Extensive Immunoinformatics study for the prediction of novel peptide-based epitope vaccine with docking confirmation against Envelope protein of Chikungunya virus: A Computational...

Article *in* Journal of biomolecular Structure & Dynamics · February 2020

DOI: 10.1080/07391102.2020.1726815

CITATIONS
25

READS
468

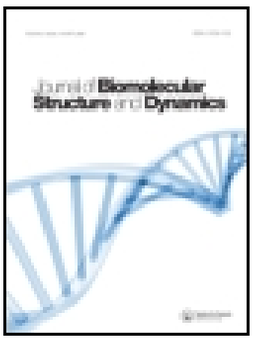
8 authors, including:

Syed Shahariar Bappy
Mawlana Bhashani Science and Technology University
2 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Sorna Sultana
Mawlana Bhashani Science and Technology University
2 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Juthi Adhikari
Universiti Brunei Darussalam
11 PUBLICATIONS **40** CITATIONS
SEE PROFILE

Shafi Mahmud
56 PUBLICATIONS **464** CITATIONS
SEE PROFILE



Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Extensive Immunoinformatics study for the prediction of novel peptide-based epitope vaccine with docking confirmation against Envelope protein of Chikungunya virus: A Computational Biology Approach

Syed Shahariar Bappy, Sorna Sultana, Juthi Adhikari, Shafi Mahmud, Md. Arif Khan, K. M. Kaderi Kibria, Md. Masuder Rahman & Abu Zaffar Shibly

To cite this article: Syed Shahariar Bappy, Sorna Sultana, Juthi Adhikari, Shafi Mahmud, Md. Arif Khan, K. M. Kaderi Kibria, Md. Masuder Rahman & Abu Zaffar Shibly (2020): Extensive Immunoinformatics study for the prediction of novel peptide-based epitope vaccine with docking confirmation against Envelope protein of Chikungunya virus: A Computational Biology Approach, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2020.1726815

To link to this article: https://doi.org/10.1080/07391102.2020.1726815

Extensive Immunoinformatics study for the prediction of novel peptide-based epitope vaccine with docking confirmation against Envelope protein of Chikungunya virus: A Computational Biology Approach

Syed Shahariar Bappy[1], Sorna Sultana[1], Juthi Adhikari[1], Shafi Mahmud[2], Md. Arif Khan*[3,4], K. M. Kaderi Kibria[1], Md. Masuder Rahman*[1], Abu Zaffar Shibly*[1]

[1]Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902.

[2] Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, Bangladesh

[3]Department of Biotechnology and Genetic Engineering, University of Development Alternative, 4/4B, Block A, Lalmatia, Dhaka-1209. 4

[4]Bio-Bio-1 Research Foundation, Sangskriti Bikash Kendra Bhaban, 1/E/1, Poribag, Dhaka 1000, Bangladesh

*Corresponding Author: Abu Zaffar Shibly, Assistant Professor, Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh. Email: zaffarshibly1987@gmail.com

Running Title: Immunoinformatics study against Chikungunya virus

Abstract

Background: Chikungunya virus (CHIKV) instigating Chikungunya fever is a global infective menace resulting in high fever, weakened joint-muscle pain, and brain inflammation. Inaccessibility and unavailability of effective drugs have led us to an uncertain arena when it comes to providing proper medical treatment to the affected people.

Method: In this study, authentic encroachment has been made concerning the peptide-based epitope vaccine designing against CHIKV. A Proteome-wide search was performed to locate a conserved portion among the accessible viral outer membrane proteins which showcase a remarkable immune response using specific immunoinformatics and docking simulation tools.

Results: Primarily, the most probable immunogenic envelope glycoproteins E1 and E2 were identified from the UniProt database depending on their antigenicity scores. Subsequently, we selected two distinctive sequences “SEDVYANTQLVLQRP” and “IMLLYPDHPTLLSYR” in both E1 and E2 glycoproteins respectively. These two sequences identified as the most potent T and B cell epitope-based peptides as they interacted with 6 and 7 HLA-I and 5 HLA-II molecules with an extremely low IC50 score that was verified by molecular docking. Moreover, the sequences possess no allergenicity and are certainly located outside the transmembrane region. In addition, the sequences exhibited 88.46% and 100.00% Conservancy, covering high population coverage of 89.49% to 94.74% and 60.51% to 88.87% respectively in endemic countries.

Conclusion: The identified peptide SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR can be utilized next for the development of peptide-based epitope vaccine contrary to CHIKV, So further documentations and experimentations like Antigen testing, Antigen production, Clinical trials are needed to prove the validity of it.

Keywords: Chikungunya, Immunoinformatics, Epitope, Peptide, Vaccine, Molecular Docking studies.

List of Abbreviations

CHIKV	Chikungunya virus
E	Envelope
HLA	Human Leukocyte Antigen
NIAID	National Institute of Allergy and Infectious Diseases
WHO	World Health Organization
FAO	Food and Agriculture Organization
TAP	Ransporter of Antigen Presentation
MHC	Major Histocompatibility Complex
APC	Antigen Presenting Cell
NSAIDs	Non-steroidal anti-inflammatory drugs.
NCBI	National Center for Biotechnology Information
CTL	Cytotoxic T Lymphocyte
IEDB-AR	Immune Epitope Database and Analysis Resource
IC_{50}	Half Maximal Inhibitory Concentration
SMM	Stabilized Matrix Method
TAP	Transport associated proteins
TMHMM	Transmembrane Helices;Hidden Markov Model / Tied Mixture Hidden Markov Model.
RCSB	Research Collaboratory for Structural Bioinformatics.
CASTp	Computed Atlas of Surface Topography of proteins
ARPs	Allergen Representative Peptides
MEME	Multiple Em for Motif Elicitation
MAST	Motif Alignment and Search Tool
MSA	Multiple Sequence Alignment.
SA	Solvent-Accessible

1. Introduction

The word Chikungunya means, “that which flexure up”, notice as the severe joint pains in the part of an infection which comes from the Makonde language. [1]. One of the arthropod-borne viruses (Chikungunya virus, Alphavirus member and Togaviridae family) causing Chikungunya fever is emerged [2, 3] and re-emerged [4, 5] as a concerning issue in recent years which spread by Aedes spp. Mosquitoes. Like other viruses, such as dengue fever, ross river fever and yellow fever it has some identical symptoms [6].

From the infected mosquitoes, CHIKV is transmitted by the urban cycle (human-mosquito-human) and sylvatic cycle (animal-mosquito-human) [6]. The day-biting mosquitoes are predominantly responsible for the transmission of CHIKV [7, 8]. Liver and joints are the CHIKV replication and expand place after being transmitted to the human body by mosquitoes [9]. The acute sign appeared after 2-4 days incubation period. Though the fever is fatal, however, its weakening symptoms are rash, headache, muscle pain, fever, arthralgia and sometimes also can be lead to inflammation in the brain, even death in some people [10, 11, 12]. Weeks to years polyarthralgia characteristics are seen at the chronic stage [13]. In infected patients, neurological disorders and eye infection are also reported, while meningoencephalitis and hemorrhagic manifestations such as haematemesis and melaena leading to death have been reported [12, 14, 15].

Almost 55 different countries in the world, millions of people are affected by Chikungunya Virus [13]. In 1950 the CHIKV was first recorded by an outbreak in Tanzania, Africa [16] and in 1958 first time in Asia was also reported [17]. Since 2000, Chikungunya disease and it’s symptoms has become a serious public health problem [18]. In 2005–2006, several Indian Ocean Islands affected by an unprecedentedly large epidemic of CHIKV(300,000 cases with 237 deaths) before spreading to India and Southeast Asia [37]. From then expansion occurred in a different country like Asia, Europe, the Americas, and Australia from Africa by infected travelers [12, 19, 20]. Moreover, in 2007, CHIKV emerged for the first time in a temperate area Italy [21] and is now a worldwide infectious threat [22]. The rapid expansion across the natural ground becoming a global public health concern [23, 24]. However, medicine or vaccination is unavailable for CHIKV until now.

Chikungunya is an analphavirus single-stranded RNA [25]. In 1984, African and Asian strains of CHIKV were used by Simizu *et al to* analyze the structural proteins of CHIKV [26]. In 2002, the CHIKV (S27, African prototype) genome sequence was fully identified by Khan *et al* [27]. The virus encodes 9 genes, consisting of coding sequences for non-structural proteins (nsP1, nsP2, nsP3, and nsP4), and five structural proteins, including a capsid (C) protein, three envelope glycoproteins (E1, E2, and E3) and a small molecule, 6K [28, 29]; also 7-methylguanosine group capped at the 5' end and the polyadenylated group at the 3' end, flanked by 5´ and 3´ sequences, these include two opening reading frames (ORFs) with junction region between them [27].

The structural analysis of CHIKV at pseudo-atomic level resolution was also reported [25], by combining cryoelectron microscopy (cryo-EM) techniques for the whole virus and X-ray

crystallography for the component of structural proteins, together with the published crystal structure of the CHIKV E1-E2 glycoprotein heterodimer [30]. The virus has 80 spikes. These spikes are made by the envelope glycoproteins E1, E2, and together with virus membrane, Transmembrane (TM) helix, and a capsid covering genome RNA to form a nucleocapsid [25]. Other alphaviruses feature, such as Sindbis virus (SINV) and Semliki Forest virus (SFV) helps to understand the role and function of CHIKV non-structural and structural proteins. Information from these two viruses, the non-structural proteins of CHIKV help to the formation of the transcription or replication complex and the negative strand production, on the other hand, the structural proteins of CHIKV related to the virus pathogenesis, these glycoproteins mainly conduce the infection [28, 31, 32, 33]. The E2 protein bind to the cell surface receptor and E1 protein helps to membrane fusion [30]. The mature CHIKV has absent E3 protein [26]. The 6K formed p62-E1 complex and before assemblage, it is deported to the plasma membrane [25].

The 439 amino acid containing E1 glycoprotein consists of 404 residues N-terminal part, 30 residues TM helix, and 5 cytoplasmic domain forming amino acids, as well as Asn141, an N-linked glycosylation site. The E2 glycoprotein consists of 364 residues that include a 26 residues TM helix, and 33 cytoplasmic domain forming amino acids. The E1 and E2 glycoprotein both divided into three domains, E1 glycoprotein has I, II and III on the other hand E2 glycoprotein has A, B, and C. Domain A is responsible for the receptor-binding process, stay between domain B and domain C. Domain B covers the fusion loop in domain II of E1 in the mature structure. Under the pH acidic environment with the combination of three E1 trimers, the virus becomes fusogenic. The fusion loop is then inserted into the host cell membrane and by this; it may interfere with the virus entry and infection of human tissue [25].

Adaptive immunity plays a major role in any virus elimination completely from the human body, although its characterization is still incomplete against CHIKV [34, 35, 36]. T cell epitopes are recognized by T lymphocytes by subsequent presentation of the antigen by HLA molecules in adaptive immunity. This generates T cell-mediated cytotoxicity and humoral immune response. With the help of helper T cells, the linear form of specific epitopes is recognized by B cell receptors (BCR) also. Then phagocytic or complementary pathways initiated after binding antibodies to their specific antigens. Memory B cells are promoted from some plasma cells which give a stable immunity against the virus [37]. Dominant immune epitopes identification of a virus is concised by vaccine design technique of epitope-based T-cell and it can be used as vaccines to produce specific immune response [38]. The effective therapeutic agent or licensed vaccine is not developed until now to cure human [20]. CHIKV is an envelope containing alphavirus. So, it has a high possibility to produce a vaccine from envelope proteins E1 and E2 that show a high immune response. Immune dominant universal B-Cell and T-cell epitopes from E1 and E2 envelope protein of CHIKV may be the best epitope-based peptide vaccine. Here, we try to find out a potential immune dominant outer membrane region of envelope protein by computational methods. The aim of this work was to facilitate laboratory-based vaccine design

for the worldwide treatment process against this pathogenic virus and find out the preventive way of Chikungunya epidemic and threat.

2. Materials and Methods

The step by step flow diagram representing the procedures of peptide-based epitope vaccine development for Chikungunya virus is illustrated in Figure 1.

