In silico B-cell and T-cell epitope-based vaccine designing against Chikungunya virus

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

Submitted by: Sheikh Anushe

Student ID: 14136002

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

66, Mohakhali, Dhaka-1212

Bangladesh

May, 2018

Declaration

I, Sheikh Anushe, declare that this thesis and the work entitled "***In silico* B-cell and T-cell epitope-based vaccine designing against Chikungunya virus**", submitted to the Department of Mathematics and Natural Sciences (MNS), BRAC University in partial fulfillment of the requirements for the degree of Bachelor of Science in Biotechnology is a record of work carried out by me under the joint supervision of my supervisors.

I further declare that this thesis has been composed solely by me and it has not been submitted, in whole or in part, in any previous institution for a degree or diploma. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

Candidate

_____

Sheikh Anushe

Certified:

Shamira Tabrejee

Supervisor,

Lecturer, Biotechnology Program

Department of Mathematics and Natural Sciences BRAC University

_____

Shamira Tabrejee

Dedicated to my Parents

Acknowledgement

The completion of my undergraduate dissertation would not have been possible without the constant guidance of Almighty in every phase of my life.

I would like to thank my **Parents** and **Saadi Khan** for everything that they have done for me and being there to support me all the time.

I am highly indebted and would like to express my appreciation to my supervisor, **Shamira Tabrejee**, Lecturer, Department of Mathematics and Natural Sciences, BRAC University for her constant supervision, guidance and immense help. Without her support, it would not have been possible to complete the study. I would also like to thank her for always being flexible and liberal with me throughout the entire period.

I wish to use this opportunity to express my sincere gratitude and thank respected **A F M Yusuf Haider**, Ph.D., Professor and Chairperson of Department of Mathematics and Natural Sciences, BRAC University, for his active cooperation and encouragement. A special word of gratitude is due to Late **Professor Dr. A. A. Ziauddin Ahmad** (Former Chairperson, Department of Mathematics and Natural Sciences) for his constant support and guidance.

I would also like to thank **Dr. Aparna Islam,** Professor, Department of Mathematics and Natural Sciences, BRAC University and **Dr. Mahboob Hossain,** Professor, Department of Mathematics and Natural Sciences, BRAC University for their enormous support and guidance.

Special gratitude and regards to **Romana Siddique**, Senior Lecturer, Department of Mathematics and Natural Sciences, BRAC University for her love, care and cooperation throughout.

Sincerely,

Sheikh Anushe

May, 2018

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Descriptions
CHIKV	Chikungunya virus
BCG	Bacillus Calmette–Guérin
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
NCBI	National Center for Biotechnology Information
MHC	Major histocompatibility complex
3D	3 Dimensional
ECSA	East-Central and South African cluster
E1	Envelope 1
Nsp	Non-structural protein
RT-PCR	Reverse transcription-polymerase chain reactions
SPPS	Solid phase peptide synthesis
MVA	Modified vaccinia Ankara
CAdVax	Complex adenovirus
MSA	Multiple sequence alignment
SMM	Stabilized matrix method
IEDB	Immune Epitope Database
SA	Structural Alphabet
GMQE	Global Model Quality Estimation
TCR	T-cell receptor
BoLA	Bovine leukocyte antigen
PVS	Protein Variability Server

Abstract

Chikungunya virus (CHIKV) is an arthropod-borne alphavirus, belonging to the Togavirus family. There is no commercial treatment or vaccine against CHIKV, despite the acute epidemics taking place in several events distributed among wide areas. In this study, we employed various computational methods to identify B-cell and T-cell epitopes from the envelope protein E1, which have the potential for vaccine development against CHIKV. By analyzing the immune parameters of the conserved sequences of E1 glycoprotein using various databases and bioinformatics tools, we identified one potential B-cell and another T-cell epitope which may be used as epitope-based peptide vaccines. Using two different B-cell epitope prediction servers, five highly similar B cell epitopes were identified from the E1 protein. Immunoinformatics analyses revealed that NTQLSEAHVEKS is a highly conserved, antigenic, surface accessible, flexible and hydrophilic B-cell epitope. Two highly conserved, non- allergenic, non-cytotoxic putative T-cell epitopes having high world population coverages were analyzed for their binding with the HLA-C 12*03 molecule. Docking simulation assay revealed that SASAKLRVL has significantly lower binding energy, which strengthened its potential as being a T-cell epitope for the epitope-based vaccine against CHIKV. This study needs more *in vivo* investigation. However, mindful of the stability and reproducibility of the immune system at choosing and acting against peptide epitopes, this study allows us to claim a B-cell and a T-cell epitope for the epitope-based peptide vaccine against the E1 protein of CHIKV with good confidence.

Chapter 1: Introduction and Literature Review

Chapter 1: Introduction

1.1 Background of the study

Chikungunya disease causes an estimated 3 million infections per year. Chikungunya is a viral disease borne by mosquitoes which was first discovered during an epidemic in southern Tanzania in 1952. It is an RNA virus belonging to the alphavirus genus of the family Togaviridae (World Health Organization, 2017). The emergence of Chikungunya was seen in Africa and since then, it has transmitted across the entire globe which caused large numbers of outbreaks that have infected millions of people in Asia, the Indian subcontinent, Europe, the Americas, and the Pacific Islands (Wahid et al., 2017). While the disease is commonly present in Africa and Asia, epidemics have been reported in Europe and the Americas since the 2000s.In 2014, more than a million suspected cases were reported. In 2014, it was said to have been occurring in Florida of the United States of America (World Health Organization, 2016).

In 2017, Pakistan continued to respond to an outbreak of Chikungunya which began in 2016 (World Health Organization, 2017). There was a recent infestation of Chikungunya in Dhaka, Bangladesh in the middle of 2017 with 196 cases confirmed from April to May (Banglajol.info, 2017). Occasionally, cases are being reported from different countries in the affected regions. The global distribution shows that CHIKV is becoming extensive at a rate which is quite alarming. CHIKV has the potential and ability to diffuse to new areas as it is a travel-associated febrile disease similar to dengue (Wahid et al., 2017).

There is no specific antiviral drug treatment for chikungunya. Treatment is aimed primarily to cure the symptoms, including the joint pain using anti-pyretics, optimal analgesics and fluids (World Health Organization, 2017). There is no commercial chikungunya vaccine yet. Vaccination is the most cost-effective way of protecting the populations prone to chikungunya in the endemic developing countries. Moreover, recent patterns in the geographic distribution of CHIKV with millions of people infected over the past years require the immediate necessity of an efficient immunization coverage rate and need further vaccine research to prevent transmission of the virus, as well as to reduce the economic problems of high healthcare costs.

Traditional vaccines have been manufactured either by using an attenuated version of the immunogen such as for Yellow Fever virus, Measles or BCG or by preparing and inactivating a pathogen as for typhoid or influenza, or by inactivating an appropriate part of it such as a toxin, in case of tetanus, and administering it in suitable quantities to trigger immunogenic response. However, there are some disadvantages to all these types of vaccines. Examples of common difficulties for attenuated vaccines are the reversion of strains to the virulent form, controlling strain properties, the presence of adventitious agents and the necessity for a cold chain. Common challenges while manufacturing inactivated vaccines are the necessity of achieving total inactivation and avoiding reactogenicity while retaining protective immunogenicity at the same time.

Peptide vaccine is an alternative approach to immunization that requires the identification of peptide epitopes on immunogens that trigger the immunogenic response and use the synthetic versions of those peptides while engineering the vaccine. Unlike traditional vaccines, peptide vaccines being entirely synthetic do not carry the risk of reversion or of incomplete inactivation and the epitopes could be selected to avoid components that give rise to unwanted side effects (World Health Organization, 1999).

1.2 Objectives of the study

The objectives of this present study to employ various bioinformatic tools in order to-

a) Evaluate the conservancy of the envelope glycoprotein E1 in all the strains of CHIKV retrieved from the NCBI database
b) Predict the antigenicity of the E1 protein
c) Predict and analyse the linear B-cell epitopes from the E1 protein
d) Analyse the conservancy of the selected B-cell epitopes in all the available strains of CHIKV
e) Determine the surface accessibility, flexibility and hydrophilicity of the selected B-cell epitopes
f) Predict and identify the linear T-cell epitopes from the E1 protein and analyse their conservancy in all the available strains of CHIKV
g) Analyse the world population coverage by the predicted T- cell epitopic peptides
h) Determine the allergenicity and toxicity of the predicted T-cell epitopes
i) Predict the 3D structures of conserved T-cell epitopes and selected MHC I allele
j) Perform the docking simulation assay of conserved T-cell epitopes with the selected MHC I allele

1.3 Literature Review

1.3.1 Chikungunya virus: an introduction

Chikungunya is an infectious disease caused by the chikungunya virus. The name "chikungunya" derives from a word in the Kimakonde language, meaning "to become contorted", and describes the stooped appearance of sufferers with joint pain (arthralgia) (World Health Organization, 2017). Common symptoms include fever, joint pain, headache, muscle pain, joint swelling, and a rash. Mostly, people get better in a week but the pain in the joints tends to last for months or longer (Cdc.gov, 2016). The risk of death is around 1 in 1,000 (Caglioti et al., 2013). The older, younger and the ones with other health problems are prone to be affected in a worse way by this disease (Cdc.gov, 2016).

The virus is transmitted through two types of mosquitoes as vectors, Aedes albopictus and Aedes aegypti (World Health Organization, 2017). They mainly bite during the day (Cdc.gov, 2016). A number of mammals and animals could serve as host for the virus including birds and rodents. We can diagnose the disease either by testing the blood for the presence of the virus's RNA or the antibodies to the virus. The symptoms could be thought to be of Dengue fever or Zika virus infection (World Health Organization, 2017).

1.3.2 Virology

Chikungunya virus (CHIKV) is a member of the alphavirus genus and belongs to the Togaviridae family. It was first discovered in Tanzania in 1953. It is an RNA virus with a positive-sense single-stranded and falls in the Group IV of Baltimore classification system. It has a genome of about 11.6kb (Weaver et al., 2014). It is a member of the Semliki Forest virus complex and is closely connected to Ross River virus, O'nyong'nyong virus, and Semliki Forest virus (Powers et al., 2001). The genome is capped in 5′ end with a 7-methylguanosine and goes through polyadenylation in the 3′ end. Its RNA encodes four nonstructural proteins (nsP1 to nsP4) and five structural proteins (C-E3-E2-6 k-E1). Genetic analysis based on the E1 envelope glycoprotein sequences showed three distinct lineages: the West African cluster, the East-Central and South African cluster (ECSA), and the Asian cluster. It is thought that CHIKV originated in West Africa, colonized other African areas, and was secondarily introduced into Asia before the 1960s. The viral situation remained stable for five decades. Strains circulating in the Western Indian Ocean in the 2000s were related to the ESCA lineage. The most notable event in CHIKV history was the occurence of an adaptive mutation, where the alanine amino acid was substituted with valine at position 226 in the E1 glycoprotein gene (E1:A226V) on an ESCA-CHIKV strain disseminating on Reunion Island after September 2005. It led the mutated CHIKV to grow independently without cholestrol and enhanced its infectivity, replication, and transmission by *Ae. albopictus,* without hampering common vectorial potential of *Ae. aegypti*. Similar genetic events took place independently in India, Gabon, and Cameroon, which suggested an evolutionary conjunction of the virus to this mosquito and subsequently leading to worldwide outbreaks and spread. To date, no variation in virulence between the different strains of CHIKV has been shown in humans (Simon et al., 2011). CHIKV can also be called an arbovirus (arthropod-borne virus), as it is transmitted by arthropods, mostly mosquitoes. Due to the serious

risk imposed by Chikungunya on humans, it was categorized as a category C priority pathogen by US National Institute of Allergy and Infectious Diseases in 2008 (Petitdemange et al., 2011).

1.3.3 Transmission

Chikungunya is usually transmitted from mosquitoes to humans. It could also spread through vertical transmission rarely, which is transmission from mother to child during pregnancy or at birth. Transmission via infected blood products and through organ donation is also theoretically possible during epidemics, though no cases have yet been reported (Burt et al., 2011).

