

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/352076626>

# In silico prediction of a highly immunogenic and conserved epitope against Zika Virus

Article in *Informatics in Medicine Unlocked* · June 2021

DOI: 10.1016/j.imu.2021.100613

---

CITATIONS  
0

READS  
77

6 authors, including:



Hossain Ahmed  
University of Development Alternative

8 PUBLICATIONS 8 CITATIONS

[SEE PROFILE](#)



Md. Abu Saleh  
University of Rajshahi

38 PUBLICATIONS 200 CITATIONS

[SEE PROFILE](#)



Shafi Mahmud  
56 PUBLICATIONS 466 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



MS Thesis [View project](#)



MS thesis project of another students [View project](#)



## In silico prediction of a highly immunogenic and conserved epitope against Zika Virus



Debasish Paul<sup>a</sup>, Imdadul Haque Sharif<sup>b</sup>, Abu Sayem<sup>c</sup>, Hossain Ahmed<sup>d</sup>, Md. Abu Saleh<sup>e,\*\*</sup>, Shafi Mahmud<sup>e,\*</sup>

<sup>a</sup> Department of Pharmacy, University of Development Alternative, Dhaka, 1209, Bangladesh

<sup>b</sup> Department of Biotechnology and Genetic Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

<sup>c</sup> Faculty of Life Science, University of Development Alternative, Dhaka, 1209, Bangladesh

<sup>d</sup> Department of Biotechnology and Genetic Engineering, University of Development Alternative, Dhaka, 1209, Bangladesh

<sup>e</sup> Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, 6205, Bangladesh

### ARTICLE INFO

#### Keywords:

Zika virus  
B-Cell epitope  
Antigenicity  
Allergenicity affinity  
Vaccine  
Global solution

### ABSTRACT

Zika virus (ZIKV) is an arboviral pathogen that belongs to the Flaviviridae subgroup and is a contemporary global concern. Recent epidemic outbreaks in Brazil have indicated that ZIKV could be responsible for post-infection neurological disorders in infants, resulting in ZIKV being considered a major threat to global health. Unfortunately, no vaccine is yet available to prevent the spread of this virus. In this study, we have applied an *in silico* approach to the identification of B-cell epitopes in the ZIKV genome. By utilizing currently available genomic data and applying multiple sequence alignments and the Immune Epitope Database (IDEb) tools, a LEFYSYKKSG epitope was identified in a highly conserved peptide region of the ZIKV polyprotein. The antigenicity, allergenicity, and affinity of the B-cell epitope were evaluated, and significant B-cell affinity against ZIKV was identified. This highly conserved epitope can be used to develop a peptide-based vaccine and can also be applied toward the development of a monoclonal antibody (mAb) for therapeutic or diagnostic purposes against ZIKV.

### Introduction

Zika virus is a well-known agent of infectious disease in humans because of its current shift from an endemic disease to a pandemic disease and the potential to become a global pandemic. Its recent spread across the Americas has provoked renewed interest in this infection and its potential complications. The virus was first identified in 1947 in rhesus macaques in the Zika forest of Uganda, near the capital city of Kampala. The virus was detected in a recent outbreak in Brazil in May 2015, and to date, the virus has been identified in more than 27 countries and territories within the region [1,2]. The abundance of the mosquito vectors *Aedes aegypti* and *Aedes albopictus*, which are known to transmit Zika virus, combined with cross-continent and international travel, are thought to be responsible for the recent outbreak of this disease in the Americas [3,6].

Zika virus is an enveloped, icosahedral arbovirus (arthropod-borne virus) that belongs to the Flaviviridae family. The genus *Flavivirus* also includes other viral pathogens that can infect humans, such as dengue,

yellow fever, and West Nile viruses [7,8]. Zika virus has two distinct Asian and African lineages [4–10]. The Zika virus in Africa is thought to have been transmitted between its primary primate hosts (monkeys and apes) through mosquito vectors to humans, who serve as secondary hosts [11,12]. By contrast, outside of Africa, humans represent the primary hosts of the Zika virus, and the virus is thought to spread through several *Aedes* mosquito vectors [3]. Phylogenetically and antigenically, the Zika virus is closely related to the *Spondweni* virus [7].