2.1 Sequences Retrieval

The amino acid sequences of CHIKV envelope (E1 and E2) protein were downloaded from UniProt (http://www.uniport.org) [39] or NCBI (http://www.ncbi.nlm.nih.gov/protein) protein database in FASTA format. E1 envelope glycoprotein and E2 envelope glycoprotein are separated from all of the other proteins of CHIKV.

2.2 Highest Antigenic Protein Selection

All of the proteins were then submitted in the VaxiJen v2.0, [40, 41] a Web-based Server, that is used for the prediction of their antigenicity. One E1 envelope glycoprotein and one E2 Envelope glycoprotein with highest antigenicity scores were selected from all of the other proteins of CHIKV those checked by VaxiJen server for next steps.

2.3 T cell Epitope Prediction

2.3.1 Peptide Processing and MHC Binding

For the prediction of human cytotoxic T lymphocyte (CTL) epitopes in any given protein, NetCTL-1.2 server [42] is designed by which epitopes was selected. CTL epitopes prediction is important for coherent vaccine design and it can reduce the amount of the experimental work. This predicts candidate epitopes based on the processing of the peptides in vivo [43] which also covers 12 HLA-I super types. The score was estimated by integrating the three predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I affinity. On the basis of the score, the best candidates were selected for advance analysis.

T-Cell Epitope calculation tools were also used for the identification of MHC-I processing (http://tools.iedb.org/processing/) [44] peptide with allele molecules or MHC-I (http://tools.iedb.org/mhci/) and MHC-II (http://tools.iedb.org/mhcii/) alleles molecule binding into the peptide at the Immune Epitope Database and Analysis Resource (IEDB-AR). Stabilized Matrix Method [45] was applied to calculate the half-maximal inhibitory concentration (IC_{50}) values that required for peptide binding to MHC-I molecules, peptide length was set to 9 amino acids earlier to the prediction. In the case of MHC-II binding analysis, the IEDB-recommended method was used for the specific HLA-DQ, HLA-DP, and HLA-DR loci. Here in, specific peptides were used to predict the MHC-II interaction on the basis of MHC-I analysis with SMM-align core [46] method. The alleles having binding affinity IC50 less than or equal 250nm were selected in both cases MHC-I and MHC-II.

2.3.2 Prediction of Transmembrane Protein Topology

The topology of a transmembrane protein refers to orientations (locations of N- and C-terminal) of membrane-spanning segments with respect to the inner or outer sides of the biological membrane occupied by the protein. The most successful method to date helps to predict the full topology of the protein[47]. Target protein uploaded in FASTA format into a hidden Markov model-based tool TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [48].

2.3.3 Allergenicity Assessment

For the evaluation of Allergenicity of new food proteins, two organizations, the United Nations Food and Agriculture Organization (FAO) and the World Health Organization (WHO) formed guidelines [49]. According to these guidelines, a protein is a potential allergen if it has either identity of 6 to 8 consecutive amino acids or greater than 35% overall sequence similarity over a window of 80 amino acids when compared with known allergens [50]. During the last 12 years, different bioinformatics tools have been established. To predict the Allergenicity of the selected epitope, AlgPred (http://crdd.osdd.net/raghava/algpred/ submission.html), an online server, [51] was implemented. Here it can predict the selected epitopes as allergen or non-allergen.

2.4 Prediction of B-cell Epitope

B-cell epitope instigates the differentiation of B-lymphocytes into memory cells and plasma [52]. Peptide sequence submitted into an epitope prediction tool ABCpred (http://crdd.osdd.net/raghava/abcpred/ABC_ submission.html) as the first step. Then Linear B-cell epitopes predicted from the given maximum immunogenic protein sequence through the B-cell epitope prediction tools of IEDB (http://tools.iedb.org/bcell/). For the B-cell epitope prediction with high accuracy, multiple tools, including the Emini surface accessibility prediction [53], Kolaskar and Tongaonkar antigenicity scale [54], Parker hydrophilicity prediction [55], Karplus and Schulz flexibility prediction [56] and lastly the Chou and Fasman beta-turn prediction tool [57] were used, because the antigenic parts of a protein belonging to the beta-turn regions [58]. Multiple tools assessment strengthens the possibility of peptide vaccine candidates.

2.5 Prediction of Epitope Conservancy

Epitope conservancy for selected epitopes was calculated using the IEDB tool [59]. Conservancy of each epitope was estimated for both filtered E1 envelope glycoprotein and Envelope glycoprotein E2 of CHIKV found in UniProtKB and some other E1 and E2 protein from a different strain. The conserved sequence was identified using Clustal Omega [60] multiple sequence alignment tool.

2.6 Population Coverage Calculation

Predicted epitopes with corresponding different HLA Class I and II alleles were submitted to the population coverage analysis tool of IEDB [61]. In case of single Population coverage, the tool predicted population coverage, HLA combinations recognized by the population, and HLA combinations identified by 90% of the population (PC90). First all of the epitopes and their respective MHC-I molecules uploaded and later population coverage area selected before submission [60].

2.7 Molecular Docking

For the molecular docking simulation study first step was to retrieving the 3D crystal structure of HLA-I and HLA-II from RCSB (Research Collaboratory for Structural Bioinformatics) [62] and then prepared for docking runs by removing the ligands from these structures, if any ligand attached with it. Then the HLA structure or PDB code and epitope sequence were input into an online docking server CABSdock (http://biocomp.chem.uw.edu.pl/CABSdock/). The CABS-dock provides an interface for modeling protein-peptide interactions using a exceedingly efficient protocol for the flexible docking of proteins to peptides [63, 64]. In case of CABS-dock, first a protein structure and a peptide sequence inputted, later, it performs a simulation to search for the binding site allowing for full flexibility of the peptide and small fluctuations of the receptor backbone.

2.8Molecular Dynamics

YASARA program [65] was employed to perform molecular dynamics simulation study where the AMBER14 force field [66] was used. The complexes were cleaned and hydrogen bonding systems were optimized. A cubic cell of 5Å was formed and periodic boundary condition was maintained. Transferable intermolecular potential 3 points or TIP3P model was applied where Na/Cl ions were added and density was 1.012 gm/cm^3. For long-range electrostatic interactions, Particle Mesh Ewald method was engaged with a distance of 8 Å. The physiological system of the interaction was set as (298K, pH 7.4, 0.9% NaCl) [67]. Initial, energy minimization of each system has been implemented by a simulated annealing method using the steepest gradient approach (5000 cycles). Then MD simulation was initialized with a time step interval of 2.50 fs [68]. Finally, MD simulation of each system was run for 50ns long to calculate RMSD (Root mean square deviation), RMSF (Root mean square fluctuation), SASA (Solvent accessible surface area), Rg(Radius of gyration) and MM-PBSA. In our simulation study, MHCII:HLA-DR(PDB ID: 1IIE)-SEDVYANTQLVLQRP, MHCII: HLA-DR(PDB ID: 1IIE) - IMLLYPDHPTLLSYR , MHCI: HLA-C*07:02(PDB ID: 5VGE)-SEDVYANTQLVLQRP and MHC I: HLA-C*07:02(PDB ID: 5VGE)-IMLLYPDHPTLLSYR complex denoted as D1, D2, D3 and D4 respectively.

3. Results

3.1 Protein Sequences Retrieval of CHIKV

Total 6 E1 envelope glycoproteins and 14 E2 envelope glycoproteins sequences from different strains of Chikungunya virus outer membrane envelope protein were retrieved from UniProtKB.

3.2 Identification of Highest Antigenic Protein Sequence of CHIKV

All the proteins were evaluated by VaxiJen server; one E1 and one E2 protein with highest score were taken from them. The protein sequence of E1 with UniProtKB ID: A0A220QT82, containing 440 amino acids, showed the highest VaxiJen score of 0.5203 and on the other hand in case of E2 with UniProtKB ID: H9BFE2, containing 240 amino acids, the

highest VaxiJen score is 0.6023. These two proteins (E1 envelope glycoprotein and E2 glycoprotein) were taken for further analysis.

3.3 T-cell Epitope Prediction CHIKV Protein

3.3.1 Peptide Processing for Epitope Candidate

The best 12 epitopes from 12 super types from the sequences of E1 envelope glycoprotein and envelope glycoprotein E2 were selected respectively based on the combinatorial score from NetCTL prediction server. The overall score was calculated combined by proteasomal cleavage efficiency, TAP transport efficiency and MHC-I binding efficiency. The results are shown in Table S1 and S2 respectively.

3.3.2 MHC class-I Epitope Binding Alleles Identification

Analysis in SMM based (stabilized matrix method) IEDB MHC-I processing prediction tool selected 12 supertype epitopes were found to be recognized by a range of MHC molecule. This shows HLA binding affinity of the epitopes in IC50 nM unit. There is an inverse relationship between binding affinity and IC50 value. Here MHC molecules with IC50 value less then 250nm for both E1 envelope glycoprotein and envelope glycoprotein E2 were selected. MHC-I processing efficiency tool of IEDB predicts an overall score for each epitope based on proteasomal cleavage efficiency, TAP transport efficiency and MHC-I binding efficiency combined. The proteasomal complex contains enzymes that digest proteins to form smaller peptides. Produced peptides are recognized by MHC class I molecules and MHC-I forms a complex with the pep-tides. The peptide-MHC class I complexes are transported to the endoplasmic reticulum, a process facilitated by transport associated proteins (TAP) before being presented to the T-cells on the plasma membrane of the cell. The higher the combined score of the peptides the better they are processed for presentation and that is critical step for creating a successful immune response. A good epitope interact with as many as MHC alleles. Thus, among the total peptides, it was found that only peptides from E1 envelope glycoprotein and peptides from envelope glycoprotein E2 have core epitope of MHC-I. To find out the most potential epitope based peptide sequence we were followed that which interacted highest MHC1 alleles in limited IC50 score. In E1 envelope glycoprotein SLDYITCEY, VYNMDYPPF, VYANTQLVL, SQAPSGFKY interact with highest 6 MHC-I alleles; and envelope glycoprotein E2 YYYELYPTM interact with highest 9 and TVNGQTVGY, LYPDHPTLL 7 MHC-I alleles. These all are interact with more than 5 MHC-I alleles. More alleles are better for good binding. The result of both envelope glycoproteins obtained from IEDB MHC-I binding analysis and processing tools are summarized in Table 1 and 2.

3.3.3 MHC Class-II Epitope Binding Alleles Identification

Both E1 and E2 envelope glycoprotein (GP) was analyzed using IEDB MHC-II binding prediction tool based on SMM-align with half-maximal inhibitory concentration (IC50) ≤ 450. A good core epitope containing peptides also interact with as many as MHC alleles. Thus, among

the total peptides, it was found that only peptides from E1 envelope glycoprotein and peptides from envelope glycoprotein E2 have core epitope of MHC-I. To find out the most potential epitope-based peptide sequence we were followed that which interact highest MHC-II alleles in limited IC50 score. In E1 envelope glycoprotein SEDVYANTQLVLQRP and envelope glycoprotein E2 IMLLYPDHPTLLSYR both interact highest 5 MHC-II alleles. Output results of the MHC-II interaction analysis are also shown in Table 3 and 4.

3.3.4 Prediction Transmembrane Protein Topology of CHIKV Protein

After submitting both target proteins in fasta format into TMHMM Server v. 2.0, a hidden Markov model-based tool, we were found that among 439amino acids, full length of E1 envelope glycoprotein 1-413 amino acids were outside, 414-436 amino acids were TMhelix and 437-439 amino acids were inside; and among 239 amino acids, full-length envelope glycoprotein E2 1-180 amino acids were outside, 181-203 amino acids were TMhelix and 204-239 amino acids were inside position. If the whole sequence is labeled as inside or outside, the prediction is that it contains no membrane helices. Results are shown in Figure S1 and Figure S2.

3.3.5 Allergenicity Assessment CHIKV Protein:

The mapping of IgE epitope(s) feature of the Algpred server makes the user to locate the epitope position in their protein. Another facilitates obtained from Bjorklund et al 2005 [69] and assign a protein allergen if it has a BLAST hit. Server search MEME/MAST allergen motifs using MAST and assign a protein allergen if it has any motif. After submitting all peptides result was found as allergen or non-allergen (Table S3 and Table S4).