1.3.4 Viral replication

The virus includes four nonstructural proteins (nsp1, nsp2, nsp3 and nsp4) that are involved in the replication of the virus. The nsp1 protein has the job of a cytoplasmic RNA capping enzyme, whereas the nsp2 protein has two separate domains with helicase and protease activity respectively. The nsp3 protein is essential for the synthesis of negative strand RNA. The nsp4 protein functions as an RNA-dependent RNA polymerase (Solignat et al. 2009). One-third of the genome from the 3 end has a polyadenylate tail that consists of 26S mRNA whose products include the structural proteins. The structural proteins of CHIKV are the surface glycoproteins (E1 and E2), capsid protein and small peptides (E3 and 6K) (Weaver et al. 2005; Powers and Logue 2007; Abere et al. 2012). The viral surface is covered with membrane-anchored spikes which are triplets of a heterodimer of envelope E1 and E2 glycoproteins (Brehin et al. 2008). E2 binds to cellular receptors in order to enter the host cell through receptor-mediated endocytosis. E1 contains a fusion peptide which, when exposed to the acidity of the endosome in eukaryotic cells, dissembles from E2 and triggers membrane fusion that releases nucleocapsids into the host cytoplasm, causing infection (Voss et al., 2010). The mature virion has 240 heterodimeric spikes of E2/E1, which after release, attach on the surface of the infected cell, where they are released by exocytosis in order to infect other cells (Weaver et al., 2014).

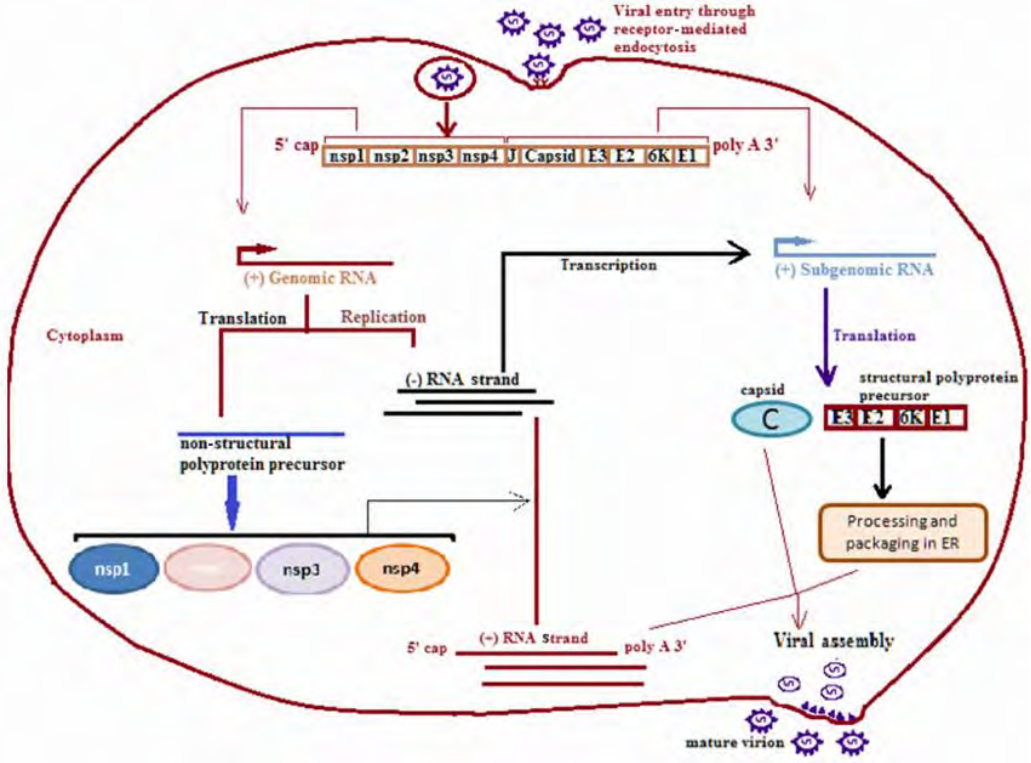


Fig 1. The CHIKV replication cycle

1.3.5 Diagnosis

Laboratory diagnosis includes the detection of the virus on early samples and/or specific antibodies of CHIKV, IgM and IgG on blood samples.

A number of quantitative reverse transcription-polymerase chain reactions (RT-PCR) are being used for CHIKV detection. There are some commercial kits which are available, sometimes with excellent sensitivity and specificity. The RNA of Chikungunya virus is detected in plasma samples within the first week after exposure to infection, and in most common cases show extremely high levels of viremia. RT-PCR can also be used to screen various fluids and tissues, including corneas or other graft tissues. CHIKV can also be collected and isolated from early samples of C6/36 or Vero cell lines. This method is only performed in laboraties maintaining Biosafety Level III protocols, and is therefore mainly used for epidemiological purposes and researches.

Anti-CHIKV antibodies can be detected in patients shortly after exposure to infection, in most cases after 5 days for IgM and only a few days later for IgG. Commercial enzyme immunoassays and immunofluorescence assays have been developed, but perform poorly when expertized.

Interpretation of serological results must be done carefully and consist of some challenges due to some reasons. Firstly, false negativity could possibly arise due to CHIKV-induced mixed cryoglobulinemia. The cross-reactivity with viruses of the Semliki Forest serocomplex could cause problems and require seroneutralization. Results should also be interpreted with caution as there is long-term persistence of anti-CHIKV IgM months after disease onset. Synchronous testing of a sample from the acute stage and a sample collected at least 3 weeks later is required to demonstrate a recent CHIKV infection in most cases. If there is any doubt assistance could be requested from an expert laboratory.

The good handling, management and care of patients affected with acute CHIKV infection are vital for public health in susceptible areas with current *Aedes* spp activity. In these areas, it is recommended promptly suspect imported or indigenous cases, use diagnostic tools adequately, isolate suspected patients, contact with the local health department rapidly, and sometimes declare cases mandatorily. The final aim is to avoid outbreaks transmitting around the new cases (Simon et al., 2011).

1.3.6 Treatment

There is no commercial vaccine to prevent or medicine to treat chikungunya virus. Treatment is aimed to get rid of the symptoms. Plenty of rest should be taken. Patients should drink fluids to prevent dehydration. Medicine such as acetaminophen (Tylenol®) or paracetamol should be taken to reduce fever and pain. Aspirin and other non-steroidal anti-inflammatory drugs should not be taken (Cdc.gov, 2016).

1.3.7 Prevention

Several measures should be taken to prevent the occurrence of chikungunya infection. Mosquito bites should mainly be prevented, especially during the day. Usage of mosquito nets is recommended. Screens should be used on windows and doors. Insect repellants could be effective but should not be used too often due do their own disadvantages (Cdc.gov, 2016).

1.3.8 Vaccines

Vaccination is the dispensation of antigenic material (a vaccine) to trigger an individual's immune system to develop adaptive immunity to a pathogen. Vaccines can prevent or ameliorate infectious disease (Liesegang, 2009; Fiore et al., 2009; Chang et al., 2009). There are many approaches of manufacturing vaccines. They are as follows.

a) Whole cell approach: Classical vaccination using whole organisms is still prevalent. They are manufactured using the whole pathogen either by inactivating (inactivated vaccine) or weakening the pathogen (live attenuated vaccine). Whole pathogen immunisations usually produce long lasting immunity. However, there are many disadvantages to it. For example, one of the main concerns is the safety as this type of vaccination could cause strong allergic or autoimmune response. Surprisingly, the allergenicity is caused not due to the presence of the pathogen but due to the contamination caused from the medium

that the microorganism was grown in (*e.g.* eggs, antibiotics). Shedding of the pathogen to the environment and infections of staff during vaccine manufacture are the other problems that have been reported. Difficulties during manufacture of some pathogens (*e.g.* malaria sporozoites), poor vaccine stability and the need for a "cold chain" are other significant disadvantages of classical vaccines. Some of the vaccines cannot even use the whole cell method (*e.g.* cancer vaccines, due to tumour being similar to healthy human cells) (Skwarczynski and Toth, 2016). There are more specific disadvantages to live attenuated and inactivated vaccines respectively.

i. Live Attenuated Vaccines: Manufacturing of vaccines by the attenuation or weakening of whole pathogens might not be ideal as the pathogen may revert back to its virulent form. One of the most significant examples of such vaccine failures was the "Lübeck disaster", when, in 1930, 67 babies among the 249 vaccinated with tuberculosis vaccine (BCG) died (Skwarczynski and Toth, 2016).

ii. Inactivated vaccines: Inactivated vaccines are produced using dead viruses and bacteria, which permits a safe method of immunising against these pathogens as they can't behave as they do normally and cause disease. However, these pathogens still consist of the identifying molecules that signal their presence to the immune system (Healthcentre.org.uk, 2016). Toxoid vaccine is a subtype of inactivated vaccine where the vaccine is manufactured using an inactivated bacterial toxin that induces immunity (Online.science.psu.edu, 2018). Some inactivated vaccines are not as efficient as their live counterparts, and while a live vaccine induces an immune response so strong that only a single dose is enough, inactivated vaccines need regular booster injections. Finally there has been an incident where the inactivation procedure has failed to safely render a virus or toxin harmless. This case involved the accidental administration of virulent smallpox, a severe condition (Healthcentre.org.uk, 2016).

b) Manufacture of vaccine using only part of the whole pathogen:

i. Subunit vaccines: Subunit vaccine utilises only the disease causing and immune response inducing part of the whole pathogen. This approach can be controlled more effectively and can be produced without the need of using the whole pathogen (*e.g.* recombinant proteins). However, they are still not completely safe and have drawbacks similar to the whole cell approach. For example, whole protein-based approach failed during the manufacture of vaccine against Group A *Streptococcus* due to potential protein-triggered autoimmunity. In addition to problems associated with protein purities, there are common stability issues, large scale protein expression difficulties, difficulties with the introduction of desired post-translational modification (*e.g.* glycosylation) into recombinant proteins and poor or undesired immune responses (inflammation, autoimmunity, *etc.*) (Skwarczynski and Toth, 2016).

ii. DNA Vaccines: This approach utilises the introduction of genetic material from particular disease causing pathogen into human cells through a specially engineered delivery system. The DNA codes for the proteins that are found on the surface of the pathogen (epitopes) and are secreted in the human cells which trigger the immune response. DNA vaccines possess a risk of potentially disrupting normal cellular processes. Secondly there is a chance that the body develops a resistance or tolerance towards the protein the vaccine introduces (Healthcentre.org.uk, 2016).

1.3.9 Peptide Vaccines

Synthetic peptide vaccines usually consist of 20–30 amino acids containing the specific epitope of an antigen related to infectious and/or chronic diseases including cancers. Peptide vaccines have no restrictions on which diseases they could target. They could work against diseases from virus infections to Alzheimer disease and even allergy (Yang and Kim, 2015). There are several more advantages to epitope-based peptide vaccines. Nowadays manufacture of peptides is simple, easily reproducible, fast and cost-effective due to recent developments in solid phase peptide synthesis (SPPS) using automatic synthesisers and implementation of microwave techniques. Chemical synthesis eliminates all the problems related to the biological contamination of the antigens.These vaccines are typically soluble in water, stable under simple storage conditions (generally does not need “cold chain”) and can be freeze-dried. Peptides can be customised very specifically according to the target. The immune responses can be directed against naturally non-immunodominant epitopes. Single peptide-based vaccine can be designed against several strains, different stages of life cycle or even different pathogens by using the multi-epitope approach. Peptide antigens are less likely to generate allergic or autoimmune responses as they lack of redundant elements (Skwarczynski and Toth, 2016).