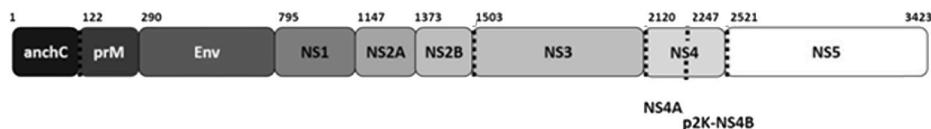
Zika virus contains a single-stranded, positive-polarity RNA genome that is approximately 11 kb in length. The open reading frame encodes a polyprotein that is proteolytically processed into three structural proteins: capsid [C], membrane [M] or pre membrane [prM], and envelope [E] proteins. The polyprotein also encodes seven non-structural proteins, which are referred to as genomic polyprotein (NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5) [7,13–15].

A recent investigation of the molecular evolution of the Zika virus indicated that the current Zika virus strains have become genetically adapted throughout their emergence in the 20th century. For example,

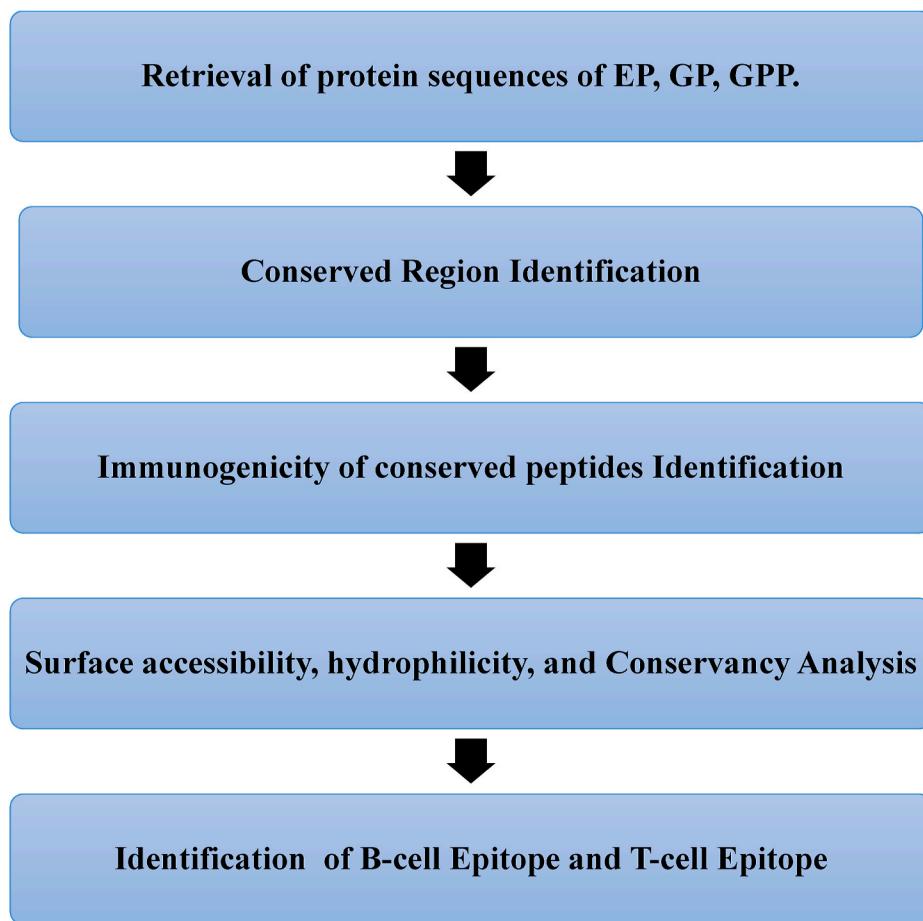
\* Corresponding author. Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, Bangladesh.