After analyzing the four prediction tools result, it was found that SEDVYANTQLVLQRP from E1 envelope glycoprotein and IMLLYPDHPTLLSYR from envelope glycoprotein E2 is the best epitope-based peptide candidate. Later, B-cell epitope condition and final analysis of these proposed epitopes conducted.

3.4 B-cell Epitope Prediction CHIKV Protein

3.4.1 ABCpred Detection of Viral Protein

After submitting ABCpred server both E1 envelope glycoprotein and envelope glycoprotein E2, we had found some sequence for the potential candidate. The B-cell epitopes predicted and ranked based on the score obtained by the trained recurrent neural network. The higher the score, the higher the probability of an epitope selected. All the epitopes counted here are above the threshold values. Here 14 mer top 12 peptides from E1 envelope glycoprotein and 14 mer top 10 peptide from envelope glycoprotein E2 was taken. The T-cell proposed epitope of E1 envelope glycoprotein conserved 11 number peptides and envelope glycoprotein E2 conserved 9 positionsin both cases (Table 5).

3.4.2 B-cell Epitope Analysis by IEDB Tools

Here a number of tools used with different analytical methods to identify the suitable B-cell epitope respectively.

I . Kolaskar and Tongaonkar Antigenicity Prediction

The physicochemical properties of peptides were determined by utilizing the Kolaskar and Tongaonkar antigenicity prediction method. The average antigenic propensity of the E1 envelope glycoprotein was 1.055 and envelope glycoprotein E2 was 1.041. The values greater than threshold were considered as a potential antigenic determinant. The T-cell proposed epitopes were considered to have the potentiality to produce the B-cell immunity. Results are summarized in Figure 2; Table 5.

II . Emini Surface Accessibility Prediction

A potent B-cell epitope must have surface accessibility. The average Emini surface accessibility of the both E1 envelope glycoprotein and envelope glycoprotein E2 was 1. The values greater than threshold were considered as potential surface accessibility career. Results are summarized in Figure 3.

III . Parker Hydrophilicity

A good B-cell epitope must have the hydrophilic regions. The average Parker hydrophilicity of the E1 envelope glycoprotein was 1.553 and envelope glycoprotein E2 was 1.228. The values greater than threshold were considered as potential hydrophilicity career. Results are summarized in Figure 4.

IV . Chou and Fasman Betaturn Prediction

The beta turns are significantly hydrophilic and frequently accessible feature carrier. These are two properties a protein of antigenic regions [70] proposed both E1 envelope glycoprotein and envelope glycoprotein E2 conserved sequences in beta turn region. Results are summarized in Figure 5.

V. Karplus and Schulz Flexibility Prediction

The flexibility of the peptide is correlated with antigenicity, already which is experimentally proved [71]. Here in Karplus and Schulz method predicted the flexibility of the epitope region. According to the result our proposed epitope conserved flexibility showing region. Results are summarized in Figure 6.

3.5 Epitope Conservancy Analysis

The Conservancy of epitopes with all selected protein sequences from different strains of Chikungunya was analyzed. It was showed that the predicted E1 protein epitope was 88.46% of protein sequence matches at identity <= 100% with 93.33% minimum identity and 100%

maximum identity where epitope is 100% conserved in 23 protein among 26 different strain protein and E2 protein epitope was 100.00% of protein sequence matches at identity <= 100% with 100.00% minimum identity and 100% maximum identity where epitope is 100% conserved in 35 protein among different strain. The result is showed in Table 6. Epitope conservancy checked and analyzed by multiple sequence alignment by clustal omega that is in Figure S3 and S4.

3.6 Population Coverage Analysis

Combined population coverage of different epitopes of CHIKV E1 envelope glycoprotein and CHIKV envelope glycoprotein E2 with their corresponding MHC-I and MHC-II alleles for different countries are shown in Table S6 and S7. Among the countries, Central Africa showed the highest population coverage of 94.74% as well as the second highest population coverage showed 93.22% in Europe, 90.55% for West Africa, 90.24% for South Africa, 89.49% for East Africa, 82.72% respectively for world population coverage for our predicted epitope of CHIKV E1 envelope glycoprotein. In case of CHIKV glycoprotein E2 epitopes, Europe showed the highest population coverage of 88.87%, as well as South Asia, showed the second highest population coverage of 81.39%, 76.98% for Central Africa, 69.99% for East Africa, 60.51% for West Africa, 84.12% respectively. Graphical view of world combined population coverage of our proposed epitope of CHIKV E1 and E2 were showed in Figure S5.

3.7 Molecular Docking of Peptide with Allele

Before performing docking 3D crystal structure of HLA-C*07:02(PDB ID: 5VGE) and HLA-DR (PDB ID: 1IIE) were retrieved from RCSB and submitted into CABSdock server with epitope sequence. For CABSdock run, there is no need of collecting epitope PEP-FOLD 3D structure. Results found from CABSdock, Epitope-allele binding structures and interacted amino acid of receptor/peptide residues are shown in Figure 7, 8, S6, and S7 as well as Table S8, S9, S10 and S11.

3.8 Molecular Dynamics

To obtain more accurate perceptions about the binding pattern and conformational variation of the systems, we utilized trajectory derived Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Solvent Accessible Surface Area (SASA), the Radius of Gyration (Rg) and Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) against time-dependent function. Therefore, RMSD of the C-alpha atom of the complex can provide evidence of complex constancy and stiffness. From Figure 9, it can be observed that all four complexes had a tendency towards high RMSD during the initial phage of simulation but it did not go beyond 3Å. Meanwhile, from 16ns this tendency is shrunken considerably and the complex seemed to be at steady state after 20ns. D1 and D4 complex had lesser RMSD profile after equilibrium whereas D3 had upper RMSD value than all other complexes. However, a little bit different scenario was depicted for the radius of gyration, and this time the D2 complex had greater Rg value during the whole simulation time. And, interestingly, for D1 and D4 complex

no fluctuations were observed which indicated the stiff nature of the complex and conformational invariability [72, 73].

To understand the flexible region of the protein we had also explored the RMSF value of the amino acid sequence of the respected protein complex where lower RMSF value indicates less variability. RMSF value illustrated in Figure 10 shows that D1 and D2 complex exhibit almost similar trends from 118-172 region which is composed of helix strand, beta, and gamma turn. Higher flexibility of the D1 and D2 complex observed from 178-192(beta and gamma turn). On the other hand, Figure 10 also demonstrated that, Ser42 (beta-turn), Arg108 (helix strand), Arg131 (beta-turn), Arg145 (helix strand), Arg151 (helix), Asp196 (beta-hairpin), Gln226 (beta-turn), Pro276 (helix) showed a high level of flexibility than other amino acid residues, and particularly in these two simulating systems, D3 had a higher RMSF peak than D4.

We also calculated SASA value from the trajectory, where all four complexes revealed a stable profile. The contracted nature of the protein complex depends on the lower deviation of the SASA profile, and similarly, a higher peak of SASA denotes the expansion of protein [74]. This law makes our findings more realistic and indicates the less deviated nature of the protein-complex system. Furthermore, in order to validate the docking score, the MM-PBSA method was implemented. From Figure 11, it can be observed that the total binding energy of all complex was in the range of (-100 to -850)kJ/mol which denotes favorable binding of the systems [75].

4. Discussion

Vaccine generation within a concise period against the growing viral attacks has become fateful [76, 77]. The procedure of epitope-based vaccine generation has been described against Dengue virus [78], Nipah virus [80], Ebola virus [81], Herpes simplex virus [82], *Mycobacterium tuberculosis* [83], human norovirus [84], Yellow fever virus [87], Infective endocarditis [88], Alkhurma hemorrhagic fever virus [89], and *Elizabethkingia anopheles* [90]. In this study, an effort is given to design epitopes that might be tested for their efficiency reducing immunity through humoral and cell-mediated immune responses for vaccine development. Several bioinformatics tools are being used to identify and characterize the potential epitopes from the antigenic protein of CHIKV that may expose efficient B-cell and T-cell mediated immunity. The whole work is followed by the methods illustrated in Figure 1. E1 and E2 protein facilitate the attachment and fusion of outer membrane protein with host cell membranes. So, these two essential involvements at the entry, these proteins were targeted for designing most potential epitopes using *in silico* computational approaches. The envelope glycoprotein sequence UniPortKB id: A0A220QT82 and H9BFE2 of CHIKV were considered as the potential and well conserved T-cell epitope. From these sequences, the NetCTL server predicted 12 epitopes from each supertype from each protein. SLDYITCEY, VYNMDYPPF, RPGYSPMVL, WLKERGASL, VYANTQLVL, IEVEGNSQL, FTGVYPFMW epitopes from CHIKV E1 envelope glycoprotein and GTNHKKWQY, LYPDHPTLL, TVNGQTVGY, VMHKKEVVL, RRCITPYEL epitopes from envelope glycoprotein E2 are the top score retaining; scores were above 1.50 (Table S1 and S2). From CHIKV E1 envelope glycoprotein SLDYITCEY,

VYNMDYPPF, VYANTQLVL, SQAPSGFKY epitopes found highest 6 HLA-I and CHIKV envelope glycoprotein E2 YYYELYPTM found 9, LYPDHPTLL and TVNGQTVGY found 7, FLLSLICCI found 6 HLA-I respectively; all these are more than 5 HLA-I(HLA-A, HLA-B and HLA-C) retaining epitope (Tables 1 and 2). Here, highest 5 HLA-II molecules could interact with the MHC-I epitope conserved peptide SEDVYANTQLVLQRP of CHIKV E1 envelope glycoprotein and IMLLYPDHPTLLSYR of CHIKV envelope glycoprotein E2 (Table 3 and 4). The best epitope interacts with more alleles because the efficiency of an epitope vaccine greatly relies on the precise interaction between epitope and HLA alleles [78]. Also, SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR both existed outside the region from the membrane (Figure 2). These were non-allergic peptides (Table S3 and Table S4). The B-cell primary and secondary immunity together were searched for SEDVYANTQLVLQRP from CHIKV E1 envelope glycoprotein and IMLLYPDHPTLLSYR from CHIKV envelope glycoprotein E2. At first, we predicted peptide by ABCpred server and found that these two epitopes are conserved in the top 10 predicted peptides (Table 5 and 6). Then bioinformatics analysis tools from the IEDB server used to predict B-cell epitopes based on the important features of protein such as beta-turns, surface accessibility, hydrophilicity, flexibility and proved to be antigenic (Figures 2, 3, 4, 5,6,and Table S5). The 9-mer epitope VYANTQLVL and LYPDHPTLL conserved peptide SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR were also good satisfactory peptide as B-cell immunity. At the time of analysis, the conservancy was found 88.46% for SEDVYANTQLVLQRP peptide in different strain of E1 protein (Table 7 and Figure S3) and 100.00% for IMLLYPDHPTLLSYR peptide in different strain of E2 protein (Table 8 and Figure S4). Due to the mutation in CHIKV like RNA viruses, a vaccine candidate epitope must come from the portion of the protein showing many conservancies that will ensure an effective long-lasting immunization [79]. Then peptide SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR were looked for population coverage. CHIKV endemic areas were the primary focus of the search. CHIKV endemic region was recorded highest combined coverage 94.74%, 90.55%, 89.49%, 90.24% and 93.22% in Central Africa, West Africa, and East Africa, South Asia, and Europe respectively for SEDVYANTQLVLQRP peptide of CHIKV E1 envelope glycoprotein and 76.98%, 60.51%, 69.99%, 81.39% and 88.87% in Central Africa, West Africa and East Africa, South Asia, and Europe respectively for IMLLYPDHPTLLSYR peptide of CHIKV envelope glycoprotein E2 (Tables S6 and S7). On the other hand, Hasan et al. proposed an epitope “YYYELYPTM” and found 69.71–72.87% coverage in different CHIKV endemic regions of Africa [85]. In some other this type of study, the highest population coverage found in Germany and that is around 80% [79, 86]. World combined coverage was found 82.72% for SEDVYANTQLVLQRP and 84.12% for IMLLYPDHPTLLSYR (Figure S5). To ensure the interaction both epitope-based peptide SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR were docked to both HLA-I allele (HLA-C*07:02) and HLA-II (HLA-DR); and confirmed that the epitope could bind effectively to HLA-I and HLA-II (Figures 7 and 8). Pairs of peptide/receptor amino acid residues during docking are showed in Table S8, S9, S10 and S11. So SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR epitope-based peptide would be the most probable vaccine candidate. To understand the binding mechanism, the flexibility of the complex, we conducted molecular dynamics simulation for 50ns. The RMSD, RMSF, Rg,

SASA and MM-PBSA profile of the four complexes were explored. The complex stability and rigidness were confirmed by assessing dynamics derived results.

