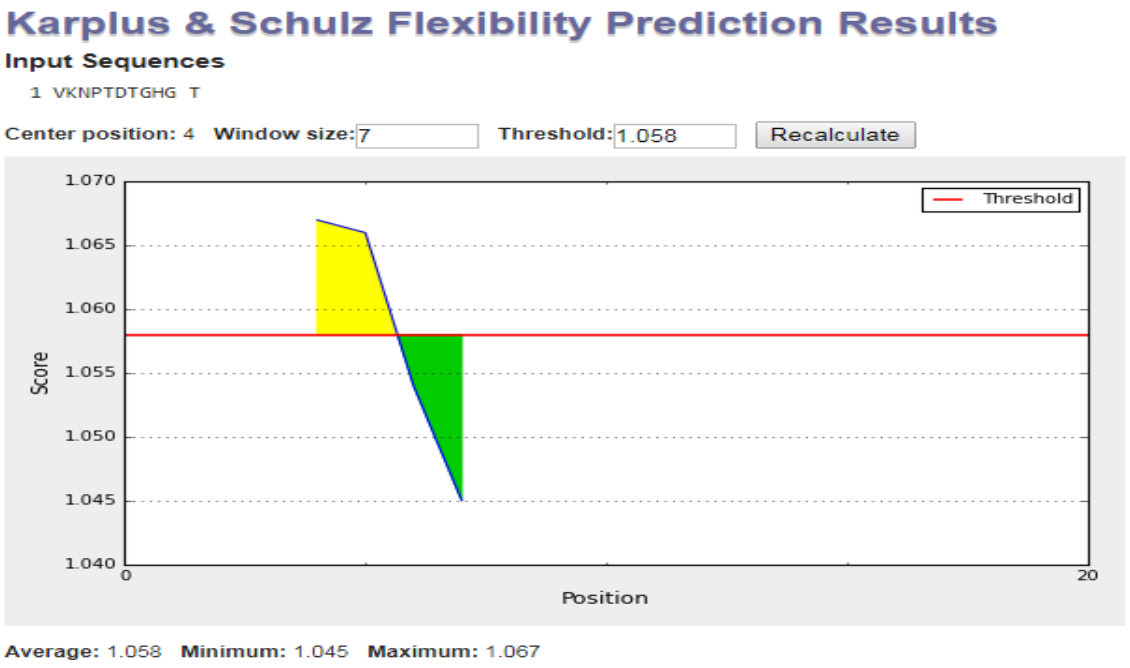

Figure 4.7.2: Results showing Karplus and Schulz Flexibility of VKNPTDTGHGT

Table 4.7.2: Peptide fragments found byKarplus and Schulz Flexibility

Predicted residue scores:

Position	Residue	Start	End	Peptide	Score
4	P	1	7	VKNPTDT	1.067
5	T	2	8	KNPTDTG	1.066
6	D	3	9	NPTDTGH	1.054
7	T	4	10	PTDTGHG	1.045

IEDB Parker hydrophilicity:

Our epitope was checked for hydrophilicity by Parker Hydrophilicity. Our window size was taken 7 with a center position of 4. The threshold was taken as 5.280. The yellow peak indicates the residues that are above the threshold. 3 out of 5 peptide fragments were above the threshold.Therefore we can conclude that VKNPTDTGHGT has the flexibility of becoming an ideal epitope.