\*\* Corresponding author. Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, Bangladesh.

E-mail addresses: [saleh@ru.ac.bd](mailto:saleh@ru.ac.bd) (Md.A. Saleh), [shafimahmudfz@gmail.com](mailto:shafimahmudfz@gmail.com) (S. Mahmud).



**Fig. 1.** Topology of the Zika Virus polyprotein. Here, anchC denotes anchored capsid protein, prM denotes precursor membrane protein, Env denotes envelope glycoprotein. The non-structural proteins are numbered from NS1 to NS5. The Dotted lines denote sites of NS3 cleavage, and the numbers in upper side indicates the amino acids along the polyprotein that separates the viral proteins. Figure Adapted from Cox et al. [15].



**Fig. 2.** Schematic representation of epitope based peptide vaccine design against Zika virus.

changes in protein glycosylation patterns might have enabled them to adapt to different hosts and vector species [7,18]. Clinically, Zika virus infection resembles dengue fever and chikungunya, often causing sub-clinical (approximately 80%) or asymptomatic infections, and even in cases that result in mild illness, symptoms may take 3–12 days to appear. Early symptoms of this disease include fever, conjunctivitis, headache, myalgia, arthralgia, maculopapular rash, widespread itchy rash, retro-orbital pain, and gastrointestinal discomfort. Symptoms may last for up to 1 week [3–12], and similar to other typical viral illnesses, the disease is self-limiting. Post-infection symptoms, including neurologic manifestations and Guillain–Barre syndrome, have been identified during epidemics in Brazil and French Polynesia [2,26,27]. Zika virus can also be transmitted through blood, blood-derived products, and semen [18–21]. The transmission of the Zika virus through saliva, urine, or respiratory droplets has not been detected.

The most threatening aspect of the Zika virus is the risk of transplacental transmission and perinatal transmission during delivery, which has been demonstrated by the presence of Zika virus RNA in amniotic fluid [17] and paired blood samples collected from the infant and mother [22]. A report from the Brazilian Ministry of Health

indicated a 20-fold increase in the incidence of microcephaly cases among infants in the northeast region of the country, which led to explorations of the potential correlation between Zika infection during pregnancy and fetal abnormalities [25,26]. Although current epidemiological data has spatially and temporally associated Zika virus infection with microcephaly, robust datasets remain necessary to unequivocally support this association. However, this potential risk has revealed the Zika virus outbreak as a serious threat to human health. On 1 February 2016, the World Health Organization (WHO) declared the recent Zika virus outbreak as a “public health emergency of international concern” due to evidence of microcephaly and other neurological disorders reported in Brazil in 2015 and in French Polynesia, New Caledonia, the Cook Islands, and Easter Island in 2013 and 2014 [7].

To date, no specific antiviral therapies are available for the Zika virus, and no vaccines have been developed. Current therapies involve symptom management, although specific agents have yet to be identified for treating the symptoms of itchiness and arthralgia. Therefore, protective measures against mosquitoes are currently the most effective weapons for minimizing the spread of Zika virus infection [23,24]. As a method of controlling the mosquito vector, in addition to the use of

insecticides, the WHO is considering the release of irradiated sterile mosquitoes. However, only a vaccine can prevent the rapid spread of Zika across the world, saving millions of newborns from the risk of abnormalities. In this article, we present the possibility of an epitope-based vaccine against a highly immunogenic and conserved region of the Zika virus genome polyprotein (GPP), which was identified using *in silico* tools (see Fig. 1).

## Material and methods

The outline of this study was shown in Fig. 2.