1.3.10 Vaccine developments for Chikungunya

There are no specific drugs to cure the disease. Treatment is aimed initially at relieving the symptoms, including the joint pain. There is no commercial Chikungunya vaccine yet (World Health Organization, 2017). However, there had been many vaccine candidates proposed for CHIKV. An inactivated vaccine approach was first used for CHIKV in the 1970s when Harrison and others used formalin to inactivate CHIKV strain 15 561. A live-attenuated vaccine for CHIKV, known as 181/clone25 or TSI-GSD-218, advanced the furthest into clinical trials. This vaccine was manufactured from the AF15561 CHIKV isolate from Thailand. A live virus vectored vaccine, MV-CHIKV, was also developed using Measles virus as virus vectors. Two replication-defective vaccine vectors, modified vaccinia Ankara (MVA) and complex adenovirus

(CAdVax) have also been developed as CHIKV vaccines. Several CHIKV vaccine candidates have been developed using the DNA vaccine approach (Erasmus, Rossi and Weaver, 2016). All these vaccine candidates were proposed but however never got approved and commercialized till now.

Chapter 2: Materials and Methods

Chapter 2: Materials and methods

2.1. Protein sequence retrieval

A total of 47 sequences of the E1 glycoprotein of Chikungunya virus were retrieved from the NCBI database (Benson et al., 2009). All the sequences were extracted from the database in the FASTA format. The dates of when they were isolated were also taken into account to cover the maximum outbreaks from the past. The length of each of the 47 sequences for CHIKV was 85 amino acids.

2.2. Variability analysis of E1 glycoproteins

To analyse the level of conservation, all the retrieved sequences were aligned by using EBI-Clustal Omega program (Sievers et al., 2011) and a multiple sequence alignment (MSA) was obtained. The MSA was visualized using Jalview (Waterhouse et al., 2009). The absolute site variability in the MSA was calculated using Protein Variability Server (PVS) (Garcia-Boronat et al., 2008). PVS makes the use of several variability metrics to compute absolute variation in a MSA.

2.3. Prediction of antigenicity of the E1 protein

In order to develop a peptide vaccine, it is essential to identify those proteins which display antigenic features. A reference E1 glycoprotein having accession number, ADV91541.1, was tested for its antigenicity. VaxiJen v2.0 server (Doytchinova and Flower, 2007) and Kolaskar & Tongaonkar method (Kolaskar and Tongaonkar, 1990) were used to predict the antigenic property of the given sequence. VaxiJen makes the use of an approach independent of alignment, to predict the antigenicity of a given protein, which is entirely based on the physiochemical properties of amino acids.

2.4. Linear B-cell epitope prediction

B cell epitope prediction can be done using various software packages and each one has its own advantage and disadvantage (Blythe and Flower, 2005; Yang and Yu, 2009). To avoid the false positive results, two popular and frequently used tools were utilized. For the prediction of B-cell epitopes, sequence was inserted in BepiPred 2.0 (Larsen et al., 2006) and ABCpred (Saha and Raghava, 2006) web servers. BepiPred 2.0 (Larsen et al., 2006) server analyzes the epitopes using a hidden Markov model and a propensity scale. Seventy five percent threshold and a window length of 12 amino acids were set as parameters for the prediction of epitopes in ABCpred server (Saha and Raghava, 2006). The predicted epitopes from these two servers were scrutinized and epitopes that were commonly recognized by both the servers were selected for further analysis.

2.5. Surface accessible regions prediction

When an antibody or a cell surface receptor binds to an epitope, an immunogenic response is triggered. An ideal epitope should be accessible to an antibody or a cell surface receptor (Caoili, 2010). In order to determine the surface accessibility of the epitope, Emini surface accessible prediction tool of IEDB was used (Vita et al., 2010; Emini et al., 1985) and the whole E1 sequence was submitted to the server. Later, the accessible regions that came as results were compared with the selected B cell epitopes. Epitopes that were found to have surface accessibility were selected and analysed for conservancy.

2.6. Conservancy analysis of B cell epitope

An ideal epitope should be conserved so that it provides a wider protection against multiple strains, and at times among species. Amino acid residues that are responsible for important functions are believed to vary less even under immune pressure. Conservancy analysis of the surface accessible epitopes with all the glycoprotein sequences was analyzed by the IEDB epitope conservancy tool (Bui et al., 2007). For predicting the conservancy, the sequence identity threshold was set to at least 80 percent.

2.7 Prediction of antigenicity of the selected B-cell epitopes

Along with the whole E1 protein showing antigenic properties, the individual epitopes should also show antigenicity separately. Therefore the selected B-cell epitopes were tested for their antigenicity in the VaxiJen v2.0 server (Doytchinova and Flower, 2007). VaxiJen makes the use of an approach independent of alignment, to predict the antigenicity of a given protein, which is entirely based on the physiochemical properties of amino acids.

2.8. Prediction of flexibility and hydrophilicity of B cell epitopes

Some studies in the past have reported that hydrophilicity and flexibility of a peptide are related to its antigenicity (Novotny et al., 1986). For this, conserved epitopes were submitted to Karplus and Schulz (KS) flexibility (Karplus and Schulz, 1985), and Parker hydrophilicity prediction tools (Parker et al., 1986) for flexibility and hydrophilicity predictions respectively (Tenzer et al., 2005). The KS method utilizes normalized B-values of Ca-atoms in protein structures for predicting protein flexibility. Because it is well made, it has been widely used for analysis of protein flexibility (Sharmin and Islam, 2014; Islam et al., 2012; Oany et al., 2014).

2.9. T cell epitope prediction and conservancy analysis

T cell epitopes were identified by NetCTL server (Larsen et al., 2005, 2007) of IEDB. The threshold was set to 0.50 and the sensitivity and specificity were set to 0.89 and 0.94 respectively. NetCTL brings out T cell epitopes by combining predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I affinity. This integrated algorithm provides overall scores for epitope prediction and based on these scores, top 10 epitopes were selected for further analysis. MHC-I alleles interacting with each of the selected epitopes were determined by MHC-I prediction server (Peters and Sette, 2005) of IEDB. Stabilized matrix method (SMM) (Peters and Sette, 2005) was used for the prediction of half maximal inhibitory concentration (IC50) of peptide binding to MHC-I alleles. The cut-off value of IC50 was set 200 nM. For the analysis of binding of the epitope to the allele, all the available MHC class I alleles were selected and the peptide lengths were set to 9 amino- acids.

Alternatively, the whole protein sequences were submitted in the IEDB MHC class II binding prediction tool (Fleri et al., 2014) as MHC class II can fit much longer peptides, at times even whole proteins. IC50 values of the epitopes binding to MHC II molecules were calculated using the Stabilized Matrix Base Method (SMM). The epitopes (containing 15 amino acid residues) that interacted with highest number of alleles were again selected. SMM-align method was applied to predict good binders and the cut-off value of IC50 was set to 100 nM. The overlapping epitopes between MHC I and MHC II binding predictions which interacted with highest number of alleles and minimum of 3 alleles were finally selected to predict epitope conservancy and population coverage (Yasmin et al., 2016). Predicted T-cell epitopes were submitted to IEDB conservancy analysis tool with a sequence identity threshold of 80 percent (Bui et al., 2007).

2.10 Analysis of allergenicity of the predicted T-cell epitopes

An ideal epitope should be a non allergen. AllerTOP v. 2.0 was used to determine the allergenicity of the selected T cell epitopes. AllerTOP is the first alignment-free server for *in silico* prediction of allergens based on the main physicochemical properties of proteins. In comparison to other servers for allergen prediction, AllerTOP outperforms them with 94% sensitivity (Dimitrov et al., 2013). To validate the results further, AllergenFP was used. It uses an alignment-free method for allergenicity prediction, based on amino acid principal properties as hydrophobicity, size, relative abundance, helix and β-strand forming propensities (Dimitrov et al., 2014).

2.11 Analysis of toxicity of the predicted T-cell epitopes

The convenience of manufacture, high specificity, and high penetration of peptides make the epitope-based peptide vaccine approach promising. However, toxicity of peptides could hamper the success of peptide-based therapy. An ideal epitope should have no or less toxicity, while having high antigenicity (Vlieghe et al., 2010). Hence, to determine the toxicity of the selected T-cell epitopes, ToxinPred web server was used (Gupta et al., 2013).

2.12 Analysis of world population coverage by the predicted T-cell epitopic peptides

To become a good vaccine candidate, prediction of T-cell epitope is not enough. The predicted peptide(s) should effectively cover human population in wide areas. To analyse the world population coverage of the separate epitopes, selected T-cell epitopic sequences with the corresponding Class I and II HLA alleles were submitted to the population coverage analysis tool of IEDB by maintaining the default analysis parameters. Population coverage analysis tool predicts and calculates the fraction of individuals who are likely to respond to a given set of epitopes with known MHC restrictions. For individual population coverage, the tool analyses and finds out the following: (1) projected population coverage, (2) average number of epitope hits/ HLA combinations recognised by the populations and (3) minimum number of epitope hits/ HLA combinations recognised by 90% of the population (PC90). These calculations are performed on the basis of of HLA genotypic frequencies mindful of non-linkage disequilibrium between HLA loci. (Bui et al., 2006).

2.13 Prediction of the 3D structures of conserved T-cell epitopes and selected HLA-C 12*03

In order to perform the docking simulation assay of the epitope and allele, 3D structures both of the peptides and corresponding alleles were needed. The 3D structures of the selected peptides were constructed using the PEP-FOLD Peptide Structure Prediction server (Thevenet et al.,2012; Maupetit et al., 2009). PEP-FOLD utilises a de novo method to predict peptide structure from amino acids, ranging from 9 to 36 residues. This procedure takes place on the basis of a Hidden Markov Model derived Structural Alphabet (SA) letter coupled the predicted series of SA letters to a greedy algorithm and a coarse-grained force field (Thevenet et al., 2012). SA letters describe conformations of four consecutive residues. The best models provided by the server were chosen for the docking assay. The MHC-I allele, HLA-C 12*03, was found to interact with majority of the predicted T cell epitopes. Therefore, the 3-D model of HLA-C 12*03 was contructed using SWISS MODEL server (Biasini et al., 2014). This homology-modeling method comprises of the following four steps: (i) template selection; (ii) target template alignment; (iii) model building; and (iv) evaluation (Schwede et al., 2003). These steps can be repeated, until a satisfying model structure is achieved. The model was

chosen based on GMQE and QMEAN scores of the model and the sequence identity and coverage of the template. The HLA-C 12*03 3D model was evaluated by PROCHECK software (Laskowski et al., 1993) and ProSA web tool (Wiederstein and Sippl, 2007). Ramachandran plot (Laskowski et al., 1993), generated using the PROCHECK software analyses and predicts the stereo-chemical property of the 3-D structure by determining residue-by-residue geometry and overall structure geometry. ProSA tool provides a Z-score which is a measurement for the quality of the model (Wiederstein and Sippl, 2007).

2.14 Docking simulation assay of conserved T-cell epitopes with the HLA-C 12*03 allele

To investigate the interaction of the T cell epitopes with the corresponding MHC class I molecules, docking analysis was performed. Computer-simulated ligand docking is a powerful technique for evaluating relative binding affinity of the ligand toward its receptor (Patronov and Doytchinova, 2013). PyRx was utilised for the docking purpose (Dallakyan and Olson, 2015). PyRx consists of a docking wizard with easy-to-use user interface which makes it a valuable tool for docking of protein to ligand. PyRx is an open source software to perform virtual screening. It is a combination of several softwares such as AutoDock Vina, AutoDock 4.2, Mayavi, Open Babel, etc. PyRx uses Vina and AutoDock 4.2 as docking softwares. For this present study, AutoDock Vina was used in PyRx for the docking purpose. Autodock Vina is a very popular and highly cited open source docking program (Quiroga and Villarreal, 2016). The 3D structure of MHC class I H-2Kb molecule docked with the octapeptide epitope, KVITFIDL (PKB1), was utilised as a positive control for the docking assay. The docked structure was determined at 2.3 Å by X-ray differaction method. The binding of KVITFIDL to H-2Kb allele causes a significantly huge conformational change which is immediately identified by a T-cell receptor (TCR) (Reiser et al., 2016). This conformational change provides a good opportunity to analyse docking simulation assay. First, the epitope was separated from the allele and then it was docked with H-2Kb. Binding energy for control epitope was calculated and compared with the energy found for the predicted T-cell epitopes. The binding of the epitope to the allele was then visualized using PyMol (Seeliger and Groot, 2010).