Though the challenging and complex process of development, peptide vaccines are better than conventional vaccines because of its safety and inexpensive nature. The study is an attempt to identify potential peptide-based epitope vaccine candidate against CHIKV using computational biology tools. This computational biology approaches with a higher probability of success will reduce the wet lab efforts. Therefore, the identified peptide SEDVYANTQLVLQRP and IMLLYPDHPTLLSYR may be utilized further for developing peptide-based epitope vaccine against CHIKV, So further documentation and experimentation is needed to prove the validity of it.

Declaration of statement

There is no potential conflict of interest to disclose with this submission.

Author contributions

Syed Shahariar Bappy: Performed Bioinformatics experiment, Data Analysis, Epitope Identification, Vaccine design and Docking Study

Sorna Sultana: Performed Bioinformatics experiment, Data Analysis

Juthi Adhikari: Data Analysis, Manuscript writing

Shafi Mahmud: Molecular dynamics, manuscript writing and editing.

Md. Arif Khan: Conceptualization, Co-Supervision, Critical Data analysis, Manuscript writing, and Reviewing

K. M. Kaderi Kibria: Helped in Bioinformatics analysis and drafting the manuscript, Reviewing

Md. Masuder Rahman: Helped in Bioinformatics analysis and drafting the manuscript, Reviewing

Abu Zaffar Shibly: Conceptualization, Experiment Design, Reviewing and performed critical revision, and overall supervision of this study.

All authors provided critical feedback and helped shape the research, analysis, and manuscript.

References

1. Robinson, M.C. An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953. Transactions of the Royal Society of Tropical Medicine and Hygiene, **1955**, 49(1), 28-32.
2. Cavrini, F.; Gaibani, P.; Pierro, A.M.; Rossini, G.; Landini, M.P.; Sambri, V. Chikungunya: an emerging and spreading arthropod-borne viral disease. The Journal of Infection in Developing Countries, **2009**, 3(10), 744-52.
3. Thiboutot, M.M.; Kannan, S.; Kawalekar, O.U.; Shedlock, D.J.; Khan, A.S.; Sarangan, G.; Srikanth, P.; Weiner, D.B.; Muthumani, K. Chikungunya: a potentially emerging epidemic?. PLoS neglected tropical diseases, **2010**, 4(4), 623.
4. Staples, J.E.; Breiman, R.F.; Powers, A.M. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. Clinical infectious diseases, **2009**, 49(6), 942-8.
5. Townson, H.; Nathan, M.B. Resurgence of chikungunya. Transactions of the Royal Society of Tropical Medicine and Hygiene, **2008**, 102(4), 308-9.
6. Singh, S.K.; Unni, S.K. Chikungunya virus: host pathogen interaction. Reviews in medical virology, **2011**, 21(2), 78-88.
7. Tomasello, D.; Schlagenhauf, P.; Chikungunya and dengue autochthonous cases in Europe, 2007–2012. Travel medicine and infectious disease, **2013**, 11(5) 274-84
8. Petersen, E.; Wilson, M.E; Touch, S.; McCloskey, B.; Mwaba, P.; Bates, M.; Dar, O.; Mattes, F.; Kidd, M.; Ippolito, G.; Azhar, E.I. Rapid spread of Zika virus in the Americas-implications for public health preparedness for mass gatherings at the 2016 Brazil Olympic Games. International Journal of Infectious Diseases, **2016**, 44, 11-5.
9. Schwartz, O.; Albert, M.L. Biology and pathogenesis of chikungunya virus. Nature Reviews Microbiology, **2010**, 8(7), 491-500.
10. Bodenmann, P.; Althaus, F.; Burnand, B.; Vaucher, P.; Pécoud, A.; Genton, B. Medical care of asylum seekers: a descriptive study of the appropriateness of nurse practitioners' care compared to traditional physician-based care in a gatekeeping system. BMC public health, **2007**, 7(1), 310.
11. Ozden, S.; Huerre, M.; Riviere, J.P.; Coffey, L.L, Afonso, P.V.; Mouly, V.; De Monredon, J.; Roger, J.C.; Amrani, M.; Yvin, J.L.; Jaffar, M.C. Human muscle satellite cells as targets of Chikungunya virus infection. PloS One, **2007**, 2(6), 527.
12. Sudeep, A.B.; Parashar, D. Chikungunya: an overview. Journal of biosciences, **2008**, 33(4), 443-9.
13. Rashad, A.A.; Mahalingam, S.; Keller, P.A. Chikungunya virus: emerging targets and new opportunities for medicinal chemistry. Journal of Medicinal Chemistry, **2013**, 57(4), 1147-66.
14. Chandak, N.H.; Kashyap, R.S.; Kabra, D.; Karandikar, P.; Saha, S.S.; Morey, S.H.; Purohit, H.J.; Taori, G.M.; Daginawala, H.F. Neurological complications of Chikungunya virus infection. Neurology India, **2009**, 57(2), 177.
15. Mahendradas, P.; Avadhani, K.; Shetty, R. Chikungunya and the eye: a review. Journal of ophthalmic inflammation and infection, **2013**, 3(1), 35.

16. Ross, R.W. The Newala epidemic: III. The virus: isolation, pathogenic properties and relationship to the epidemic. Epidemiology & Infection, **1956**, 54(2), 177-91.
17. Hammon, W.M.; Rundnick, A.; Sather, G.E. Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. Science, **1960**, 131(3407), 1102-3.
18. Charrel, R.N.; de Lamballerie,X.; Raoult, D. Chikungunya outbreaks-the globalization of vectorborne diseases. New England Journal of Medicine, **2007**, 356(8), 769.
19. Parashar, D.; Cherian, S. Antiviral perspectives for chikungunya virus. BioMed Research International, **2014**.
20. Weaver, S.C; Osorio, J.E.; Livengood, J.A.; Chen, R.; Stinchcomb, D.T. Chikun-gunya virus and prospects for a vaccine. Expert Review of Vaccines, **2012**, 11, 1087–1101.
21. Rezza, G.; Nicoletti, L.; Angelini, R.; Romi, R.; Finarelli, A.C.; Panning, M.; Cordioli, P.; Fortuna, C.; Boros, S.; Magurano, F.; Silvi, G. Infection with chikungunya virus in Italy: an outbreak in a temperate region. The Lancet, **2007**, 370(9602), 1840-6.
22. Enserink, M. Infectious diseases. Chikungunya: no longer a third world disease. Science, **2007**, 318, 1860–1.
23. Pialoux, G.; Gaüzère, B.A.; Jauréguiberry, S.; Strobel, M. Chikungunya, an epidemic arbovirosis. The Lancet infectious diseases, **2007**, 7(5), 319-27.
24. Burt, F.J.; Rolph, M.S.; Rulli, N.E.; Mahalingam, S.; Heise, M.T. Chikungunya: a re-emerging virus. The Lancet, **2012**, 379(9816), 662-71.
25. Sun, S.; Xiang, Y.; Akahata, W.; Holdaway, H.; Pal, P.; Zhang, X.; Diamond, M.S.; Nabel, G.J.; Rossmann, M.G. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. Elife, **2013**, 2.
26. Simizu, B.; Yamamoto, K.; Hashimoto, K.; Ogata, T. Structural proteins of Chikungunya virus. Journal of virology, **1984**, 51(1), 254-8.
27. Khan, A.H.; Morita, K.; del Carmen Parquet, M.; Hasebe, F.; Mathenge, E.G.; Igarashi, A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. Journal of General Virology, **2002**, 83(12), 3075-84.
28. Jose, J.; Snyder, J.E.; Kuhn, R.J. A structural and functional perspective of alphavirus replication and assembly. Future microbiology, **2009**, 4(7), 837-56.
29. Strauss, J.H.; Strauss, E.G. The alphaviruses : gene expression, replication, and evolution. Microbiological Reviews, **1994**, 58(3), 491–562.
30. Voss, J.E.; Vaney, M.C.; Duquerroy, S.; Vonrhein, C.; Girard-Blanc, C.; Crublet, E.; Thompson, A.; Bricogne, G.; Rey, F.A. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature, **2010**, 468(7324), 709-12.
31. Gould, E.A.; Coutard, B.; Malet, H.; Morin, B.; Jamal, S.; Weaver, S.; Gorbalenya, A.; Moureau, G.; Baronti, C.; Delogu, I.; Forrester, N. Understanding the alphaviruses: recent research on important emerging pathogens and progress towards their control. Antiviral research, **2010**, 87(2), 111-24.
32. Wang, Y.F.; Sawicki, S.G.; Sawicki.; D.L. Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. Journal of virology, **1994**, 68(10), 6466-75.
33. Rungrotmongkol, T.; Nunthaboot, N.; Malaisree, M.; Kaiyawet, N.; Yotmanee, P.; Meeprasert, A.; Hannongbua, S. Molecular insight into the specific binding of ADP-ribose

to the nsP3 macro domains of chikungunya and Venezuelan equine encephalitis viruses: molecular dynamics simulations and free energy calculations. Journal of Molecular Graphics and Modelling, **2010**, 29(3), 347-53.

34. Couderc, T.; Khandoudi, N.; Grandadam, M.; Visse, C.; Gangneux, N.; Bagot, S.; Prost, J.F.; Lecuit, M. Prophylaxis and therapy for Chikungunya virus infection. The Journal of infectious diseases, **2009,** 200(4), 516-23.
35. Her, Z.; Malleret, B.; Chan, M.; Ong, E.K.; Wong, S.C.; Kwek, D.J.; Tolou, H.; Lin, R.T.; Tambyah, P.A.; Rénia, L.; Ng, L.F. Active infection of human blood monocytes by Chikungunya virus triggers an innate immune response. The Journal of Immunology, **2010**, 184(10), 5903-13.
36. Labadie, K.; Larcher, T.; Joubert, C.; Mannioui, A.; Delache, B.; Brochard, P.; Guigand, L.; Dubreil, L.; Lebon, P.; Verrier, B.; de Lamballerie, X. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. The Journal of clinical investigation, **2010**, 120(3), 894-906.
37. Poggesi, I.; Benedetti, M.S.; Whomsley, R.; Lamer, S.L.; Molimard, M.; Watelet, J.B. Pharmacokinetics in special populations. Drug metabolism reviews, 2009, 41(3), 422-54.
38. Patronov, A.; Doytchinova, I. T-cell epitope vaccine design by immunoinformatics. Open biology, **2013**, 3(1), 120-139.
39. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M. UniProt: the universal protein knowledgebase Nucleic Acids Res, **2004**, 32, D115 D119.
40. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis work bench. Bioinformatics, **2009**, 25(9, 1189-91.
41. Doytchinova, I.A.; Flower, D.R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics, **2007**, 8(1), 4.
42. Larsen, M.V.; Lundegaard, C.; Lamberth, K.; Buus, S.; Lund, O.; Nielsen, M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinformatics, **2007**, 8(1), 424.
43. Larsen, M.V.; Lundegaard, C.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O.; Nielsen, M. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. European journal of immunology, **2005**, 35(8), 2295-303.
44. Tenzer, S.; Peters, B.; Bulik, S.; Schoor, O.; Lemmel, C.; Schatz, M.M.; Kloetzel, P.M.; Rammensee, H.G.; Schild, H.; Holzhütter, H.G. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cellular and Molecular Life Sciences CMLS, **2005**, 62(9), 1025-37.
45. Peters, B.; Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BioMed Central, **2005**, 6(1), 132.
46. Nielsen, M.; Lundegaard, C.; Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BioMed Central, **2007**, 8(1), 238.