### Retrieval of protein sequences

A total of 65 primary amino acid sequences for three different proteins encoded by various strains of Zika virus isolated from 18 different countries were retrieved from the UniProt Knowledge Base (UniProtKB database: <http://www.uniprot.org>) [29] and NCBI ([www.ncbi.nih.gov](http://www.ncbi.nih.gov)). Among the 65 obtained sequences, 40, 5, and 20 were for the envelope protein (EP), glycoprotein (GP), and GPP, respectively. The obtained sequences, which are highly reviewed and manually annotated, were stored in FASTA format for the immunoinformatics analysis.

### Identification of conserved regions

Multiple sequence alignments of the stored protein sequences were conducted using the EBI Clustal Omega program (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [30]. Multiple sequence alignment-derived amino acid sequences that covered the highest numbers of identical and similar amino acid sequences, without gaps, were chosen as preserved regions.

### Immunogenicity of conserved peptides

Identifying apparent immunogenic epitopes in an antigenic protein sequence might reduce the necessary wet lab analysis required for vaccine development. To generate an immune response, the B-cell epitope plays a vital role in interacting with the B lymphocytes [31]. The Immune Epitope Database (IEDB) tools were used to identify the essential properties necessary for B-cell antigenicity, including the Kolaskar and Tongaonkar antigenicity scale [32], Emini surface accessibility prediction [33], Karplus and Schulz flexibility prediction [34], BepiPred linear epitope prediction analysis [35], and Chou and Fasman beta-turn prediction analysis [36]. Parker hydrophilicity prediction [37] in IEDB was also performed, at the default threshold, to analyze the hydrophilicity of the predicted epitopes.

### Prediction of surface-accessible epitopes

The surface-accessible epitopes of the selected preserved peptides were predicted using the Emini surface accessibility prediction tool [31], which is among the available B-cell epitope prediction tools. The IEDB tool was used for this purpose, with a default threshold level of 1.0. A similar threshold has been used for previous vaccine design studies [38,43,44].

### Prediction of epitope hydrophilicity

The preserved epitope was analyzed to determine the hydrophilicity of the predicted epitopes using the Parker hydrophilicity prediction tool [37] in IEDB tools, with a default threshold level of 3.448.

### Prediction of epitope conservancy

An IEDB analysis tool used was used to perform epitope conservancy prediction [39] for all predicted epitopes. Each epitope's conservancy

Sequence	Protein Name	Position
IAPAYSRIGVSNRDEVFGMSGGTWVVLHGGCVTVMAQDKPTIDELYTTVSNMAEVRSYCYEASISDMASDSCRPTQGEAYLDKCSDTIQVC	Glycoprotein (GP)	1–98
DAHARRQTVVVLGSQEGAVHTALAGALEAMDGAKGRIFSGHLKCRLKMDKIRLKGYSYSLCTAAFTIKPAETLHGTVTVEVQYAGTDGCK	Envelope Protein (EV)	116–210
GWGEAGALITAATSTLVEGSPNPKWNNSTATSLCNIFRGSYLAGSLYTVTRNAGLVKRREGGTGETLGKWRKARLNQMSALEFSYKSIGTEVCREEARRAILKDGVATGGHAVSRGSAKLRLWL	Genome Polyprotein (GPP)	2459–2585

levels were calculated based on protein sequence identities.

#### T-cell epitope anticipation and MHC-I restriction analysis

NetCTL server was used to identify the potential T-cell epitopes of the selected protein groups. First, potential T-cell epitopes were recognized by the NetCTL server [47]. The default approach was utilized to predict epitope characteristics, such as primary histocompatibility complex class I (MHC-I)-binding, proteasomal C-terminal cleavage, and transporter of antigenic peptide (TAP) transport effectiveness. A total of 12 MHC-I supertypes were used to identify epitopes. The sensitivity and specificity of epitope identification were 0.89 and 0.94, respectively, at a threshold level of 0.5. The ability of the MHC-I [48] allele to bind with the proposed epitopes was determined using the T-Cell Epitope Expectation Instruments from the Immune Epitope Database and Analysis Resource (IEDB-AR). The half-maximal inhibitory concentration ( $IC_{50}$ ) for peptide binding with MHC-I was assessed using the Stabilized Matrix Method, utilizing a preselected 9.0-mer epitope.