2.15 Validation of workflow

Since the workflow used here included various computational tools developed by different platforms, so the validation of the workflow was required. A conserved B-cell epitope, SVQYHPL, was identified from the envelope protein of the reticuloendotheliosis virus through wet lab screening (Xue et al., 2012). The protein sequence of the envelope protein was retrieved from the NCBI database and fed into the workflow for to check whether the workflow can identify and qualify that as a B-cell epitope. Six T-cell epitopes were identified through wet lab from the polyprotein precursor sequence of West Nile Virus (McMurtrey et al., 2008). The

protein sequence of the polyprotein was retrieved from the NCBI database and fed into the workflow to check whether the workflow can identify and qualify those as T-cell epitopes.

Chapter 3: Results

Chapter 3: Results

3.1 Glycoprotein E1 is conserved in most pathogenic Chikungunya strains

The E1 protein sequence was of a length of 85 amino acids. To predict the degree of conservation, Multiple Sequence Alignment by Clustal Omega (Sievers et al., 2011) and protein variability analysis (Garcia-Boronat et al., 2008) were performed. From the Multiple Sequence Alignmnent, E1 glycoprotein was found to be well conserved in most of the 47 strains. The absolute variability computed by Protein Variability Server (Garcia-Boronat et al., 2008) revealed 77 fully conserved nucleotides, which comprise of more than 90% of the length of E1 protein (Fig. 2b). Two regions with amino acid positions 1-17 and 20-76 were conserved in all the strains.

Fig. 2a. Jalview Image of the MSA of all the 47 available sequences of CHIKV E1 protein

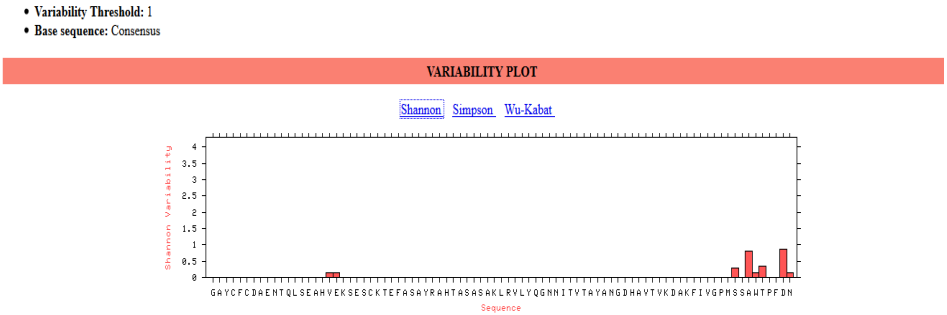


Fig. 2b. Protein variability index of E1 protein determined by using PVS server. The conservancy threshold was 1.0 in this analysis. X axis indicates the amino acid position in sequences and Y axis indicates the Shannon entropy.

3.2 E1 protein is highly antigenic

To be a vaccine candidate, a protein must be antigenic enough to provoke sufficient immune response. Evaluation of E1 protein sequence having accession number ADV91541.1 by VaxiJen server (Doytchinova and Flower, 2007) identified it as a probable antigen with a value of 0.4949. The threshold for the antigenicity of virus was 0.4. A window size of 7 amino acids was set to determine the antigenicity of the central amino acid for each of residue of E1 protein in the Kolaskar & Tongaonkar antigenicity prediction tool (Kolaskar and Tongaonkar, 1990). It revealed most of the amino acid residues out of 85 in the protein were above the threshold value of 1.00. (Fig. 3). The maximum and minimum scores were 1.153 and 0.952 for residues at position 2 and 9 respectively, with an average of 1.033.

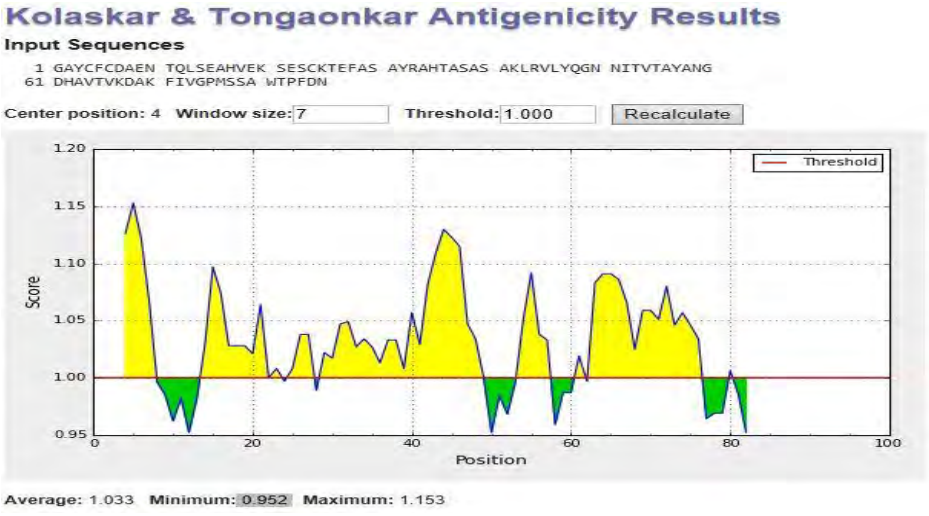


Fig 3. The E1 protein was found to be highly antigenic. The threshold is 1.00 and residues in yellow regions are antigenic in nature

3.3. NTQLSEAHVEKS, KSESCKT, AHVEKSESC, NGDHAVT and DAENTQLS were recognized as B-cell epitopes by all the prediction tools

Two different software packages were utilised for the B cell epitope prediction. The number of peptides recognized by BepiPred and ABCpred were 6 and 10, respectively (Table 1A). Epitopes recognized by the two prediction tools were taken into account for further analysis. Five antigenic epitopes of protein were found common to the four prediction tools (Table 1B). The locations of these epitopes are 10-21, 7-14, 59-65, 16-26 and 20-26.

Table 1A. B-cell epitopes recognized by BEPIPRED 2.0 and ABCPRED

No.	Sequence	Length	Start	End
1	NTQLSEAHVEKS	12	10	21
2	DAENTQLS	8	7	14
3	NGDHAVT	7	59	65
4	AHVEKSESCKT	11	16	26
5	KSESCKT	7	20	26

Table 1B: Five B-cell epitopes found common in both prediction servers along with their lengths and sequence positions

BEPIPRED 2.0	ABCPRED
SSAWT	GNNITVTAYANG
DAENTQLSEAHVEKSESCKT	NTQLSEAHVEKS
DAENTQLSEAHVEKSESCKT	YCFCDAENTQLS
NGDHAVTVK	VTAYANGDHAVT
DAENTQLSEAHVEKSESCKT	AHVEKSESCKTE
DAENTQLSEAHVEKSESCKT	KSESCKTEFASA
	SAKLRVLYQGNN
	TASASAKLRVLY
	ASAYRAHTASAS
	CKTEFASAYRAH

3.4. NTQLSEAHVEKS, KSESCKT, AHVEKSESC and DAENTQLS were found to have surface accessibility properties

At threshold cutoff 1.0, the surface accessibility of the E1 protein was determined by Emini surface accessibility prediction tool (Emini et al., 1985), and 2 peptides of length of 6 amino acids and 9 amino acids were found to have scores above the threshold. They were ENTQLS and AHVEKSESC. These peptides were compared with the selected five epitopes (Table 2 and Fig. 4). Four epitopes among those five were found to have consensus sequences with the two predicted surface accessible peptides meaning that they overlap with the Emini predicted peptides as they reside in the same region of the protein E1. They were NTQLSEAHVEKS, KSESCKT, AHVEKSESC and DAENTQLS. These four epitopes were analyzed for conservancy.

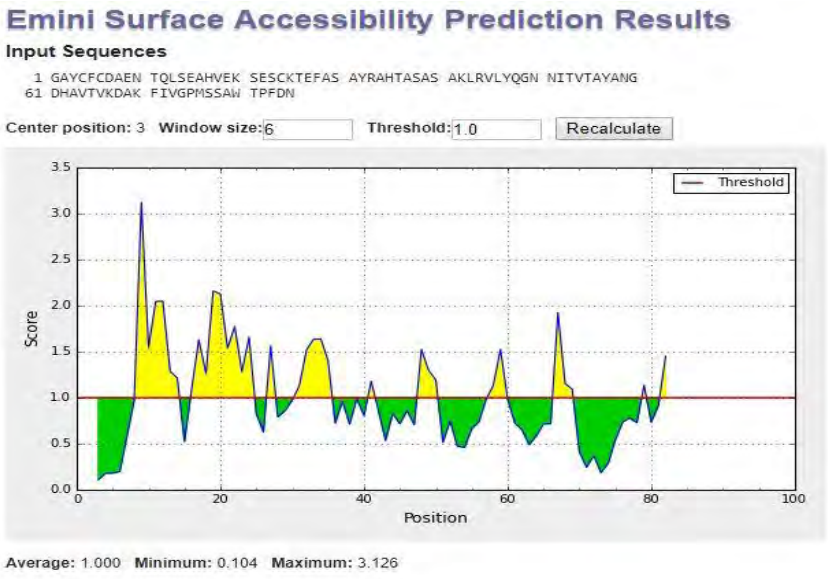


Figure 4: Surface accessibility of E1 protein. The horizontal red line indicates the surface accessibility cutoff and the yellow regions above this line are surface accessible epitopes.

Table 2: Predicted surface accessible antigenic sites by using Emini surface accessibility prediction analysis.

No.	Peptide	Position	Length
1	ENTQLS	9-14	6
2	AHVEKSESC	16-24	9

3.5 NTQLSEAHVEKS, KSESCKT, AHVEKSESC and DAENTQLS are highly similar among all sequences

The use of conserved epitopes provides broader protection across multiple strains, or even species, than epitopes derived from highly variable genomic regions. So, in an epitope-based vaccine development, an ideal epitope should be highly similar or conserved. The conservancies of the four B cell epitopes were evaluated by the IEDB conservancy analysis tool (Bui et al. 2007). Among four, all four epitopes were found to be highly similar in all E1 protein sequences (Table 3)

Table 3: Consensus sequences between the predicted B cell epitopes and Emini surface peptides along with their conservancy using IEDB conservancy analysis.

No.	Sequence	Length	Percent of protein sequence matches at identity <=100%	Minimum Identity	Maximum Identity
1	NTQLSEAHVEKS	12	97.87% (46/47)	83.33%	100 %
2	DAENTQLS	8	100 % (47/47)	100 %	100 %
3	KSESCKT	7	100 % (47/47)	100 %	100 %
4	AHVEKSESCKT	11	97.87% (46/47)	81.82%	100 %

3.6 NTQLSEAHVEKS and DAENTQLS are highly antigenic

VaxiJen server (Doytchinova and Flower, 2007) identified NTQLSEAHVEKS and DAENTQLS as probable antigens with values of 0.4374 and 1.1035 respectively. The threshold for the antigenicity was 0.4 set as default parameter. The other two epitopes did not qualify as antigens in the server so they are eliminated from further analysis.

3.7 NTQLSEAHVEKS is highly flexible and hydrophilic in nature

Flexibility and accessibility are two fundamental properties of an epitope to induce an immune response (Novotny et al., 1986). Among two highly similar epitopes, NTQLSEAHVEKS was found to be highly flexible in Karplus and Schulz flexibility prediction analysis (Karplus and Schulz, 1985) (Fig. 5), whereas DAENTQLS was found to show no flexibility. With a window size of 7 amino acids and centre position as 4, two peptides, NTQLSEA and TQLSEAH were found to have flexibility above the threshold of 0.991 (Table 4). Then, NTQLSEAHVEKS was assessed for its hydrophilicity by IEDB Parker hydrophilicity analysis (Parker et al., 1986). With a window size of 7 amino acids and centre position as 4, NTQLSEA was found to have hydrophilicity above the threshold of 3.036. Therefore NTQLSEAHVEKS epitope was found to be hydrophilic in nature (Fig 6).