47. Von Heijne, G. A day in the life of Dr K. or how I learned to stop worrying and love lysozyme: a tragedy in six acts. Journal of molecular biology, **1999**, 293(2), 367-79.
48. Krogh, A.; Larsson, B.; Von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. Journal of molecular biology, **2001**, 305(3), 567-80.
49. Alimentarius, C. Codex principles and guidelines on foods derived from biotechnology. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organisation: Rome, **2003**.
50. Stadler, M.B.; Stadler, B.M. Allergenicity prediction by protein sequence. The FASEB Journal, **2003**, 17(9), 1141-3.
51. Saha, S.; Raghava, G.P. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research, **2006**, 34(2), 202-9.
52. Nair, D.T.; Singh, K.; Siddiqui, Z.; Nayak, B.P.; Rao, K.V.; Salunke, D.M. Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. The Journal of Immunology, **2002**, 168(5), 2371-82.
53. Emini, E.A.; Hughes, J.V.; Perlow, D.; Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology, **1985**, 55(3), 836-9.
54. Kolaskar, A.S.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters, **1990**, 276(1-2), 172-4.
55. Parker, J.M.; Guo, D.; Hodges, R.S. 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, **1986**, 25(19), 5425-32.
56. Karplus, P.A.; Schulz, G.E. Prediction of chain flexibility in proteins. The Science of Nature, **1985**, 72(4), 212-3.
57. Chou, P.Y.; Fasman, G.D. Empirical predictions of protein conformation. Annual review of biochemistry, **1978**, 47(1), 251-76.
58. Rini, J.M.; Schulze-Gahmen, U.; Wilson, I.A. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science, **1992**, 255(5047), 959-65.
59. Bui, H.H.; Sidney, J.; Li, W.; Fusseder, N.; Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BioMed Central, **2007**, 8(1), 361.
60. Sievers, F.; Wilm A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J.D.; Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **2011,** 7:539
61. Bui, H.H.; Sidney, J.; Dinh, K.; Southwood, S.; Newman, M.J.; Sette, A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC bioinformatics, **2006**, 7(1), 153.

62. Berman, H.M.; West brook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne P.E. The Protein Data Bank. Nucleic Acids Research, **2000**,28, 235-42.
63. Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. Nucleic acids research, **2015**, 43(1), 419-24.
64. Blaszczyk, M.; Kurcinski, M.; Kouza, M.; Wieteska, L.; Debinski, A.; Kolinski, A.; Kmiecik, S. Modeling of protein–peptide interactions using the CABS-dock web server for binding site search and flexible docking. Methods, **2016**, 93, 72-83.
65. Krieger, E.; Darden, T.; Nabuurs, S.B.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins: Structure, Function, and Bioinformatics. **2004** Dec 1;57(4):678-83..
66. Case, D.A.; Darden, T.A.; Cheatham, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Pearlman, D.A.; Crowley, M.; Walker, R.C. AMBER 9.0 Software package.
67. Krieger, E.; Nielsen, J.E.; Spronk, C.A.; Vriend, G. Fast empirical pKa prediction by Ewald summation. Journal of molecular graphics and modeling,**2006**, 1;25(4):481-6.
68. Krieger, E.; Vriend, G. New ways to boost molecular dynamics simulations. Journal of computational chemistry. 2015 May 15;36(13):996-1007.
69. Björklund, Å.K.; Soeria-Atmadja, D.; Zorzet, A.; Hammerling, U.; Gustafsson, M.G. Supervised identification of allergen-representative peptides for in silico detection of potentially allergenic proteins. Bioinformatics, **2005**, 21(1), 39-50.
70. Rose, G.D.; Glerasch, L.M.; Smith, J.A. Turns in peptides and proteins. In Advances in protein chemistry, **1985**, 37 1-109.
71. Novotný, J.; Handschumacher, M.; Haber, E.; Bruccoleri, R.E.; Carlson, W.B.; Fanning, D.W.; Smith, J.A.; Rose, G.D. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). Proceedings of the National Academy of Sciences, **1986**, 83 (2), 226-30.
72. Islam, M.J.; Parves, M.R.; Mahmud, S.; Tithi, F.A.; Reza, M.A. Assessment of structurally and functionally high-risk nsSNPs impacts on human bone morphogenetic protein receptor type IA (BMPR1A) by computational approach. Computational biology and chemistry,**2019**, 1, 80:31-45.
73. Mahmud, S.; Parves, M.R.; Riza, Y.M.; Sujon, K.M.; Ray, S.; Tithi, F.A.; Zaoti, Z.F.; Alam, S.; Absar, N. Exploring the potent inhibitors and binding modes of phospholipase A2 through in silico investigation. Journal of Biomolecular Structure and Dynamics, **2019**, 22, 1-1.
74. Yang, C.; Jas, G.S.; Kuczera, K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, **2004**, 11, 1697(1-2), 289-300.
75. Shahbaaz, M.; Nkaule, A.; Christoffels, A. Designing novel possible kinase inhibitor derivatives as therapeutics against Mycobacterium tuberculosis: An in silico study. Scientific reports, **2019**, 13, 9(1), 1-2.

76. Groot, D.; Rappuoli, R. Genome-derived vaccines. Expert Review of Vaccines, **2004**, 3(1), 59–76.
77. Marshall, S.J. Developing countries face double burden of disease. Bull. World Health Organization, **2004**, 82 (7), 556.
78. Sabetian, S.; Nezafat, N.; Dorosti, H.; Zarei, M.; Ghasemi, Y. Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus, Journal of Biomolecular Structure and Dynamics, **2019**, 37:10, 2546-2563.
79. Trainor, N.B.; Crill, W.D.; Roberson, J.A.; Chang, G.J. Mutation analysis of the fusion domain region of St. Louis encephalitis virus envelope protein. Virology, **2007**, 360 (2), 398–406.
80. Ravichandran, L.; Venkatesan, A.; Febin Prabhu Dass, J. Epitope-based immunoinformatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV). J Cell Biochem, **2018**.
81. Khan, M.A.; Hossain, M.U.; Rakib-Uz-Zaman, S.M.; Morshed, M.N. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. Scand J Immunol, **2015**, 82, 25–34.
82. Hasan, M.;Islam, S.;Chakraborty, S.; Mustafa,A.H.; Azim, K.F.; Joy, Z.F.; Hossain,M.N.; Foysal, S.H.; Hasan, M.N. Contriving a chimeric polyvalent vaccine to prevent infections caused by herpes simplex virus (type-1 and type-2): an exploratory immunoinformatic approach. Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2019.1647286
83. Wu, M.; Li, M.; Yue, Y.; and Xu, W. DNA vaccine with discontinuous T-cell epitope insertions into HSP65 scaffold as a potential means to improve immunogenicity of multi-epitope Mycobacterium tuberculosis vaccine. Microbiology and Immunology, **2016**, 60(9), 634-645.
84. Azim, K.F., Hasan, M., Hossain, M.N., Somana, S.R., Hoque, S.F., Bappy, M.N., Chowdhury, A.T., Lasker, T. Immunoinformatics approaches for designing a novel multi epitope peptide vaccine against human norovirus (Norwalk virus). Infection, Genetics and Evolution, **2019,** 1(74), 103936.
85. Hasan, M.A.; Khan, M.A.; Dattaa, A.; Mazumder, M.H.H.; Hossain, M.U. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. Molecular Immunology, **2015**, 65, 189–204
86. Hossain M.U.; Keya, C.A.; Das, K.C.; Hashem, A.; Omar, T.M.; Khan, M.A.; Rakib U. Zaman, S.M.; Salimullah, M. An immunopharmacoinformatics approach in development of vaccine and drug candidates for West Nile virus. Front Chem, **2018**, 6, 246.
87. Oliveira Tosta, S.F.; Passos, M.S.; Kato, R.; Salgado, A.; Xavier, J.; Jaiswal, A.K.; Soares, S.C.; Azevedo, V.; Giovanetti, M.; Tiwari, S.; Alcantara, L.C.J. Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches, Journal of Biomolecular Structure and Dynamics, **2020**, DOI: 10.1080/07391102.2019.1707120
88. Priyadarshini, V.; Pradhan, D.; Munikumar, M.; Swargam, S.; Umamaheswari, A.; Rajasekhar, D. Genome-based approaches to develop epitope-driven subunit vaccines

against pathogens of infective endocarditis, Journal of Biomolecular Structure and Dynamics, **2014**, 32:6, 876-889.

89. Aziz Ul-Rahman; Shabbir, M.A.B. In silico analysis for development of epitopes-based peptide vaccine against Alkhurma hemorrhagic fever virus, Journal of Biomolecular Structure and Dynamics, **2019**, DOI: 10.1080/07391102.2019.1651673
90. Nain, Z.; Abdulla, F.; Rahman, M.M.; Karim, M.M.; Khan, M.S.A.; Sayed, S.B.; Mahmud, S.; Rahman, S.M.R.; Sheam, M.M.; Haque, Z.; Adhikari, U.K. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis* using immunoinformatic approaches, Journal of Biomolecular Structure and Dynamics, **2019**, DOI: 10.1080/07391102.2019.1692072

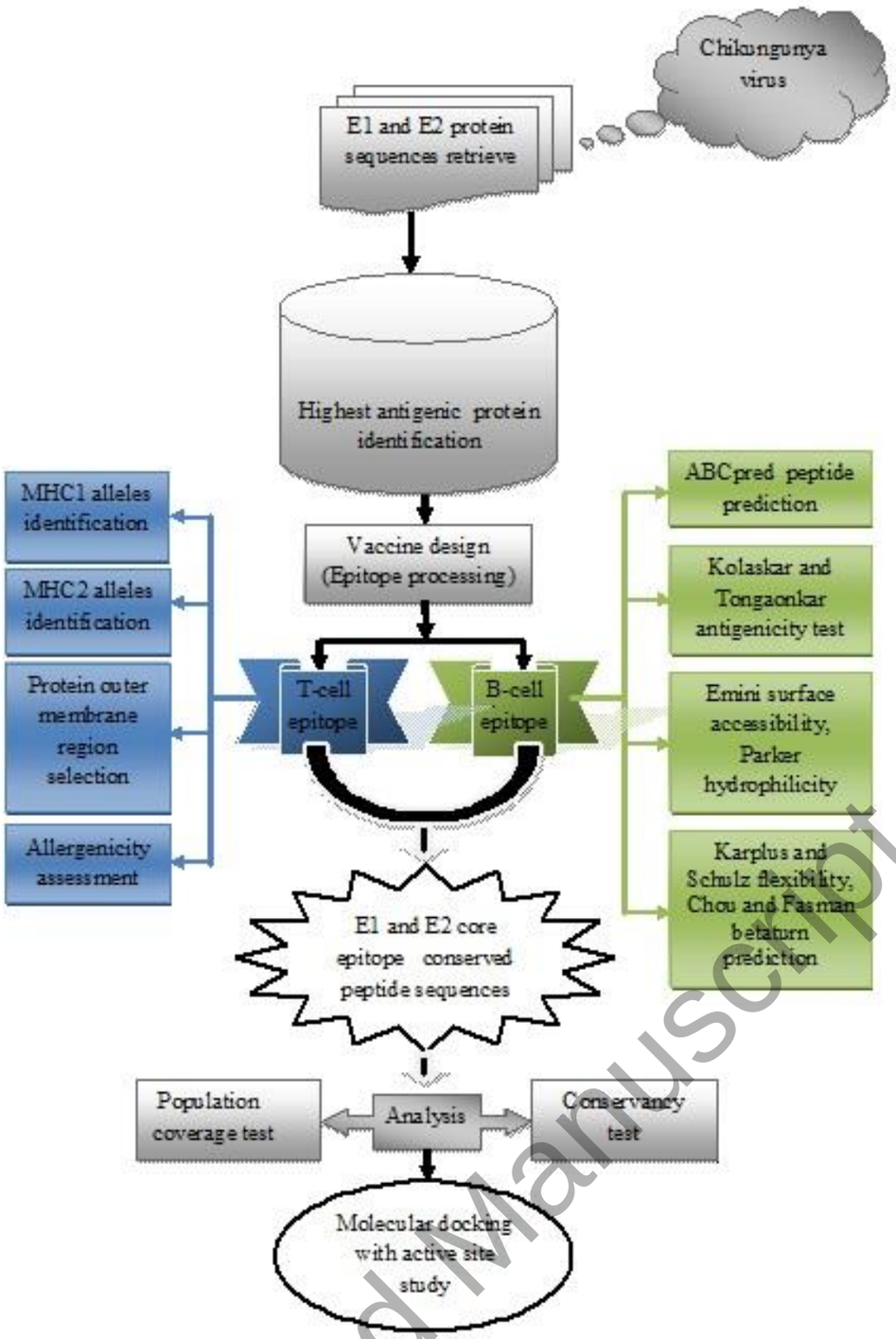


Fig. (1). Graphical representation of Epitope based Peptide vaccine design against Chikungunya virus