#### Allergenicity and antigenicity assessment

The allergenicity of the conserved sequences identified using the EBI Clustal Omega program was predicted with AlgPred [49], an online server for predicting allergenic proteins and mapping IgE epitopes. The antigenicity of the proposed epitopes was verified using VexiJen server v2<sup>50</sup>.

#### Assessment of the epitope affinity toward B-cells

The affinity of the epitope toward B-cells was evaluated using the online docking server CABSDock (<http://biocomp.chem.uw.edu.pl/CABSDock/>) [51]. The CABSDock server provides an interface for docking and presenting protein-peptide interactions using an exceptionally resourceful protocol for modeling the flexible docking of proteins with peptides [51,52].

## Results

#### Multiple sequence alignment

The primary amino acid sequences for 40 EP, 5 GP, and 20 GPP proteins were retrieved and analyzed to identify conserved regions. These protein sequences were obtained from various strains of Zika virus obtained in 18 countries, including Brazil, Guatemala, Senegal, Côte d'Ivoire, Burkina Faso, Central African Republic, Gabon, Chile, Philippines, Thailand, Colombia, China, Indonesia, Haiti, Uganda, Norway, Russia, and Suriname (Supp File 1–3). The identified conserved regions, including the primary amino acid sequences, are summarized in Table 1.

#### The identification of B-cell epitopes

A number of investigation tools based on protein sequences were utilized to identify potential B-cell epitopes. The conserved region's physicochemical properties, which are determined by the amino acid sequence, were evaluated by the Kolaskar and Tongaonkar antigenicity tool. The average antigenic propensity value was determined to be 1.006, with a maximum value of 1.110 and a minimum value of 0.879. The threshold for antigenic determination of conserved regions was 1.00, with all values greater than 1.00 indicative of potential antigenic determinants. The presence of beta-turns in a protein can play a significant role in initiating antigenicity because beta-turns can introduce hydrophilic properties and affect surface accessibility [40]. In addition, peptide flexibility is thought to have a strong correlation with antigenicity [41]. The BepiPred linear epitope prediction tool was used to discern the presence of linear B-cell epitopes in the conserved regions.

**Table 2**

Linear B-cell epitope, antigenic regions, and surface accessible epitopes predicted from the conserved region of glycoprotein (GP).

Peptide	B-cell epitope analysis	Length
DFVEGMSGGT	BepiPred analysis (linear B-cell epitope)	10
AQDKPTV		7
SDMASDSRCPTQGEAYLDKQSDTQY		25
YSIRCIGVSNRD	IEDB analysis (antigenic sites)	12
TWVDVVLEHGGCVTVMA		17
PTVDIELVTTTV		12
VRSYCYEASI		10
AQDKPTV	Surface accessible epitope	7
GEAYLDKQSDT		11

**Table 3**

Linear B-cell epitope, antigenic regions, and surface accessible epitopes predicted from the conserved region of Envelope Protein (EP).

Peptide	B-cell epitope analysis	Length
DAHAKR	BepiPred analysis (linear B-cell epitope)	6
SQEAVH		7
EAEMDGAKGR		10
KIPAETLHG		9
KRQTVVVLGSQEGAVHTALAGAL	IEDB analysis (antigenic sites)	23
LFSGHLKCRKLMDK		14
RLKGVSYSLCTAAFTFTK		18
AEMDGAK	Surface accessible epitope	7
LKMDKRL		8

**Table 4**

Linear B-cell epitope, antigenic regions, and surface accessible epitopes predicted from the conserved region of Genome Polyprotein (GPP).