Table 4. Predicted peptides with high flexibility and values above threshold of 0.991

No.	Sequence	Position	Length	Score
1	NTQLSEA	10-16	7	1.024
2	TQLSEAH	11-17	7	1.007

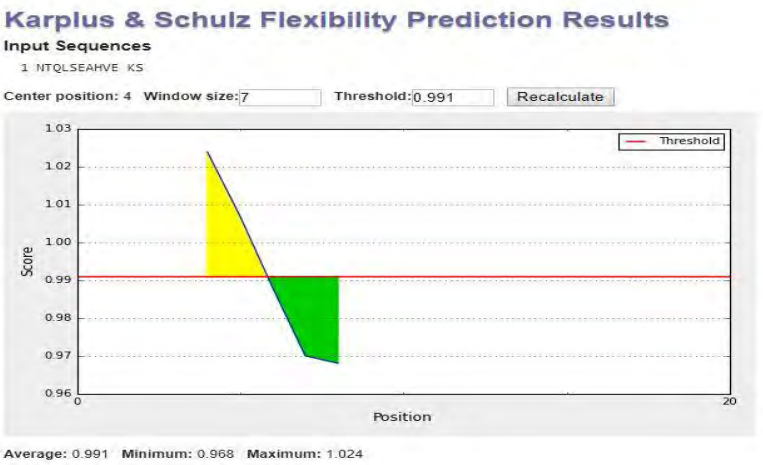


Fig 5: Flexibility of the NTQLSEAHVEKS epitope. Amino acids of this epitope were found to be above the threshold level.

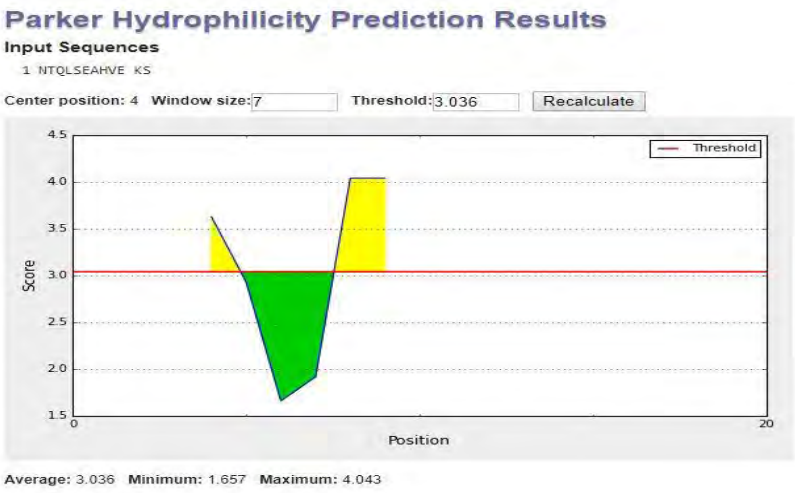


Fig 6: Hydrophilicity of the NTQLSEAHVEKS epitope. Many of the residues of this selected epitope were found to be hydrophilic in nature. Residues above the cutoff 3.036 (horizontal red line) are in the yellow region.

3.8 T-cell epitopes, ASAKLRVLY, LYQGNNITV, SASAKLRVL, KTEFASAYR, HTASASAKL, GNNITVTAY and DAKFIVGPM were found which were highly conserved

In terms of the chosen parameter settings, NetCTL server (Larsen et al., 2005) identified several potential T cell epitopes, but only 10 epitopes were chosen on the basis of high combinatorial scores and high conservancy among the available strains (Table 5). MHC class I alleles that interact with the selected 10 epitopes were determined by MHC-I binding prediction server based on IC_{50} cutoff values of 200 (Peters and Sette, 2005; Peters et al., 2003) (Table 6). On the other hand, the whole sequence of the E1 protein was submitted in the IEDB MHC class II binding prediction tool (Fleri et al., 2014). The peptides of length of 15 amino acids were selected on the basis of the IC50 cut off value of 100 and on the basis of the highest number of alleles they interacted with. The overlapping seven T-cell epitopes between MHC I and MHC II binding predictions which interacted with highest number of alleles (minimum 3 alleles) were finally selected to predict allergenicity and toxicity (Yasmin et al., 2016) (Table 7)

Table 5. 10 epitopes were selected from NetCTL based on the high combatorial scores and their conservancy.

No.	Epitope sequence	AA position	Threshold	Supertype	Percent of protein sequence matches at identity <=100%
1.	ASAKLRVLY	39-47	2.3235	A1	100.00% (47/47)
2.	KSESCKTEF	20-28	1.2177	B58	100.00% (47/47)
3.	LYQGNNITV	46-54	1.1382	A24	100.00% (47/47)
4.	HTASASAKL	35-43	1.0939	B39	100.00% (47/47)
5.	GNNITVTAY	49-57	1.0476	B62	100.00% (47/47)
6.	SASAKLRVL	38-46	1.0328	B7	100.00% (47/47)
7.	FCDAENTQL	5-18	0.9236	B39	100.00% (47/47)
8.	KTEFASAYR	25-33	0.8863	A3	100.00% (47/47)
9.	QLSEAHVEK	12-20	0.8455	A3	97.87% (46/47)
10.	DAKFIVGPM	68-76	0.7999	A26	100.00% (47/47)

Table 6. Predicted T-cell epitopes along with their interacting MHC-I alleles

Epitope Sequence	Allele	IC50 (nM)
ASAKLRVLY	HLA-C*12:03	47.67
	HLA-C*03:03	56.05
	HLA-A*30:02	98.05
	HLA-A*29:02	179.37
KSESCKTEF	HLA-C*12:03	60.15
LYQGNNITV	HLA-C*14:02	5.83
	HLA-C*12:03	10.97
	HLA-C*03:03	182.23
HTASASAKL	HLA-B*15:02	19.73
	HLA-C*03:03	20.40
	HLA-A*68:02	60.48
	HLA-C*15:02	93.42
	HLA-C*12:03	128.89
GNNITVTAY	HLA-C*12:03	38
	HLA-C*14:02	19
	HLA-B*15:02	16
	HLA-C*03:03	39
SASAKLRVL	HLA-C*03:03	4.55
	HLA-C*15:02	24.86
	HLA-C*12:03	33.21
	HLA-C*07:01	179.54
	HLA-C*06:02	187.85
	HLA-B*15:02	188.46
FCDAENTQL	HLA-C*12:03	17.55
	HLA-C*08:02	55.74
	HLA-C*03:03	100.60
	HLA-B*15:02	101.91
KTEFASAYR	HLA-A*31:01	16.42
	HLA-C*12:03	38.48
	HLA-C*15:02	55.39
	HLA-A*68:01	68.94
	HLA-C*14:02	91.75
	HLA-C*03:03	155.10
QLSEAHVEK	HLA-C*03:03	48.04
	HLA-C*12:03	55.87
DAKFIVGPM	HLA-C*12:03	2.93
	HLA-C*03:03	8.60
	HLA-C*14:02	162.79
	HLA-B*15:02	198.26

Table 7. Seven shortlisted T-cell overlapping epitopes between MHC I and MHC II binding predictions

Epitope	MHC- I Allele	Peptide Sequence	MHC-II Allele
ASAKLRVLY	HLA-C*12:03 HLA-C*03:03 HLA-A*30:02 HLA-A*29:02	RAHTASASAKLRVLY AHTASASAKLRVLYQ	HLA-DRB5*01:01
		RAHTASASAKLRVLY	HLA-DRB1*07:01
LYQGNNITV	HLA-C*14:02 HLA-C*12:03 HLA-C*03:03	KLRVLYQGNNITVTA SAKLRVLYQGNNITV AKLRVLYQGNNITVT LRVLYQGNNITVTAY	HLA-DRB1*13:02
HTASASAKL	HLA-B*15:02 HLA-C*03:03 HLA-A*68:02 HLA-C*15:02 HLA-C*12:03	ASAYRAHTASASAKL SAYRAHTASASAKLR AYRAHTASASAKLRV YRAHTASASAKLRVL	HLA-DRB1*01:01
		ASAYRAHTASASAKL SAYRAHTASASAKLR	HLA-DRB1*09:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AHTASASAKLRVLYQ SAYRAHTASASAKLR AYRAHTASASAKLRV	HLA-DRB5*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AYRAHTASASAKLRV	HLA-DRB1*07:01
		ASAYRAHTASASAKL	HLA-DRB1*04:01
GNNITVTAY	HLA-C*12:03 HLA-C*14:02 HLA-B*15:02 HLA-C*03:03	LRVLYQGNNITVTAY	HLA-DRB1*13:02
SASAKLRVL	HLA-C*03:03 HLA-C*15:02 HLA-C*12:03 HLA-C*07:01 HLA-C*06:02 HLA-B*15:02	YRAHTASASAKLRVL	HLA-DRB1*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY AHTASASAKLRVLYQ	HLA-DRB5*01:01
		YRAHTASASAKLRVL RAHTASASAKLRVLY	HLA-DRB1*07:01
KTEFASAYR	HLA-A*31:01 HLA-C*12:03 HLA-C*15:02 HLA-A*68:01 HLA-C*14:02 HLA-C*03:03	KTEFASAYRAHTASA CKTEFASAYRAHTAS SCKTEFASAYRAHTA ESCKTEFASAYRAHT	HLA-DRB1*11:01
DAKFIVGPM	HLA-C*12:03 HLA-C*03:03 HLA-C*14:02 HLA-B*15:02	KDAKFIVGPMSSAWT DAKFIVGPMSSAWTP VKDAKFIVGPMSSAW TVKDAKFIVGPMSSA	HLA-DRB1*01:01

3.9 ASAKLRVLY, LYQGNNITV, SASAKLRVL and KTEFASAYR were selected as non-allergens out of the seven T-cell epitopes

Two servers, AllerTOP v. 2.0 (Dimitrov et al., 2013) and AllergenFP were used to predict the allergenicity of the seven T-cell epitopes. Out of the seven, ASAKLRVLY, LYQGNNITV, SASAKLRVL and KTEFASAYR were found to be non-allergens. HTASASAKL, GNNITVTAY and DAKFIVGPM were found to be allergens and so were eliminated from the further analysis.

3.10 ASAKLRVLY, LYQGNNITV, SASAKLRVL and KTEFASAYR were found to be non-toxic

The toxicity of the selected four T-cell epitopes were analysed by the ToxinPred server (Gupta et al., 2013). All of the selected four T-cell epitopes, ASAKLRVLY, LYQGNNITV, SASAKLRVL and KTEFASAYR, were found to be non-toxic to cell proving their potential as candidate vaccines.

3.11 SASAKLRVL and KTEFASAYR were selected with the highest world population coverages

The four selected T-cell epitopes were analysed for their world population coverage in the Population Coverage analysis tool of IEDB (Bui et al., 2006). Population coverage by the most probable epitopes varied between 26 % – 65 % when MHC class I and II alleles were considered (Table 8).Top two epitopes, SASAKLRVL and KTEFASAYR, with the highest world population coverages of 65.89% and 40.04% were chosen for further generation of 3D models by PEPFOLD server.

Table 8: Four T-cell epitopes with their world population coverages

Epitope Sequence	World Population Coverage
ASAKLRVLY	37.15%
LYQGNNITV	26.05%
SASAKLRVL	65.89%
KTEFASAYR	40.04%

3.12 3D structures of the predicted epitope peptides and HLA-C 12*03 were predicted and validated

In order to perform the molecular docking simulation assay, the 3-D structure of both the epitopes and MHC molecule are required. PEPFOLD Peptide Structure Prediction server was used for generating 3-D structure of two T cell epitopes SASAKLRVL and KTEFASAYR (Fig. 7A and B). As both epitopes were predicted to interact with HLA-C 12*03 allele, the 3-D structure of the allele was generated in the SWISS MODEL server by homology modeling (Biasini et al., 2014) (Fig. 7C). Based on the GQME and QMEAN4 scores of 0.74 and 1.02 respectively, best model was selected for which the template was 5w69.3.A. The target and template sequences showed 98.19% sequence identity. The model was further validated using Ramachandran plot and Z-score. Ramachandran plot generated by Procheck software showed that 93.2% residues were in the favorable region (Fig. 7D). A good model would be expected to have over 90% in the most favoured regions. The G- factor was found to be -0.04 which is normal as it is not below -0.5. To check whether our 3D structure is within the range of scores normally found for native proteins of similar sizes, ProSAz-score was calculated (Wiederstein and Sippl, 2007). Z- score which provides an overall model quality, was -9.27 that confirmed the good quality of the generated model (Parvege et al., 2016). (Fig. 7E).