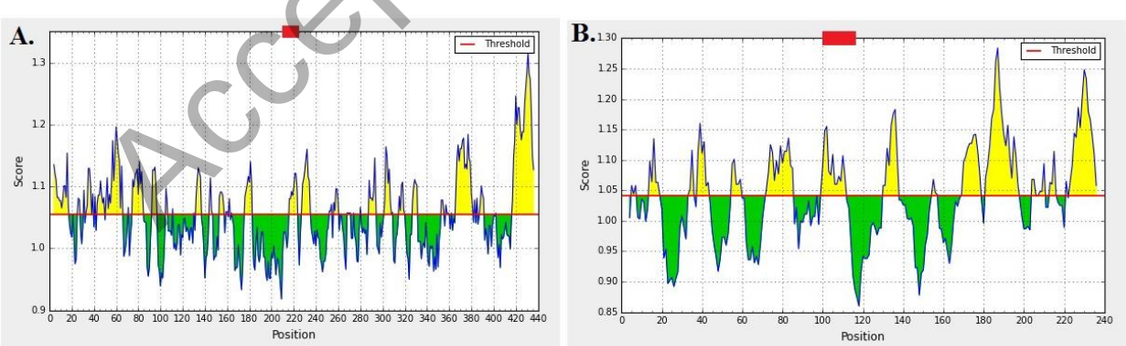


Fig. (2). Kolaskar and Tongaonkar antigenicity prediction result of the E1 envelope glycoprotein and E2 envelope glycoprotein. The X- and Y-axis represent the sequence position and antigenic propensity score, respectively. **A.** Here, the minimum and maximum propensity

scores are 0.918 and 1.316. The yellow regions above the threshold (1.055) are antigenic. The selected peptide region (SEDVYANTQLVLQRP) marked as red. **B.** Here, the minimum and maximum propensity scores are 0.860 and 1.284. The yellow regions above the threshold (1.041) are antigenic. The selected peptide region (IMLLYPDHPTLLSYR) marked as red.

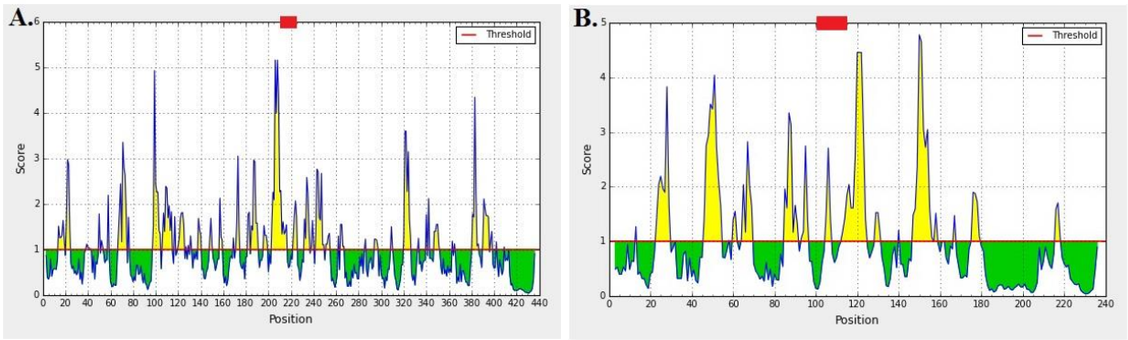


Fig. (3). Emini surface accessibility prediction result of the E1 envelope glycoprotein and E2 envelope glycoprotein. The X- and Y-axis represent the sequence position and surface probability, respectively. **A.** Here, the minimum and maximum propensity scores are 0.037 and 5.167. The yellow regions above the threshold (1.00) are surface accessible. The selected peptide region (SEDVYANTQLVLQRP) marked as red. **B.** Here, the minimum and maximum propensity scores are 0.034 and 4.785. The yellow regions above the threshold (1.00) are surface accessible. The selected peptide region (IMLLYPDHPTLLSYR) marked as red.

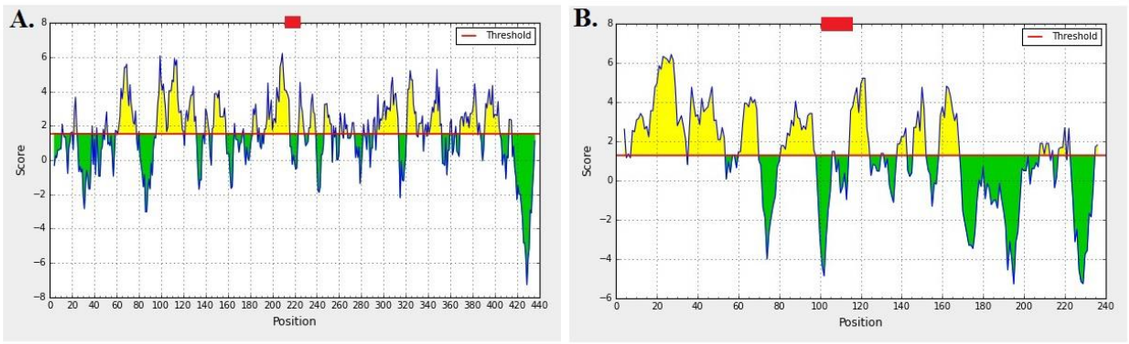


Fig. (4). Parker hydrophilicity prediction result of the E1 envelope glycoprotein and E2 envelope glycoprotein. The X- and Y-axis represent the sequence position and hydrophilicity scale, respectively. **A.** Here, the minimum and maximum propensity scores are -7.286 and 6.229. The yellow regions above the threshold (1.553) are hydrophilic. The selected peptide region (SEDVYANTQLVLQRP) marked as red. **B.** Here, the minimum and maximum propensity scores are -5.286 and 6.443. The yellow regions above the threshold (1.288) are hydrophilic. The selected peptide region (IMLLYPDHPTLLSYR) marked as red.

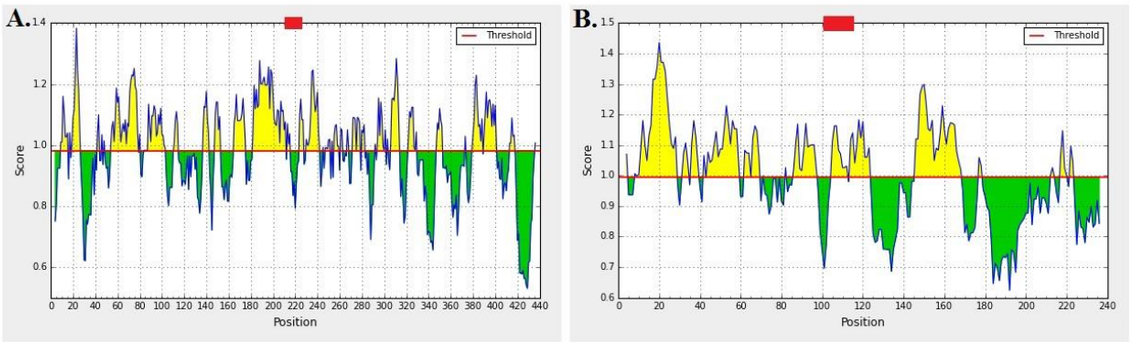


Fig. (5).Chou and Fasman beta-turn prediction result of the E1 envelope glycoprotein and E2 envelope glycoprotein. The X- and Y-axis represent the sequence position and turn score, respectively. **A.** Here, the minimum and maximum propensity scores are 0.530 and 1.383. The yellow regions above the threshold (0.982) are good. The selected peptide region (SEDVYANTQLVLQRP) marked as red. **B.** Here, the minimum and maximum propensity scores are 0.624 and 1.436. The yellow regions above the threshold (0.995) are good. The selected peptide region (IMLLYPDHPTLLSYR) marked as red.

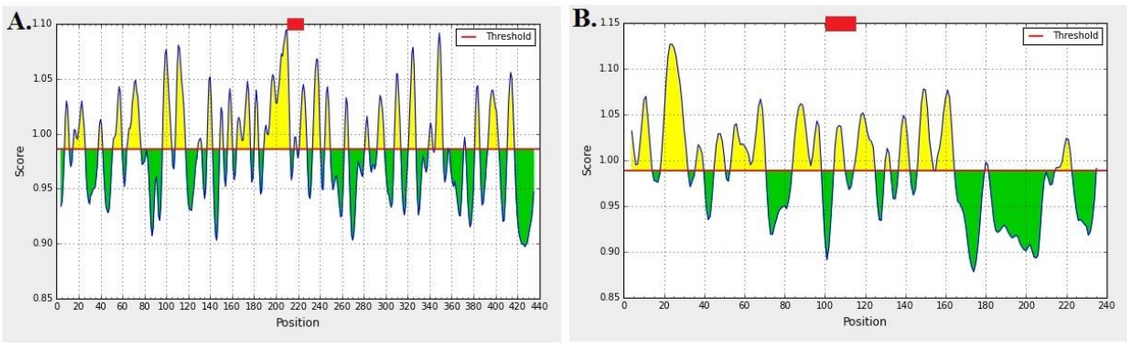
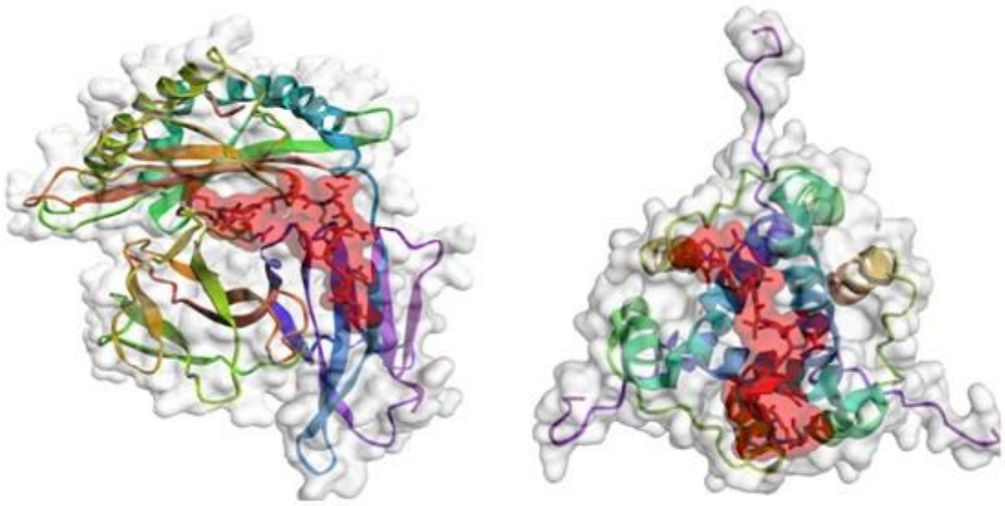
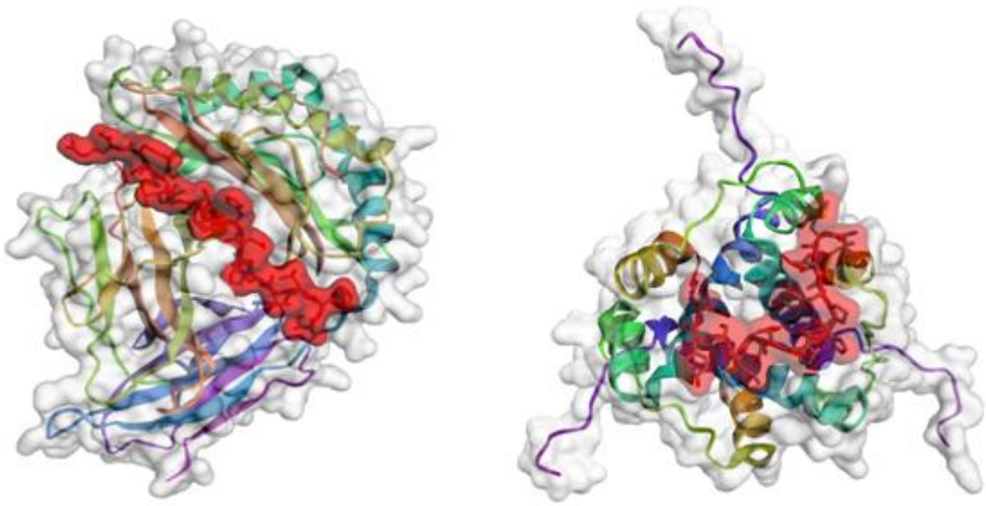


Fig. (6). Karplus and Schulz flexibility prediction result of E1 envelope glycoprotein and E2 envelope glycoprotein. Here, X-axis and Y-axis represent sequence position and flexibility score, respectively. **A.** Here, the minimum and maximum propensity scores are 0.897 and 1.095. The flexible yellow regions above the threshold (0.986) are shown in yellow. The selected peptide region (SEDVYANTQLVLQRP) marked as red. **B.** Here, the minimum and maximum propensity scores are 0.878 and 1.127. The yellow regions above the threshold (0.989) are flexible. The selected peptide region (IMLLYPDHPTLLSYR) marked as red.