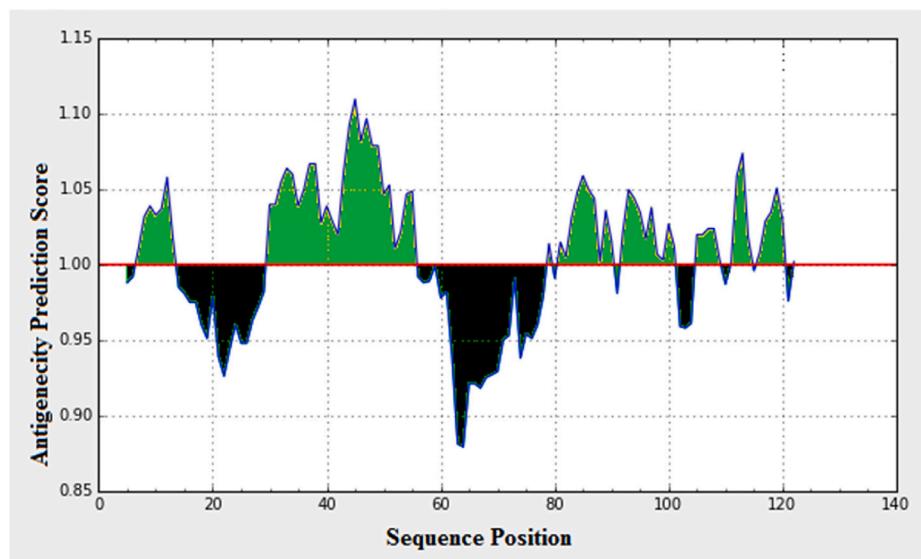
Peptide	B-cell epitope analysis	Length
GWGEA	BepiPred analysis (linear B-cell epitope)	5
STLWEGSPNKYWNSSTAT		18
VKRGGGGTGETLGEK		15
KSGIT		5
RRALKDGVATGGHAVSRGS		19
ALITAAT	IEDB analysis (antigenic sites)	7
ATSLCNIFRGSYLAGASLIYTVRTRNA		26
SALEFVSYKK		10
GITEVCREEA		10
EGSPNPKYWNS	Surface accessible epitopes	11
TLGEKWKRARN		11
FYSYKKG		8
REEARRA		7
SPNKYWNSS	Chou and Fasman Beta turn	9
LEFYSYKKGITE		13
YSYKKSG	Karplus and Schulz Flexibility	7
GGGTGETLGEKWKA		14
WEGSPNKY	Parker Hydrophilicity	8
WNSSTATSLCN		11
EFYSYKKGITEVCREEARRA		21
VSRGSAK		7

#### Glycoprotein (GP)

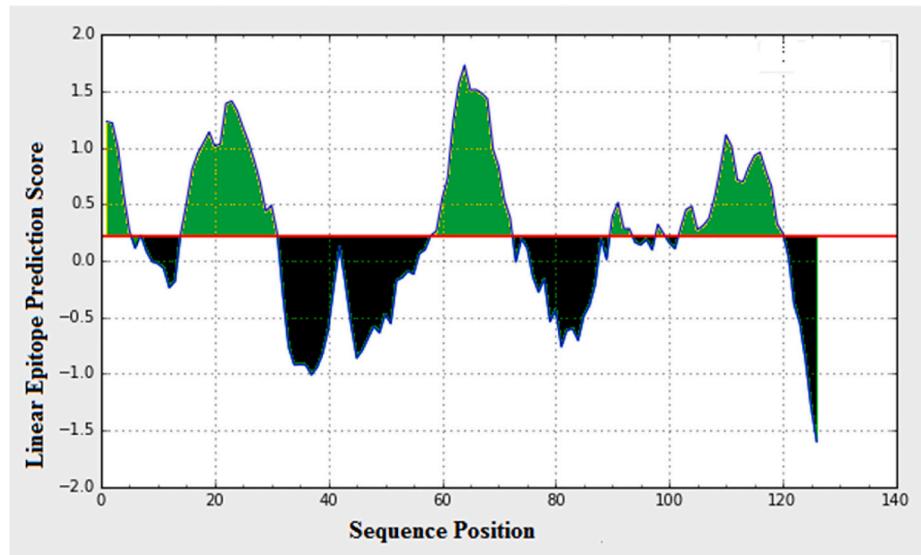
The linear B-cell epitopes, antigenic sites predicted using IEDB tools, and surface-accessible epitopes from the predefined preserved GP sequence are summarized in Table 2. We identified three potential linear B-cell epitopes, but only one epitope (AQDKPTV) met the criteria of being a surface-accessible epitope. Therefore, none of the identified GP epitopes satisfied the criteria for potential B-cell epitopes.