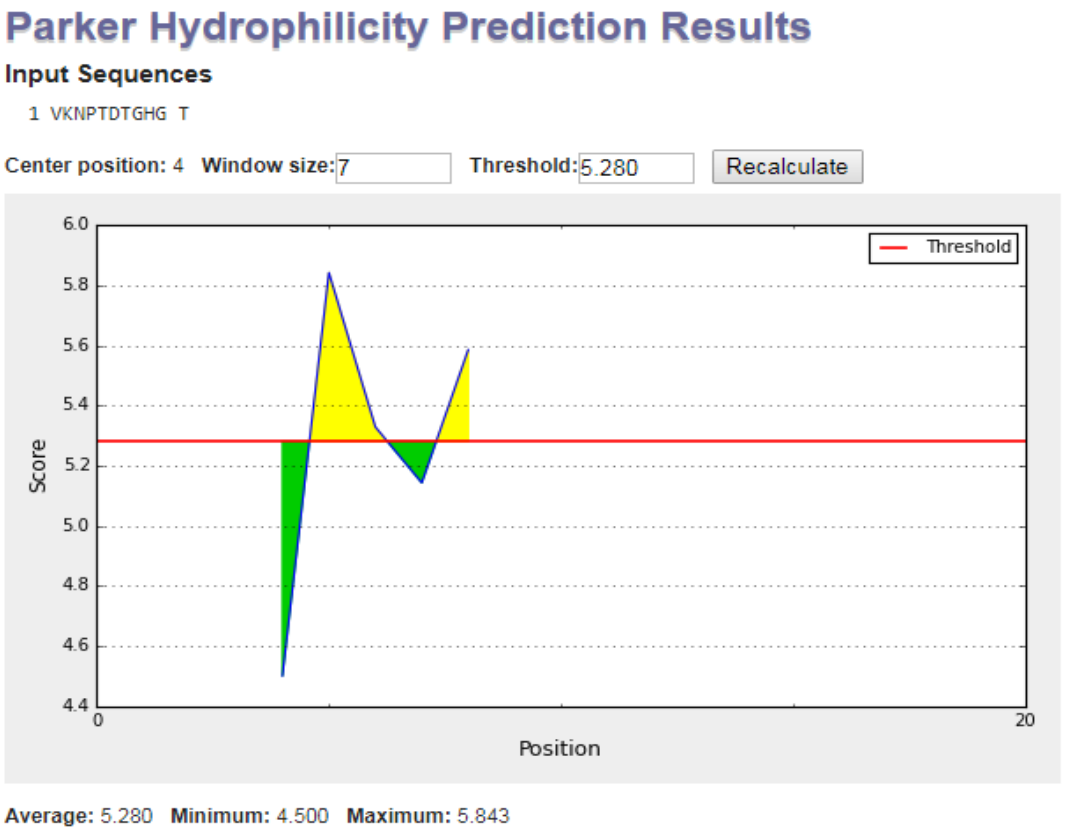

Figure 4.7.3: Results showing Parker Hydrophilicity of VKNPTDTGHGT

Table 4.7.3: Peptide fragments found byParker Hydrophilicity

Predicted residue scores:

Position	Residue	Start	End	Peptide	Score
4	P	1	7	VKNPTDT	4.5
5	T	2	8	KNPTDTG	5.843
6	D	3	9	NPTDTGH	5.329
7	T	4	10	PTDTGHG	5.143
8	G	5	11	TDTGHGT	5.586

Chapter 5: Discussion

Chapter 5: Discussion

The Yellow fever virus is a member of Flavivirus genus and Flaviviridae family. Its genome consists of a single positive-stranded RNA molecule with an approximate weight of 11kb. It consists of several structural proteins such as capsid, membrane with its precursor, envelope. The structural E protein of YFV is the most immunogenic part which is required for receptor binding and membrane fusion. A major component of virion surface and thus is usually targeted for antigenic purposes.Native E protein presents itself as a homodimer but goes through various conformational during receptor mediated endocytosis. The conformational change occurs in the low pH environment of the endosome during the fusion of viral envelope with the endosomal membrane.

The E protein sequence of 30 YFV strains were extracted from NCBI database and their MSA was done using CLUSTAL OMEGA. The conserved region of E protein was found to be 337 amino acids.It starts from the position of 108 and ends at 445. This stretch of amino acids was checked for antigenicity by a tool named Vaxijen v2.0 server. The results indicates that the region is antigenic with a value of 0.4588 and thus we can proceed to find epitopes from that conserved part of E protein.

The variability of the region was checked using Protein variability server (PVS). The results show that our conserved sequence has little variability. Another tool named Jalview was used to visualize our conserved region and later we introduced our fragment inside epitope prediction tools such as Bepipred and BCPREDS server. Bepipred predicted 10 epitopes as we indicated a size range to be equal to or greater than 6 amino acids whereas BCPREDS predicted 14 epitopes from our conserved E protein region. Common epitopes between both servers were taken for further investigation. We found four common epitopes which were checked for antigenicity by Vaxijen v2.0 server. 3 out of 4 epitopes were found to be antigenic by Vaxijen v2.0 server. Those 3 epitopes were again checked for conservancy using IEDB conservancy analysis tool. Conservancy is checked to see whether or not our peptide is present in all 30 sequences. Normally atleast 80% conservancy is required for that particular epitope.

Out of the 3 epitopes only the epitope VKNPTDTGHGT showed 96.67 conservancy. Our epitope has a length off 11 amino acids which starts from the position 307 and ends at 317. This epitope was further checked for surface accessibility, flexibility and hydrophilicity. Emni surface accessibility tool was used to check for surface accessibility and 3 out of 6 peptide fragments were above the threshold 1.0. Thus ensuring it to have the desired surface accessibility. Karplus and Schulz flexibility prediction analysis showed that 2 out of 4 peptide fragments in the VKNPTDTGHGT to be above the threshold. IEDB Parker hydrophilicity also showed that 3 out of 5 peptide fragments of our epitope to be above the threshold. Hence we can conclude that VKNPTDTGHGT is antigenic and conserved. It has surface accessibility, flexibility and hydrophilicity of an ideal epitope that could be further checked in the wet lab work.