(I) (II)

Fig. (7). Binding confirmation of predicted E1 envelope glycoprotein epitope to both MHC-I and MHC-II alleles. (I) epitope and MHC-I allele interaction and (II) epitope and MHC-II allele interaction.



(I) (II)

Fig. (8). Binding confirmation of predicted envelope glycoprotein E2 epitope to both MHC-I and MHC-II alleles.(I) epitope and MHC-I allele interaction and (II) epitope and MHC-II allele interaction.

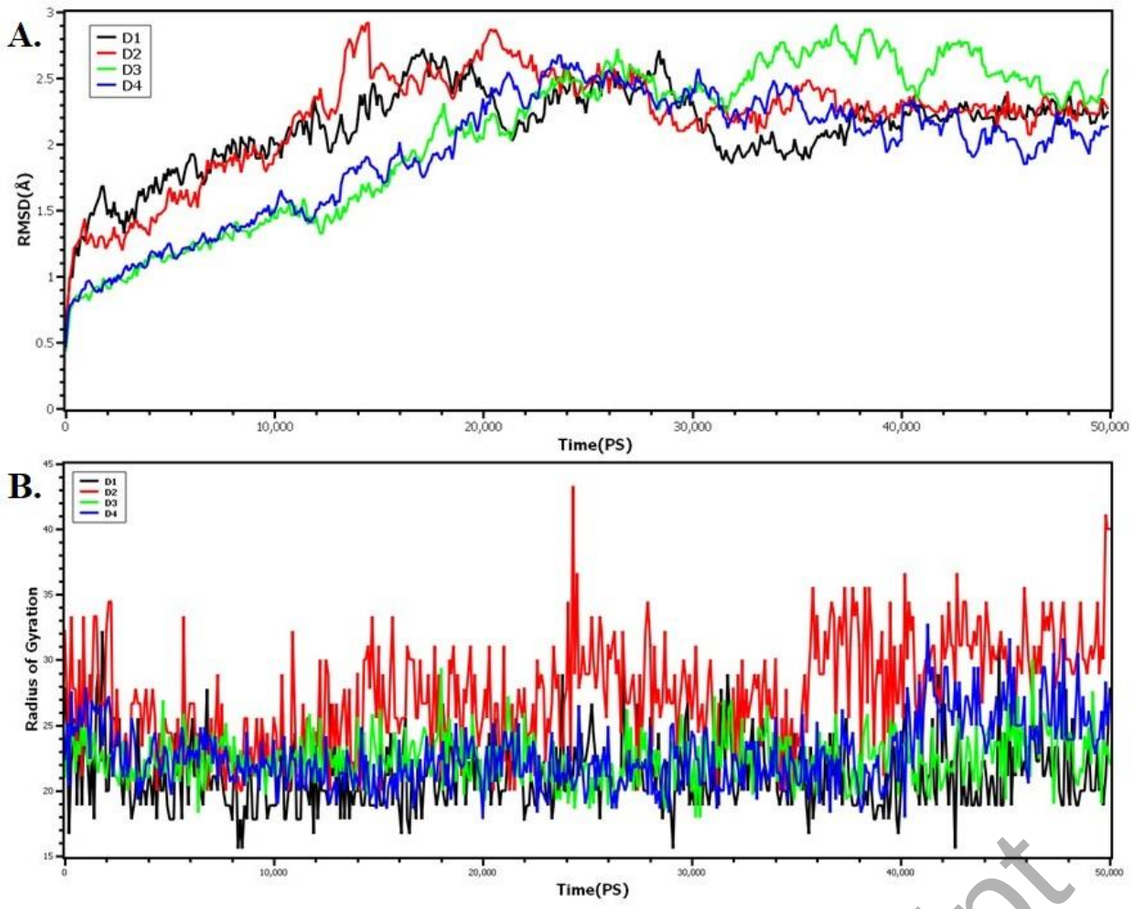


Fig. (9). (A) complex stability checked by RMSD (root mean square deviation), (B) rigidness of four biological system (D1, D2, D3 and D4 complex) evaluated through radius of gyration or Rg (50ns).

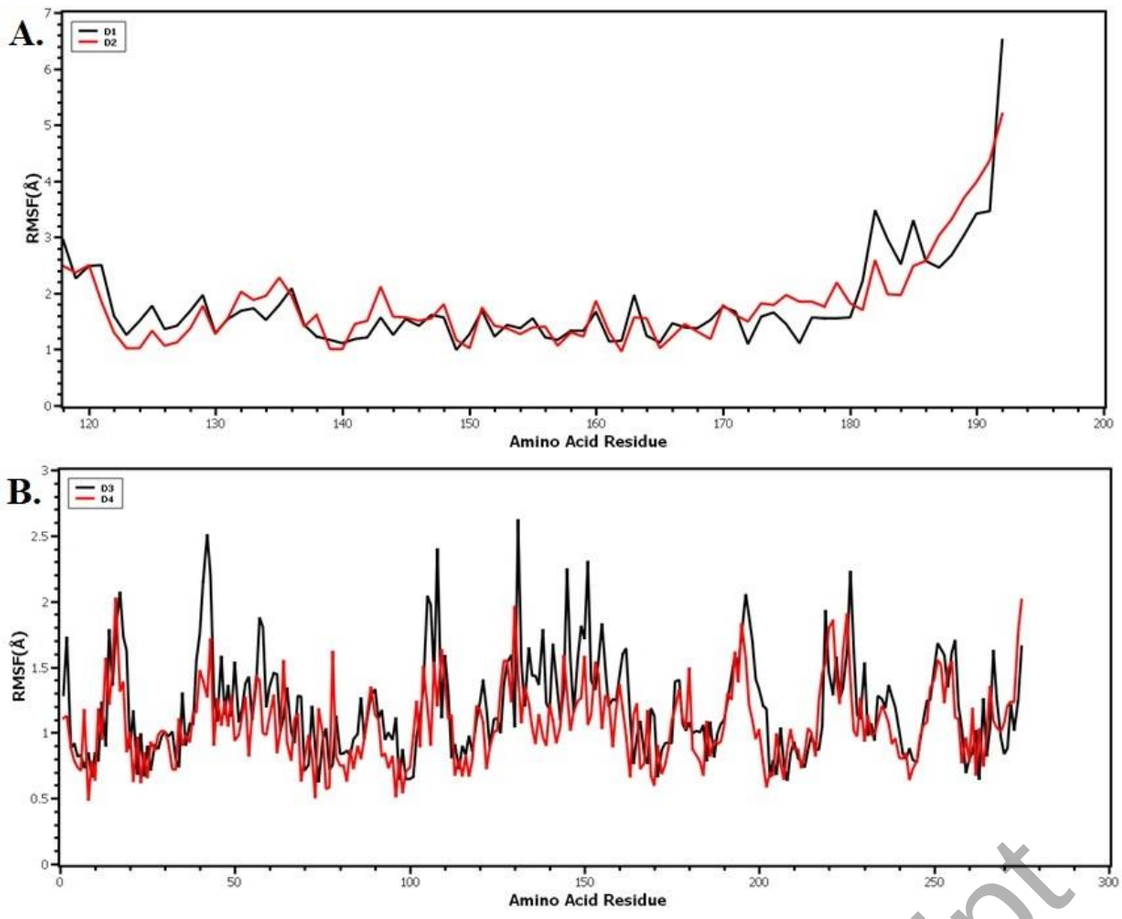


Fig. (10). Complex flexibility assessed by Root mean square fluctuation (rmsf) per residue in molecular dynamics simulation.

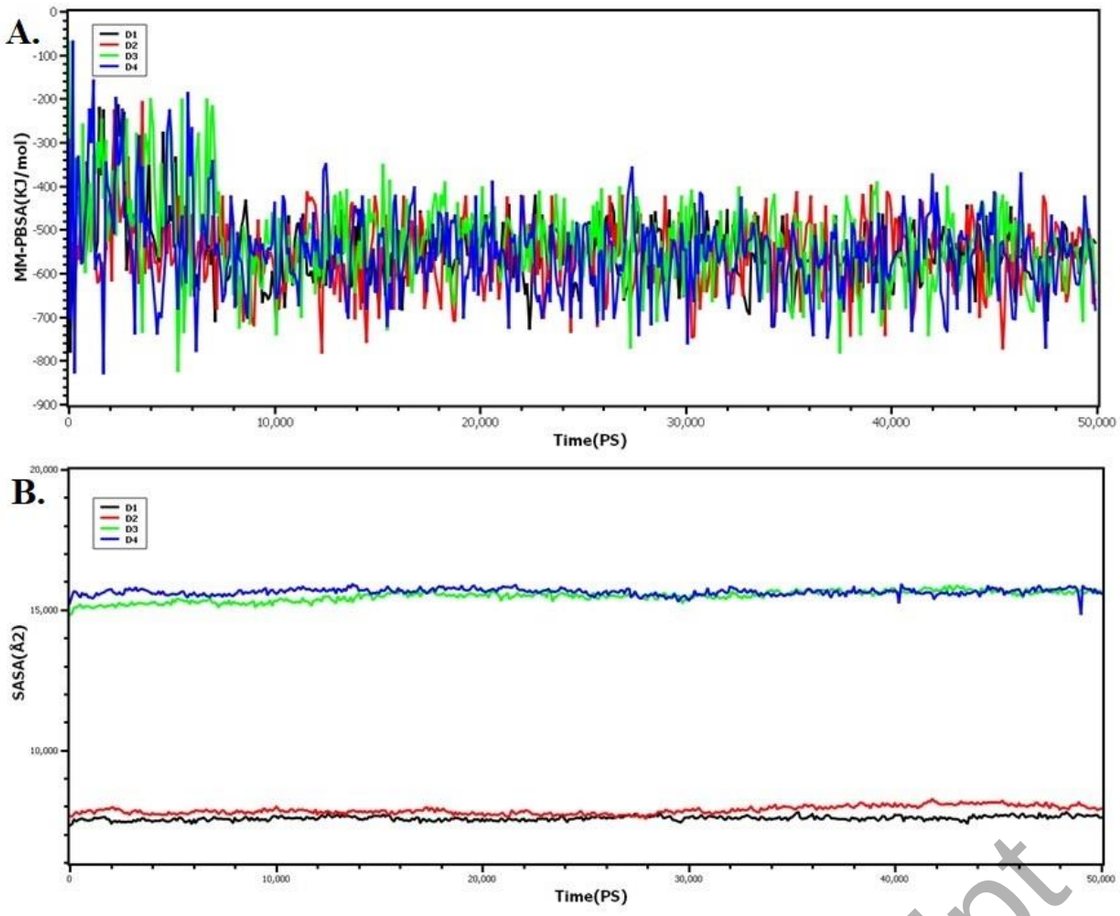


Fig. (11). Illustration of (A) MM-PBSA results in the form of different free binding energies of the complex, (B) changes occur in the surface area of four complex during simulation time (SASA values).

Table 1. MHC-I alleles and IC50 score for T cell epitope processing prediction of E1 Envelope glycoprotein.