Fig 7A. Predicted 3D Structure of KTEFASAYR

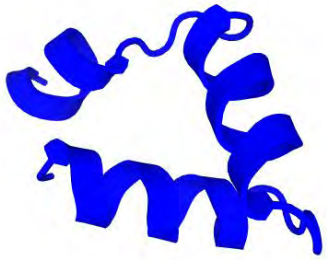


Fig 7B. Predicted 3D Structure of SASAKLRVL

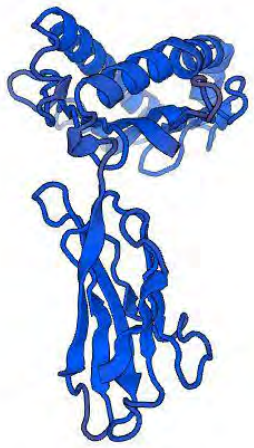
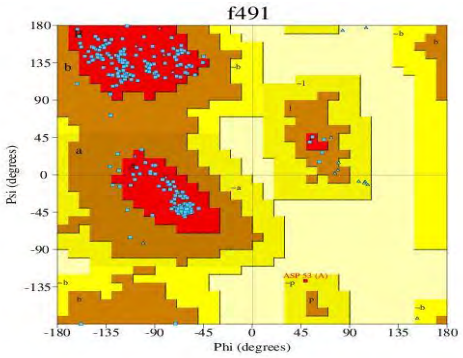


Fig 7C. Predicted 3D structure of the HLA-C 12*03 allele



1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	221	93.2%
Additional allowed regions	[a,b,l,p]	15	6.3%
Generously allowed regions	[~a,~b,~l,~p]	1	0.4%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		237	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		20	
Proline residues		14	
Total number of residues		273	

2. G-Factors

Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	-0.07	
Chi1-chi2 distribution	0.01	
Chi1 only	-0.11	
Chi3 & chi4	0.43	
Omega	-0.57*	
		-0.14
Main-chain covalent forces:-		
Main-chain bond lengths	0.29	
Main-chain bond angles	-0.06	
		0.09
OVERALL AVERAGE		-0.04

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.

Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

Fig 7D. Ramachandran plot of HLA-C 12*03 along with statistics showing residues in the most favorable and disallowed regions and the G-factor for the model

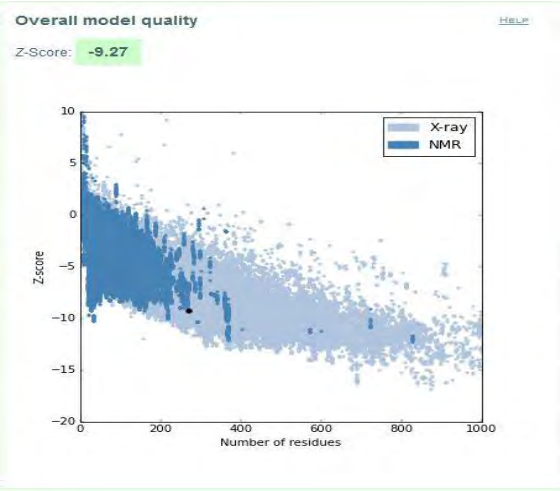


Fig 7E. Z-score for quality of the 3D structure HLA-C 12*03

3.13. T cell epitope SASAKLRVL bound with the HLA-C 12*03 allele with a binding energy of -14.1 kcal/mol

At first binding models for both of the conserved T cell epitopes with the HLA-C 12*03 were generated using the AutoDOCK Vina tool in PyRx (Dallakyan and Olson, 2015). Free energy of binding was estimated by Autodock Vina tool according to the following equation:

Free energy of binding= Intermol energy+Internal energy+Torsional energy-Unbound energy

In docking analysis, intermolecular forces included Van der Waals forces, Hydrogen bonds, solvation and electrostatic energy. The energy values calculated for binding both of the epitopes, SASAKLRVL and KTEFASAYR, to the binding groove of the HLA-C 12*03 were -14.1 kcal/mol and -13.9 kcal/mol, respectively.

Control peptide "KVITFIDL" bound to the H-2kb allele with the binding energy of -14.3 kcal/mol. As lower binding energy favors the formation of stable interaction, we can expect that SASAKLRVL will interact with MHC-I molecules *in vivo* readily.

The binding of the epitope SASAKLRVL to the HLA-C 12*03 allele was then visualized using PyMol (Fig. 8A).

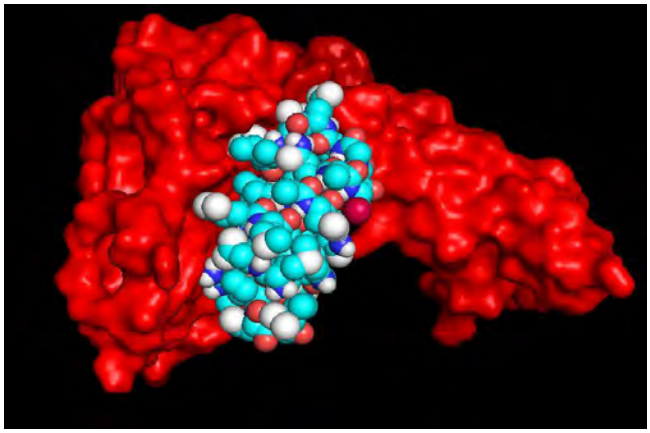


Fig 8A

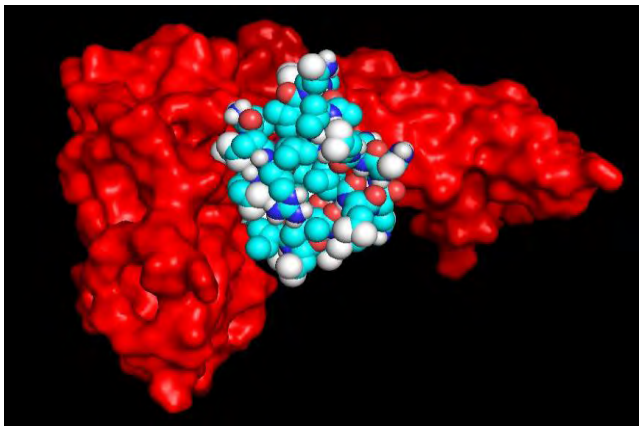


Fig 8B

Fig 8. Docking simulation assays of SASAKLRVL (Fig 8A) and KTEFASAYR (Fig 8B) to the HLA-C 12*03 allele

3.14 Designed workflow synchronizes with experimental results

The workflow, designed to predict epitopes, used here was validated by positive controls. Xue et al. (2012) mapped a linear B-cell epitope, SVQYHPL, located in the envelope protein of Reticuloendotheliosis virus. When the sequence of the envelope protein was fed into the workflow for the B-cell epitope, the same B-cell epitope, SVQYHPL, was successfully identified. McMurtrey et al. (2008) mapped six linear T-cell epitopes, SVGGVFTSV, RLDDDGNFQL, YTMDGEYRL, SLFGQRIEV, SLTSINVQA and ATWAENIQV in the polyprotein sequence of West Nile virus NY99. When the sequence of the polyprotein was fed into the workflow, it identified all T-cell epitopes except RLDDDGNFQL.

Chapter 4: Discussion and conclusion

Chapter 4: Discussion

Chikungunya virus causes 3 million infections per year. In recent years, the virus has changed from being unknown to becoming a worldwide public health hazard affecting millions of people throughout the tropical and sub-tropical world. It has also become a frequent cause of travel-associated febrile illness. Since 2004, huge urban epidemics producing considerable morbidity in a broadening geographical area have occurred throughout the topical and sub-tropical world (Peterson and Powers et al., 2016). The global distribution shows that CHIKV is becoming extensive at a rate which is quite alarming. CHIKV has the potential and ability to diffuse to new areas as it is a travel-associated febrile disease (Wahid et al., 2017). There is no commercial chikungunya vaccine yet and there is no antiviral drug treatment for it. Therefore the development of vaccine for Chikungunya has become quite vital (World Health Organization, 2017). A vaccine typically contains an agent that is similar to a disease-causing microorganism, and is usually made from weakened or killed forms of the microbe, its toxins or one of its surface proteins (World Health Organization, n.d). Vaccines made from whole cells of organisms or parts of them have many disadvantages to them. The safety of this form of vaccination is one of the major concerns as it may cause autoimmune or strong allergic responses. Interestingly, allergic shock is often caused not by the presence of pathogen itself but rather, it is caused by contamination from the medium on which microorganism was grown (*e.g.* eggs, antibiotics). Attenuation or inactivation of such vaccines might not be perfect as the pathogen may revert back to its virulent state. Subunit vaccines being more controllable than whole cell vaccines are still not perfectly safe, and cause side effects and production difficulties similar to whole pathogen strategies, like manufacturing difficulties and poor or undesired immune response (Skwarczynski and Toth, 2016).

Therefore, only using the minimal antigenic epitopes to induce the desired immune response seems like the most sensible and safest method to develop vaccines. The synthetic epitope-based peptide vaccines may have such a potential. Epitopes have become suitable vaccine candidates as it is comparatively easier to construct and produce and there is absence of infectious potential and chemical stability (Naz and Dabir, 2007; Purcell et al., 2007). They have the potential to not only render protection against the disease but also behave as the therapeutic tool to treat them. An epitope, also known as antigenic determinant, is that part of an antigen which is recognized by the immune system, specifically by antibodies, B cells, or T cells. The epitope is the specific part of the antigen to which an antibody binds (Huang and Honda, 2006). The use of only a minimal component of microbes which is able to provide long term protection against the pathogen is becoming the trend in vaccine development. Thus, fully synthetic epitope-based peptide vaccines are the potential future of vaccine development. This type of vaccine may not replace the recent tendancy in development of recombinant protein-based vaccines in the near future; however, exciting development in peptide-based immunogens is already occurring (Skwarczynski and Toth, 2016).

Though sequences from pathogens provide several potential vaccine candidates, it can be assumed that only one in 100 to 200 peptides binds to a particular MHC in real (Yewdell and Bennink, 1999). Therefore, a good computational resolving method could significantly narrow down the number of peptides that have to be manufactured and tested (De Groot et al., 2002). With the creation of various bioinformatics tools and the availability of massive sequence data, epitope-based peptide vaccine design against highly conserved antigenic protein has become very convenient. In the present study, we utilised various bioinformatics tools to find out the epitopes highly conserved in all strains of Chikungunya virus. Computational determination of conserved epitope for the manufacture of peptide vaccines is not only less time consuming, but is also very economic (Parvege et al., 2016). The ground behind the epitope-based vaccine is the chemical synthesis of identified B-cell and T-cell epitopes, which are immunodominant and can trigger specific immune responses (Patronov and Doytchinova, 2013). T cells (thymus cells) and B cells (bone marrow- or bursa-derived cells) are the main cellular components of the adaptive immune response. T cells play roles in cell-mediated immunity, whereas B cells are primarily involved in humoral immunity (relating to antibodies) (Janeway et al., 1999). In our study, epitopes inducing positive and desirable cell mediated immune responses for both B- and T-cells were predicted. The identification of B-cell epitopes is rather important for immunodetection and immunotherapeutic applications. This is because, an epitope being the minimal immune unit, is potent enough to generate a strong humoral immune response with no adverse side effects to human body (Sun et al., 2013). Most of the vaccines are developed based on B cell immunity, but presently, T cell epitope-based vaccine has been drawn much more interest when it comes to vaccines against viruses, as CD8+ T-cells elicit strong immune responses in the host against viral infections (Shrestha and Diamond, 2004). For the design of novel vaccines, we should have perfect knowledge about the antigenic regions that are recognized not only by B cells, but also by $CD8^+$ cytolytic T cells and $CD4^+$ T helper cells of the respective host (Gerner et al., 2009). For these reasons, we identified B-cell epitopes as well as T-cell epitopes.