Chapter 6: Conclusion

Chapter 6: Conclusion

Yellow fever virus belongs to the family Flaviviridae, genus Flavivirus. The YFV genome is single stranded, positive sense RNA, which encodes for a polyprotein. Attenuated Yellow Fever vaccines 17D/ 17DD are the only vaccines available to prevent the yellow fever infection. Both of them can elicit T-cell responses and neutralizing antibodies. Some toxic side effects and allergic reactions have been reported at some places. Whole cell vaccines are not taken as an alternative because they are expensive and time consuming to produce. DNA vaccines induce an antibody production, limited to protein immunogens and there is a chance of insertion of foreign DNA into the host genome. This may cause the cell to become cancerous. Whereas epitope based vaccine for YF can be used as a better alternative as it has high immunogenicity, hence may elicit a faster immune response. It is safer to use and can be easily stored. In addition, it can be readily produced at a cheaper price than the conventional YF vaccines.Among the viral proteins, the E protein is the most targeted one due to its high antigenic potential. It is involved in various purposes such as receptor binding site for viral attachment, fusion, penetration, haemagglutination, host range and cell tropism. The epitope found from the conserved E region predicted by the tools shows it has all the properties to be an ideal candidate. Hence, this peptide can be experimentally tested to check whether it induces a strong immune response or not. There should be more *in silico* work done in order to find specific epitopes for such viruses. It is our quickest way to find epitopes that could be lethal against pathogenic viruses.

References:

Barros, M. C. E. S., Galasso, T. G. C. M., Chaib, A. J. M., Degallier, N., Nagata, T., & Ribeiro, B. M. (2011). Yellow fever virus envelope protein expressed in insect cells is capable of syncytium formation in lepidopteran cells and could be used for immunodetection of YFV in human sera. *Virology Journal*, *8*(1), 261. https://doi.org/10.1186/1743-422X-8-261

Beck, A. S., & Barrett, A. D. T. (2015). Current status and future prospects of yellow fever vaccines. *Expert Review of Vaccines*, *14*(11), 1479–92. https://doi.org/10.1586/14760584.2015.1083430

Emini EA, Hughes JV, Perlow DS, Boger J. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol***55**:836-839

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. *A new bioinformatics analysis tools framework at EMBL-EBI (2010)* Nucleic acids research 2010 Jul, 38 Suppl: W695-9. https://doi:10.1093/nar/gkq313

Jens Erik Pontoppidan Larsen, Ole Lund and Morten Nielsen. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006; 2: 2.

Kametani, Y., Miyamoto, A., Tsuda, B., & Tokuda, Y. (2015). B Cell Epitope-Based Vaccination Therapy. *Antibodies*, *4*(3), 225–239. https://doi.org/10.3390/antib4030225

Keenan, B. P., & Jaffee, E. M. (2012). Whole cell vaccines--past progress and future strategies. *Seminars in Oncology*, *39*(3), 276–86. https://doi.org/10.1053/j.seminoncol.2012.02.007

Kuno, G., Chang, G. J., Tsuchiya, K. R., Cropp, C. B., Kuno, G., Chang, G. J., … Karabatsos, N. (1998). Phylogeny of the Genus Flavivirus Phylogeny of the Genus Flavivirus, *72*(1), 72–83.

Karplus PA, Schulz GE. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften***72**:212-213

Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett***276**:172-174

Lambert, D. M., & Fowler, C. J. (2005). Therapeutic Applications, *48*(16), 1077–1088.

Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. *Immunome Res***2**:2

mukhopadhyay2005 (1). (n.d.).

Nogueira, R. T., Nogueira, A. R., Pereira, M. C., Rodrigues, M. M., Galler, R., & Bonaldo, M. C. (2011). Biological and immunological characterization of recombinant Yellow Fever 17D Viruses expressing a Trypanosoma cruzi Amastigote Surface Protein-2 CD8 + T cell epitope at two distinct regions of the genome. *Virology Journal*, *8*, 127. https://doi.org/10.1186/1743-422X-8-127

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*, *3*(1), 120139. https://doi.org/10.1098/rsob.120139

Parker JM, Guo D, Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry***25**:5425-5432.

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins D*Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Molecular Systems Biology 7 Article number: 539 https://doi:10.1038/msb.2011.75

Volk, D. E., May, F. J., Gandham, S. H. A., Anderson, A., Von Lindern, J. J., Beasley, D. W. C., … Gorenstein, D. G. (2009). Structure of yellow fever virus envelope protein domain III. *Virology*, *394*(1), 12–18. https://doi.org/10.1016/j.virol.2009.09.001

Wu, J., Bera, A. K., Kuhn, R. J., Smith, L., & Smith, J. L. (2005). Structure of the Flavivirus Helicase : Implications for Catalytic Activity , Protein Interactions , and Proteolytic Processing Structure of the Flavivirus Helicase : Implications for Catalytic Activity , Protein Interactions , and Proteolytic Processing. *Journal of Virology*, *79*(16), 10. https://doi.org/10.1128/JVI.79.16.10268