Epitope	MHC Allele (IC50)	No. of allele
SLDYITCEY	HLA-C*12:03 (1.81)(7.6); HLA-C*05:01(1.57)(13.2); HLA-C*14:02(0.84)(71.7); HLA-A*29:02(0.44)(177.7); HLA-B*15:02(0.44)(178.3); HLA-A*01:01(0.4)(196.1);	06
ALASAEFRV	HLA-A*02:01(-0.32)(28.3); HLA-A*02:06(-0.66)(62.3); HLA-C*12:03(-0.67)(64.3)	03
AIIKYTASK	HLA-C*12:03(-0.28)(25.7); HLA-A*11:01(-0.44)(37.2); HLA-C*03:03(-0.71)(68.5); HLA-A*03:01(-1.13)(179); HLA-A*30:01(-1.22)(221.8)	05
VYNMDYPPF	HLA-C*14:02(1.58)(4.3); HLA-C*07:02(0.74)(29.5); HLA-A*23:01(0.62)(39); HLA-A*24:02(0.35)(72.7); HLA-C*03:03(0.32)(78.1); HLA-C*12:03(0.28)(85.7)	06
STALASAEF	HLA-C*03:03(1.08)(21.9); HLA-B*15:02(0.82)(40.3); HLA-A*32:01(0.34)(119.7); HLA-B*15:01(0.31)(129.4); HLA-C*12:03(0.29)(134.3)	05
RPGYSPMVL	HLA-B*07:02(0.56)(25.6); HLA-C*03:03(0.44)(3.1); HLA-C*12:03(0.19)(61.1); HLA-B*15:02(-0.28)(177.5)	04
WLKERGASL	HLA-B*08:01(0.73)(11.1); HLA-C*12:03(0.48)(19.9); HLA-B*15:02(0.33)(27.7); HLA-C*03:03(0.03)(55.7); HLA-C*14:02(-0.32)(124.6)	05
GQFGDIQSR	HLA-C*12:03(0.5)(13.8); HLA-C*03:03(-0.39)(108.3); HLA-A*31:01(-0.61)(180.5)	03
VYANTQLVL	HLA-C*14:02(1)(9.1); HLA-C*07:02(0.27)(48.2); HLA-B*15:02(0.21)(55.5); HLA-C*03:03(0)(91.1); HLA-C*12:03(-0.05)(101.2); HLA-A*23:01(-0.37)(213.5)	06
IEVEGNSQL	HLA-C*03:03(0.37)(23); HLA-B*15:02(0.28)(28.1); HLA-B*40:01(0.1)(42.4); HLA-C*12:03(-0.12)(71.3)	04
FTGVYPFMW	HLA-B*58:01(0.41)(21.1); HLA-C*12:03(0.34)(25); HLA-C*03:03(-0.14)(74.7); HLA-B*57:01(-0.24)(93.9); HLA-B*53:01(-0.42)(141.9)	05
SQAPSGFKY	HLA-B*15:02(1.01)(39.7); HLA-C*12:03(0.99)(41.4); HLA-A*30:02(0.75)(70.9); HLA-C*03:03(0.59)(102.9); HLA-B*15:01(0.59)(104.5); HLA-A*29:02(0.53)(119.1)	06

Table 2. MHC-I alleles and IC50 score for T-cell epitope processing prediction of Envelope glycoprotein E2.

Epitope	MHC-1 Allele (IC50)	No. of allele
GTNHKKWQY	HLA-C*05:01(1.1)(31); HLA-C*12:03(1.05)(34.5); HLA-A*30:02(0.38)(162.4); HLA-A*29:02(0.23)(227.9)	04
FLLSLICCI	HLA-A*02:01(0.68)(7.8); HLA-C*03:03(0.43)(13.9); HLA-A*02:06(0.19)(24.2); HLA-C*14:02(-0.08)(45.5); HLA-C*12:03(-0.46)(108.4); HLA-B*15:02(-0.66)(169.9)	06
TVNGQTVGY	HLA-C*05:01(1.67)(12.4); HLA-C*12:03(1.33)(27.5); HLA-A*29:02(0.75)(103.9); HLA-B*35:01(0.69)(120.7) HLA-C*07:01(0.63)(139); HLA-C*03:03(0.58)(155.8) HLA-A*30:02(0.48)(195.2)	07
LYPDHPTLL	HLA-C*14:02(1.01)(10.6); HLA-C*12:03(0.33)(50.7) HLA-B*15:02(0.12)(82.3); HLA-C*06:02(0.01)(107.8) HLA-A*24:02(-0.14)(151.6);HLA-C*07:02(-0.29)(213.4); HLA-C*07:01(-0.32)(229.2)	07
YPTMTVVVV	HLA-C*12:03(-0.01)(22.1); HLA-C*03:03(-0.72)(114.2); HLA-B*07:02(-0.82)(144.2)	03
VMHKKEVVL	HLA-C*14:02(0.41)(48.8); HLA-C*03:03(0.08)(105.8); HLA-C*12:03(-0.18)(189.3)	03
RRCITPYEL	HLA-C*07:02(0.37)(68.4); HLA-C*12:03(0.35)(72.5); HLA-C*03:03(0.19)(103.9); HLA-B*39:01(0.09)(131.3); HLA-B*15:02(-0.04)(175.9)	05
YYYELYPTM	HLA-C*14:02(1.4)(1.1); HLA-C*07:02(0.41)(10.3); HLA-C*03:03(0.21)(16.3); HLA-C*12:03(0.03)(24.8); HLA-A*24:02(-0.45)(74.6); HLA-A*23:01(-0.5)(84.4); HLA-B*15:02(-0.58)(101.7); HLA-C*06:02(-0.78)(160.3); HLA-B*35:01(-0.85)(187.3)	09
YELYPTMTV	HLA-C*12:03(-0.32)(27.4); HLA-C*03:03(-0.59)(51); HLA-B*40:02(-0.89)(101.4); HLA-B*40:01(-1.04)(143.2); HLA-A*02:06(-1.19)(203)	05
AAGTNHKKW	HLA-C*12:03(1.02)(8.2); HLA-C*03:03(0.58)(22.7); HLA-B*57:01(0.27)(46.3); HLA-B*58:01(-0.19)(134.1) HLA-C*15:02(-0.33)(184.3)	05
KARNPTVTY	HLA-C*12:03(1.97)(7.5); HLA-A*30:01(1.09)(56.4); HLA-B*15:02(0.62)(166.4); HLA-A*30:02(0.62)(168.8); HLA-B*15:01(0.51)(214.3);	05

Table 3. Peptide for epitope and MHC-II alleles depending on IC50 score of CHIKV E1 Envelope glycoprotein

Core epitope	**Peptide**	**MHC-II Allele(IC50)**	**No. of allele**
AIIKYTASK	DFGGVAIIKYTASKK	HLA-DRB5*01:01(47);	01
WLKERGASL	SGFKYWLKERGASLQ	HLA-DRB1*07:01(292); HLA-DRB1*01:01(41);	02
VYANTQLVL	SED**VYANTQLVL**QRP	HLA-DPA1*03:01/ DPB1*04:02(96); HLA-DPA1*02:01/ DPB1*01:01(38); HLA-DPA1*01:03/ DPB1*02:01(99); HLA-DRB1*13:02(284); HLA-DPA1*01/ DPB1*04:01(202)	05
IEVEGNSQL	REAEIEVEGNSQLQ	HLA-DPA1*02:01/ DPB1*01:01(140)	01
FTGVYPFMW	KVFTGVYPFMWGGAY	HLA-DRB5*01:01(91); HLA-DRB1*01:01(184)	02

Table 4. Peptide for epitope and MHC-II alleles depending on IC50 score of CHIKV Envelope glycoprotein E2.

Epitope	Peptide	MHC-II Allele(IC50)	No. of allele
FLLSLICCI	ATVPFLLSLICCIRT	HLA-DRB1*07:01(199); HLA-DRB1*01:01(152)	02
TVNGQTVGY	GNVKITVNGQTVGYK	HLA-DQA1*05:01/ DQB1*03:01	01
LYPDHPTLL	IML**LYPDHPTLL**SYR	HLA-DRB1*03:01(251); HLA-DRB1*07:01(243); HLA-DRB1*01:01(234); DRB3*01:01(333); HLA-DRB1*13:02(428);	05
YPTMTVVVV	YPTMTVVVVSVATFI	HLA-DRB5*01:01(276); HLA-DRB1*07:01(27); HLA-DRB1*01:01(80)	03
RRCITPYEL	RRCITPYELTPGATV	HLA-DRB1*09:01(105); HLA-DRB1*07:01(52)	
YYYELYPTM	IILYYYELYPTMTVV	HLA-DPA1*01:03/ DPB1*02:01(73); HLA-DPA1*01/ DPB1*04:01(220); HLA-DPA1*02:01/ DPB1*01:01(150); HLA-DRB1*04:01(136)	04
YELYPTMTV	YYYELYPTMTVVVVS	HLA-DRB1*09:01(138); HLA-DRB1*04:01(201); HLA-DRB1*04:05(275); HLA-DRB1*07:01(120);	05
KARNPTVTY	KARNPTVTYGKNQVI	HLA-DRB1*07:01(224); HLA-DRB1*13:02(277)	02

Table 5. B-cell epitope based Peptide sequence and their Artificial Neural Network (ANN) score of E1 and E2 Envelope glycoprotein from ABCpred server.

ANN of E1 Envelop Glycoprotein			
Rank	**Sequence**	**Start position**	**Score**
1	FKYWLKERGASLQH	240	0.89
2	SCKVFTGVYPFMWG MSSAWTPFDNKIVV	77 166	0.87
3	VVYKGDVYNMDYPP AHTASASAKLRVLY	178 124	0.86
4	LASAEFRVQVCSTQ	360	0.85
5	NSQLQISFSTALAS	349	0.84

	SMTNAVTIREAEIE HVPYSQAPSGFKYW	332 230	
6	SLDYITCEYKTVIP VLQRPSAGTVHVPY	43 220	0.83
7	GNMPISIDIPDAAF TPFDNKIVVYKGDV	274 171	0.82
8	QIATNPVRAMNCAV	260	0.81
9	ECKDKSLPDYSCKV AAYANGDHAVTVKD	67 145	0.79
10	PDAAFTRVVDAPSL	283	0.78
11	DFGGVAIIKYTASK EVSACTHSSDFGGV **VYANTQLVL**QRPSA YKTLVNRPGYSPMV QGNNITVAAYANGD	311 302 213 15 138	0.77
12	ISFSTALASAEFRV	253	0.76
ANN of E2 Envelop Glycoprotein			
1	GQTVGYKCNCGGSN	10	0.92
2	IHIPFPLANVTCRV	71	0.90
3	WVMHKKEVVLTVPT	126	0.89
4	KCNCGGSNEGQTTT	16	0.88
5	YRNMGEEPNYQEEW	113	0.83
6	KITVNGQTVGYKCN QTTTDKVINNCKVD HEIILYYYELYPTM	5 26 169	0.82
7	ELYPTMTVVVVSVA LTVPTEGLEVTWGN	177 135	0.81
8	WQYNSPLVPRNAEL	51	0.80
9	L**LYPDHPTLL**SYRN	102	0.79
10	GLEVTWGNNEPYKY	141	0.78

Table 6. Conservancy of different epitopes of CHIKV E1 envelope glycoprotein protein in different strain of this virus

Conservancy of E1 Envelop Glycoprotein			
Protein sequence	**Percent of protein sequence matches at identity <=100%**	**Minimum identity**	**Maximum identity**
DFGGVAIIKYTASKK	66.67% (6/26)	20.00%	100.00%
SGFKYWLKERGASLQ	92.31% (24/26)	20.00%	100.00%
SEDVYANTQLVLQRP	88.46% (23/26)	93.33%	100.00%
REAEIEVEGNSQLQI	73.08% (19/26)	26.67%	100.00%
KVFTGVYPFMWGGAY	53.85% (14/26)	20.00%	100.00%
Conservancy of E2 Envelop Glycoprotein			
GNVKITVNGQTVGYK	40.00% (14/35)	20.00%	100.00%
IMLLYPDHPTLLSYR	100.00% (35/35)	100.00%	100.00%
KARNPTVTYGKNQVI	94.29%(33/35)	93.33%	100.00%