We predicted one highly conserved B-cell epitope and another T-cell epitope which may be used to develop universal vaccines to prevent all types Chikungunya infection. We designed the vaccine against CHIKV with *in-silico* methods taking envelope glycoprotein E1 into consideration. The reason for choosing envelope glycoprotein is because of its functions as these proteins are important for viral attachment to the host cell surface, and also for facilitating immune response in the host cell. The protein is found on the viral capsid and virion membrane. To identify the best probable B- and T-cell peptides which could be used to design an effective vaccine, different *in-silico* approaches were taken into consideration in this study. It was stated that E1-targeted antibodies are more likely to cross-react with other alphaviruses and thus are nonspecific compared with E2 protein targeted antibodies (Islam et al., 2012). However, we have analysed the conservancy of the E1 protein and it was 90 % conserved among all the 49 strains of E1 protein found in the NCBI database, and therefore epitopes that have been predicted are unique to the CHIKV E1 protein. Our main purpose to develop an epitope-based peptide

vaccines against CHIKV is expected to be served by our study but additionally if it also works against other alphaviruses, it should not be the cause of any problem.

Surface glycoproteins are prone to undergoing frequent mutations to defeat host defense. First step in an epitope-based vaccine design is to identify the sequences which are highly conserved. The degree of variability or similarity of specific proteins provides important insights about its evolution, structure function, and immunology. Hence, at first, we determined the degree of conservation of E1 protein. MSA and absolute variability site analysis demonstrated that E1 protein is highly conserved in all strains of CHIKV. To be a vaccine candidate, a protein must be antigenic enough to provoke sufficient immune response. The antigenicity of the E1 protein was analysed by the VaxiJen server (Doytchinova and Flower, 2007) and Kolaskar & Tongaonkar antigenicity prediction tool (Kolaskar and Tongaonkar, 1990) and the sequence was found to be highly antigenic. Several B cell epitope prediction software packages are currently available. Each software has its own set of data and employs a particular procedure for prediction (Blythe and Flower, 2005; Yang and Yu, 2009). Several B cell epitope prediction methods have been developed in recent years, but performances of prediction of these methods are still not so great (Greenbaum et al., 2007). To avoid the false positive results, we utilized two different B-cell epitope prediction methods and only the five epitopes which were found common to all prediction were selected for further analysis. The five epitopes were NTQLSEAHVEKS, KSESCKT, AHVEKSESC, NGDHAVT and DAENTQLS. B-cell epitopes must be accessible enough to bind to antibodies for generating immune responses (Caoili, 2010). Here, we identified the two surface accessible peptides, ENTQLS and AHVEKSESC, and compared these peptides with the five selected B-cell epitopes. Four of the five B-cell epitopes, NTQLSEAHVEKS, KSESCKT, AHVEKSESC and DAENTQLS, overlap with the two predicted surface accessible peptides as they are present in the same region of E1 protein. The use of conserved epitopes provides broader protection across multiple strains, or even species, than epitopes which are derived from highly variable genomic regions. So, in an epitope-based vaccine development, an ideal epitope should be highly similar or conserved. The conservancy of the selected four B-cell epitopes were analysed in the IEDB conservancy analysis tool (Bui et al., 2007) and all the epitopes were found to be highly conserved. Two of the epitopes, DAENTQLS and KSESCKT, are 100% conserved in all the 47 strains and the other two, NTQLSEAHVEKS and AHVEKSESC, are conserved in 46 strains with 97.87% conservancy. Along with the whole E1 protein showing antigenic properties, the individual epitopes should also show antigenicity separately. Therefore the selected four B-cell epitopes were tested for their antigenicity in the VaxiJen v2.0 server (Doytchinova and Flower, 2007). NTQLSEAHVEKS and DAENTQLS were found to be highly antigenic with antigenic scores of 0.4374 and 1.1035 respectively, which were above the threshold value of 0.4. The other two epitopes did not qualify as antigens and therefore were eliminated from further analysis.

Flexibility and accessibility are two fundamental properties of an epitope to generate an immune response (Novotny et al., 1986). Between NTQLSEAHVEKS and DAENTQLS, NTQLSEAHVEKS was found to be highly flexible in Karplus and Schulz flexibility prediction analysis (Karplus and Schulz, 1985) and highly hydrophilic in the Parker hydrophilicity prediction tools (Parker et al., 1986).

Therefore, through all the analyses, we can say that NTQLSEAHVEKS can qualify as being a B-cell epitope for the construction of epitope-based peptide vaccine for Chikungunya disease. In the NTQLSEAHVEKS epitope, the single letter data-base code N stands for amino acid Asparagine, T for Threonine, Q for Glutamine, L for Leucine, S for Serine, E for Glutamic acid, A for Alanine, H for Histidine, V for Valine and K for Lysine.

The T-cell epitope prediction server, NetCTL, identified several CD8+ epitopes and based on their prediction scores and their conservancy, top ten epitopes were selected for further analysis. For the cytotoxic T-cells to be activated it is required that epitopes bound to Major Histocompatibility (MHC) class I molecules are recognized (Sette et al., 2001). The job of MHC molecules is to bind epitopes derived from pathogens and present them on the cell surface to be recognised by the appropriate T-cells. The effects are almost always detrimental to the pathogen. Virus-infected cells are killed, macrophages are activated to kill bacteria residing in their intracellular vesicles and B-cells are stimulated to create antibodies that abolish or neutralize extracellular pathogens (Janeway et al., 1999). T-cells recognize their particular antigen in terms of major histocompatibility complex (MHC) molecules and the genes which code for these molecules are highly variable within an outbred population of a particular species. For instance, in cattles, the bovine leukocyte antigen (BoLA) MHC class II molecules that are expressed are DR and DQ. Three of the genetic loci which code for these molecules are polymorphic. Till now, 106 DRB3, 46 DQA, and 52 DQB alleles have been reported. Therefore, the identification of T-cell epitopes is crucial to the MHC alleles available in a particular species, an event called MHC restriction (Gerner et al., 2009).

To identify binding of MHC-I alleles to each of the ten selected epitopes, a sophisticated machine learning method called stabilized matrix method (SMM) (Peters and Sette, 2005) was employed along with the binding affinity, IC_{50}. MHC I alleles interacting with the selected T-cell epitopes were predicted based on IC_{50} cutoff values of 200 in the IEDB MHC class I binding prediction tool (Peters and Sette, 2005; Peters et al., 2003). On the other hand MHC II alleles interacting with epitopes from the whole E1 protein were predicted through the IEDB MHC class II binding prediction tool based on IC_{50} cutoff values of 100 (Fleri et al., 2014). The IC50 or the half maximal inhibitory concentration is the concentration of an inhibitor where the response (or binding) is reduced by half (GraphPad, n.d.). The overlapping seven T-cell epitopes between MHC I and MHC II binding predictions which interacted with highest number of alleles (minimum 3 alleles) were finally selected to predict allergenicity and

toxicity (Yasmin et al., 2016). They were ASAKLRVLY, LYQGNNITV, SASAKLRVL, KTEFASAYR, HTASASAKL, GNNITVTAY and DAKFIVGPM.

We needed to make sure that the selected seven T-cell epitopes did not show any allergenicity and thus their allergenicity were predicted by two servers AllerTOP v. 2.0 (Dimitrov et al., 2013) and AllergenFP (Dimitrov et al., 2014). Out of the seven epitopes, ASAKLRVLY, LYQGNNITV, SASAKLRVL and KTEFASAYR were found to be non-allergens. HTASASAKL, GNNITVTAY and DAKFIVGPM were found to be allergens and so were eliminated from the further analysis.

Because peptides have high specificity, high penetration and possess ease of manufacturing , peptide vaccines have emerged as promising therapeutic tools for many fatal diseases (Thundimadathil, 2012). But, one of the limitations that disrupt the efficacy of peptide-based therapies is their toxicity. Hence, toxicity for the four selected epitopes was assessed by ToxinPred tool (Gupta et al., 2013) and all the four selected epitopes were found to be non-cytotoxic.

An ideal T-cell epitope should effectively cover human population to see if the MHC alleles that the epitope interacts with, are present in the population around the world. To find out the population coverage of the individual epitopes, predicted epitopic sequences with the corresponding Class I and II HLA alleles were submitted to the population coverage analysis tool of IEDB (Bui et al., 2006). Top two epitopes, SASAKLRVL and KTEFASAYR, with the highest world population coverages of 65.89% and 40.04% were chosen for further generation of 3D models by PEPFOLD server. 60% and higher of world population coverage for both MHC I and MHC II alleles is a very good value (Yasmin et al., 2016).

For the molecular docking simulation assay, the 3-D structure of both the epitopes and MHC molecule are required. PEPFOLD Peptide Structure Prediction server was used for generating 3-D structure of two T cell epitopes SASAKLRVL and KTEFASAYR. As both epitopes were predicted to interact with HLA-C 12*03 allele, the 3-D structure of the allele was generated by homology modeling in the SWISS MODEL server (Biasini et al., 2014).

Ligand docking simulated on computer is a fast and powerful technique to evaluate the relative binding affinity of the ligand towards its receptor. Computer- simulated molecular docking was performed by AutoDock Vina in PyRx to investigate the intermolecular interactions between the epitopes and the HLA-C 12*03 allele. From the outcomes of docking simulation assay, it was found that the binding energy of the SASAKLRVL epitope was -14.1 kcal/mol, which almost resembles the binding energy of the control peptide, KVITFIDL which was -14.3 kcal/mol . A positive control is always needed in an experiment to compare and validate the results with. Therefore a positive control of KVITFIDL

(PKB1) complexed with H-2Kb molecule (Reiser et al., 2016) was used to validate the docking of the T-cell epitopes to the allele. As lower binding energy favors stable intermolecular interaction, SASAKLRVL could be a potential T-cell epitope. In the SASAKLRVL epitope, the single letter data-base code S stands for amino acid Serine, A for Alanine, K for Lysine, L for Leucine, R for Arginine and V for Valine.

Although bioinformatics has restructured the research in fields of biomedical science, some doubts still exist in the scientific community whether the prediction concords with the *in-vivo* experimental result (Parvege et al., 2016) . Therefore, in-vivo experimental results for B-cell and T-cell epitopes were fed into our workflow and the validity of our *in-silico* procedure and outcomes was confirmed.

Stability and antigenicity of our peptide vaccine can be further enriched attaching these epitopes to adjuvants (Olesen et al., 2009). Adjuvants (Latin word *adjuvare*, meaning "to help or aid") were first coined by Ramon as "substances used in combination with a specific antigen that produced a more robust immune response than the antigen alone" (Ramon, 1924). Some adjuvants licensed for human use are Alum, MF59, AS03 and virosomes (Lee and Nguyen, 2015). To defend the concerns regarding actual immunogenicity, stability, efficacy and delivery method of these epitopes inside human bodies, both *in vitro* and *in vivo* experiments are essential (Parvege et al., 2016).

Conclusion

This study determines potential epitopes for designing an epitope-based peptide universal vaccine for all pathogenic strains of Chikungunya Virus. NTQLSEAHVEKS is predicted to be a B-cell epitope and SASAKLRVL is predicted as a T-cell epitope.

References

Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J. and Sayers, E. (2009). GenBank. *Nucleic Acids Research*, 37(Database), pp.D26-D31.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research, 42*(W1).

Blythe, M. and Flower, D. (2009). Benchmarking B cell epitope prediction: Underperformance of existing methods. *Protein Science*, 14(1), pp.246-248.