#### Envelope protein (EP)

The linear B-cell epitopes, antigenic sites predicted using IEDB tools, and surface-available epitopes from the recognized preserved EP sequence are summarized in Table 3. Four potential linear B-cell



**Fig. 3.** Antigenicity of the conserved peptide of Genome Polyprotein. The conserved peptide was found to be highly antigenic in the IEDB analysis. Most of the residues were found above the threshold 1.00. Residues in the Green region are antigenic and in the green region are below the threshold (red line).



**Fig. 4.** BepiPred linear epitope prediction of Genome Polyprotein. The linear epitope residues of the conserved peptide which are above the cut off are located in the Green region. The red horizontal line indicates linear cutoff (1.000).

epitopes were identified. One of the predicted linear B-cell epitopes from this region, SQEGAVH, overlapped with another antigenic site identified by the Kolaskar and Tongaonkar prediction; however, no similarities were found with the surface-accessible epitopes. Thus, none of the identified EP epitopes satisfied the criteria for potential B-cell epitopes.

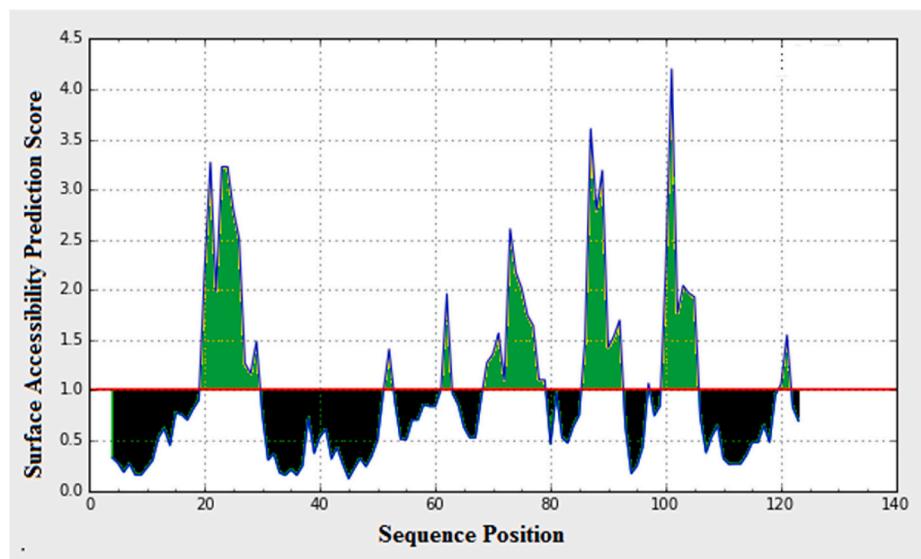
#### Genome polyprotein (GPP)

The linear B-cell epitopes, antigenic sites predicted using IEDB tools, surface-available epitopes, Chou and Fasman beta-turn epitopes, Karplus and Schulz flexibility-characterized epitopes, and Parker hydrophilicity-characterized epitopes from the predefined preserved GPP sequence are summarized in Table 4. For GPP, we identified 4 antigenic sites, 4 surface-accessible epitopes, 6 potential linear B-cell epitopes, 2 beta-turn epitopes, 2 flexible epitopes, and 3 hydrophilic epitopes. One of the predicted surface-accessible epitopes FYSYKKSG, partially overlapped with several predicted epitopes, including an