Bréhin, A., Rubrecht, L., Navarro-Sanchez, M., Maréchal, V., Frenkiel, M., Lapalud, P., Laune, D., Sall, A. and Desprès, P. (2008). Production and characterization of mouse monoclonal antibodies reactive to Chikungunya envelope E2 glycoprotein. *Virology*, 371(1), pp.185-195.

Bui, H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics, 7*(1), 153.

Bui, H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics, 8*(1), 361.

Burt, F., Rolph, M., Rulli, N., Mahalingam, S. and Heise, M. (2012). Chikungunya: a re-emerging virus. *The Lancet*, 379(9816), pp.662-671.

Caglioti C, e. (2018). *Chikungunya virus infection: an overview. - PubMed - NCBI*. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/pubmed/23912863 [Accessed 29 Apr. 2018].

Caoili, S. (2010). B-cell epitope prediction for peptide-based vaccine design.*Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology - BCB '10.*

Cdc.gov. (2018). *Symptoms, Diagnosis, & Treatment | Chikungunya virus | CDC*. [online] Available at: https://www.cdc.gov/chikungunya/symptoms/index.html [Accessed 29 Apr. 2018].

Chang, Yuli; Brewer, Noel T.; Rinas, Allen C.; Schmitt, Karla; Smith, Jennifer S. (July 2009). Evaluating the impact of human papillomavirus vaccines. *Vaccin*e. 27 (32): 4355–62.

Dallakyan, S., & Olson, A. J. (2015). Small-Molecule Library Screening by Docking with PyRx. *Methods in Molecular Biology Chemical Biology*, 243-250

Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics, 14*(Suppl 6).

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2013). AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics, 30*(6), 846-851.

Doytchinova, I. and Flower, D. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1), p.4.

Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*, *55*(3), 836–839.

Erasmus, J., Rossi, S. and Weaver, S. (2016). Development of Vaccines for Chikungunya Fever. *Journal of Infectious Diseases*, 214(suppl 5), pp.S488-S496.

Fiore, Anthony E.; Bridges, Carolyn B.; Cox, Nancy J. (2009). Seasonal influenza vaccines. *Current Topics in Microbiology and Immunology*. 333: 43–82.

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Frontiers in Immunology, 8*.

Garcia-Boronat, M., Diez-Rivero, C., Reinherz, E. and Reche, P. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery.*Nucleic Acids Research*, 36(Web Server), pp.W35-W41.

Gerner, W., Hammer, S. E., Wiesmuller, K., & Saalmuller, A. (2009). Identification of Major Histocompatibility Complex Restriction and Anchor Residues of Foot-and-Mouth Disease Virus-Derived Bovine T-Cell Epitopes. *Journal of Virology, 83*(9), 4039-4050.

Groot, A. S., Sbai, H., Aubin, C. S., Mcmurry, J., & Martin, W. (2002). Immuno-informatics: Mining genomes for vaccine components. *Immunology and Cell Biology, 80*(3), 255-269.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE, 8*(9).

Healthcentre.org.uk. (2018). *Guide to Vaccines | Types, Administration & Protection Against Disease & Illness*. [online] Available at: http://www.healthcentre.org.uk/vaccine/ [Accessed 29 Apr. 2018].

Huang, J., & Honda, W. (2006). CED: A conformational epitope database. *BMC Immunology, 7*(1), 7.

Islam, R., Sakib, M. S., & Zaman, A. (2012). A computational assay to design an epitope-based peptide vaccine against chikungunya virus. *Future Virology, 7*(10), 1029-1042.

Janeway, C. (1999). *Immunobiology: The immune system in health and disease*. London: Harcourt Brace & Company.

Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften, 72*(4), 212-213.

Kolaskar, A. and Tongaonkar, P. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*, 276(1-2), pp.172-174.

Larsen, J., Lund, O. and Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Research*, 2(1), p.2.

Laskowski, R. A., Macarthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*(2), 283-291.

Lee, S., & Nguyen, M. T. (2015). Recent Advances of Vaccine Adjuvants for Infectious Diseases. *Immune Network, 15*(2), 51.

Liesegang, Thomas J. (August 2009). Varicella zoster virus vaccines: effective, but concerns linger. *Canadian Journal of Ophthalmology*. 44 (4): 379–84.

Maupetit, J., Derreumaux, P., & Tuffery, P. (2009). PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic Acids Research, 37*(Web Server).

Mcmurtrey, C. P., Lelic, A., Piazza, P., Chakrabarti, A. K., Yablonsky, E. J., Wahl, A., . . . Hildebrand, W. H. (2008). Epitope discovery in West Nile virus infection: Identification and immune recognition of viral epitopes. *Proceedings of the National Academy of Sciences, 105*(8), 2981-2986.

Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., . . . Rose, G. D. (1986). Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proceedings of the National Academy of Sciences, 83*(2), 226-230.

Oany, A., Pervin, T., & Emran, A. (2014). Design of an epitope-based peptide vaccine against spike protein of human coronavirus: An in silico approach. *Drug Design, Development and Therapy,* 1139.

Olesen, O. F., Lonnroth, A., & Mulligan, B. (2009). Human vaccine research in the European Union. *Vaccine, 27*(5), 640-645.

Parker, J. M., Guo, D., & Hodges, R. S. (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*(19), 5425-5432.

Parvege, M. M., Rahman, M., Nibir, Y. M., & Hossain, M. S. (2016). Two highly similar LAEDDTNAQKT and LTDKIGTEI epitopes in G glycoprotein may be useful for effective epitope based vaccine design against pathogenic Henipavirus. *Computational Biology and Chemistry, 61*, 270-280.

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology, 3*(1), 120139-120139.

Peters, B., & Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics, 6*(1), 132.

Petersen, L. R., & Powers, A. M. (2016). Chikungunya: Epidemiology. *F1000Research*.

Petitdemange, C., Becquart, P., Wauquier, N., Béziat, V., Debré, P., Leroy, E. and Vieillard, V. (2011). Unconventional Repertoire Profile Is Imprinted during Acute Chikungunya Infection for Natural Killer Cells Polarization toward Cytotoxicity. *PLoS Pathogens*, 7(9), p.e1002268.

Powers, A. and Logue, C. (2007). Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*, 88(9), pp.2363-2377.

Powers, A., Brault, A., Shirako, Y., Strauss, E., Kang, W., Strauss, J. and Weaver, S. (2001). Evolutionary Relationships and Systematics of the Alphaviruses. *Journal of Virology*, 75(21), pp.10118-10131.

Purcell, A. W., Mccluskey, J., & Rossjohn, J. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery, 6*(5), 404-414.

Quiroga, R., & Villarreal, M. A. (2016). Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *Plos One, 11*(5).

Rajesh, K. N. (2007). Peptide vaccines against cancer, infectious diseases, and conception. *Frontiers in Bioscience, 12*(1), 1833.

Ramon G. (1924). Sur la toxine et sur I'anatoxine diphtheriques. *Ann. Inst. Pasteur* 38, 1–10

Reiser, J., Grégoire, C., Darnault, C., Mosser, T., Guimezanes, A., Schmitt-Verhulst, A., . . . Housset, D. (2002). A T Cell Receptor CDR3β Loop Undergoes Conformational Changes of Unprecedented Magnitude Upon Binding to a Peptide/MHC Class I Complex. *Immunity, 16*(3), 345-354.

Saha, S. and Raghava, G. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, 65(1), pp.40-48.

Schwede, T. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research, 31*(13), 3381-3385.

Seeliger, D., & Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-Aided Molecular Design, 24*(5), 417-422.

Sette, A., Chesnut, R., & Fikes, J. (2001). HLA expression in cancer: Implications for T cell-based immunotherapy. *Immunogenetics, 53*(4), 255-263.

Sharmin, R., & Islam, A. B. (2014). A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. *BMC Bioinformatics, 15*(1), 161.

Shrestha, B., & Diamond, M. S. (2004). Role of CD8 T Cells in Control of West Nile Virus Infection. *Journal of Virology, 78*(15), 8312-8321.

Sievers, F., Wilm, A., Dineen, D., Gibson, T., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. and Higgins, D. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), pp.539-539.

Simon, F., Javelle, E., Oliver, M., Leparc-Goffart, I. and Marimoutou, C. (2011). *Chikungunya Virus Infection*.

Skwarczynski, M. and Toth, I. (2016). Peptide-based synthetic vaccines. *Chemical Science*, 7(2), pp.842-854.

Solignat, M., Gay, B., Higgs, S., Briant, L. and Devaux, C. (2018). *Replication cycle of chikungunya: A re-emerging arbovirus*.

Sun, P., Ju, H., Liu, Z., Ning, Q., Zhang, J., Zhao, X., . . . Li, Y. (2013). Bioinformatics Resources and Tools for Conformational B-Cell Epitope Prediction. *Computational and Mathematical Methods in Medicine, 2013*, 1-11.

Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., . . . Holzhütter, H. (2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage,TAP transport and MHC class I binding. *CMLS Cellular and Molecular Life Sciences, 62*(9), 1025-1037.

Thevenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., & Tuffery, P. (2012). PEP-FOLD: An updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Research, 40*(W1).

Thundimadathil, J. (2012). Cancer Treatment Using Peptides: Current Therapies and Future Prospects. *Journal of Amino Acids, 2012*, 1-13.

Vaccines. (2017, October 05). Retrieved from http://www.who.int/topics/vaccines/en/

Vita, R., Zarebski, L., Greenbaum, J. A., Emami, H., Hoof, I., Salimi, N., . . . Peters, B. (2009). The Immune Epitope Database 2.0. *Nucleic Acids Research, 38*(Suppl_1).

Vlieghe, P., Lisowski, V., Martinez, J., & Khrestchatisky, M. (2010). Synthetic therapeutic peptides: Science and market. *Drug Discovery Today, 15*(1-2), 40-56.

Voss, J., Vaney, M., Duquerroy, S., Vonrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G. and Rey, F. (2010). Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature*, 468(7324), pp.709-712.

Wahid, B., Ali, A., Rafique, S. and Idrees, M. (2018). *Global expansion of chikungunya virus: mapping the 64-year history*.

Waterhouse, A., Procter, J., Martin, D., Clamp, M. and Barton, G. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), pp.1189-1191.

Weaver, S. (2014). Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. *PLoS Neglected Tropical Diseases*, 8(6), p.e2921.

Weaver, S., Osorio, J., Livengood, J., Chen, R. and Stinchcomb, D. (2012). Chikungunya virus and prospects for a vaccine. *Expert Review of Vaccines*, 11(9), pp.1087-1101.

Who.int. (2018). [online] Available at: http://www.who.int/biologicals/publications/trs/areas/vaccines/peptide/WHO_TRS_889_A1.pdf?ua=1 [Accessed 29 Apr. 2018].

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*(Web Server).

World Health Organization. (2018). *Chikungunya*. [online] Available at: http://www.who.int/en/news-room/fact-sheets/detail/chikungunya [Accessed 29 Apr. 2018].

Xue, M., Shi, X., Zhang, J., Zhao, Y., Cui, H., Hu, S., . . . Wang, Y. (2012). Identification of a Conserved B-cell Epitope on Reticuloendotheliosis Virus Envelope Protein by Screening a Phage-displayed Random Peptide Library. *PLoS ONE, 7*(11).

Yang, H. and Kim, D. (2015). Peptide Immunotherapy in Vaccine Development. *Peptide and Protein Vaccines*, pp.1-14.

Yang, X. and Yu, X. (2009). An introduction to epitope prediction methods and software. *Reviews in Medical Virology*, 19(2), pp.77-96.

Yasmin, T., Akter, S., Debnath, M., Ebihara, A., Nakagawa, T., & Nabi, A. H. (2016). In silico proposition to predict cluster of B- and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of C. jejuni. *In Silico Pharmacology, 4*(1).

Yewdell, J. W., & Bennink, J. R. (1999). Immunodominance In Major Histocompatibility Complex Class I–Restricted T Lymphocyte Responses. *Annual Review of Immunology, 17*(1), 51-88.

50% of what? How exactly are IC50 and EC50 defined? (n.d.). Retrieved from https://www.graphpad.com/support/faq/50-of-what-how-exactly-are-ic50-and-ec50-defined/