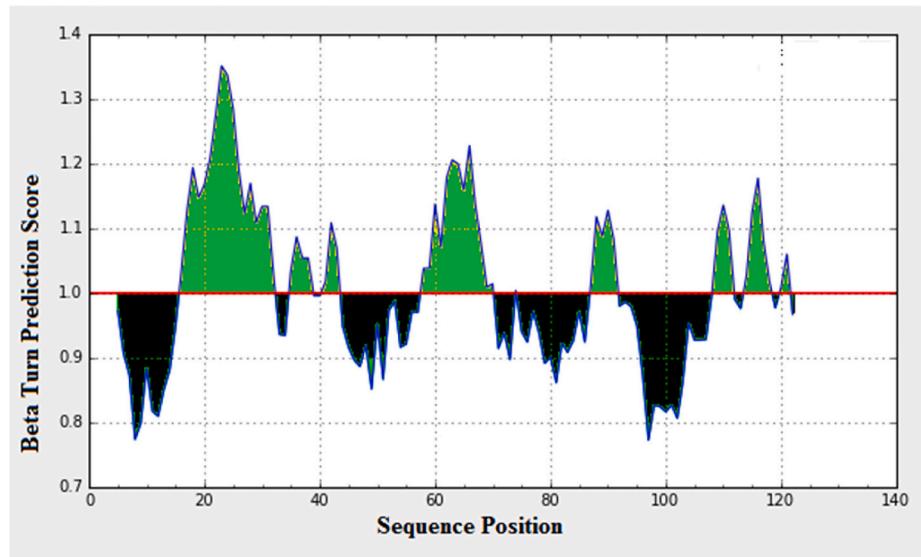
antigenic site predicted by the Kolaskar and Tongaonkar analysis (SALEFSYKK), a Chou and Fasman beta-turn epitope (LEFY-SYKKSGITE), a Karplus and Schulz flexibility-characterized epitope (YSYKKSG), and a Parker hydrophilicity-characterized epitope (EFY-SYKKSGITEVCREEARRA). Graphical representations of these B-cell epitope characteristics, which were predicted using IEDB analysis tools, are shown in Figs. 3–8. Based on the analysis of the obtained results, the LEFYSYKKSG epitope of GPP appeared to completely satisfy the necessary criteria to represent a potential B-cell epitope for the Zika virus.

#### Analysis of epitope conservancy

The conservancies of predicted epitopes, AQDKPTV in GP, SQEGAVH in EP, and LEFYSYKKSG in GPP, were determined by IEDB conservancy analysis resources. All of the analyzed epitopes in each protein were found to be 100% conserved (Fig. 9).



**Fig. 5.** Conserved peptide's surface accessibility of Genome Polyprotein. The surface accessible residues of the conserved peptide which are above the cut off are located in the Green region. The red horizontal line indicates surface accessibility cutoff (1.000).



**Fig. 6.** Chou and Fasman Beta turn epitope prediction of Genome Polyprotein. The beta turn epitope residues of the conserved peptide which are above the cut off are located in the Green region. The red horizontal line indicates beta turn cutoff (1.000).

#### Determination of T-cell epitopes and MHC limitations

A collaborative approach to the analysis of the 12 MHC-I supertypes that recognize T-cell epitopes was applied using the NetCTL server [33]. The five epitopes from each protein with the highest scores were selected from the combined results of the NetCTL server. These selected epitopes were then analyzed for MHC binding. MHC alleles demonstrated high affinities ( $IC_{50} < 200$  nM) for the proposed T-cell epitopes, as assessed by the Stabilized Matrix Method, as shown in Table 5.

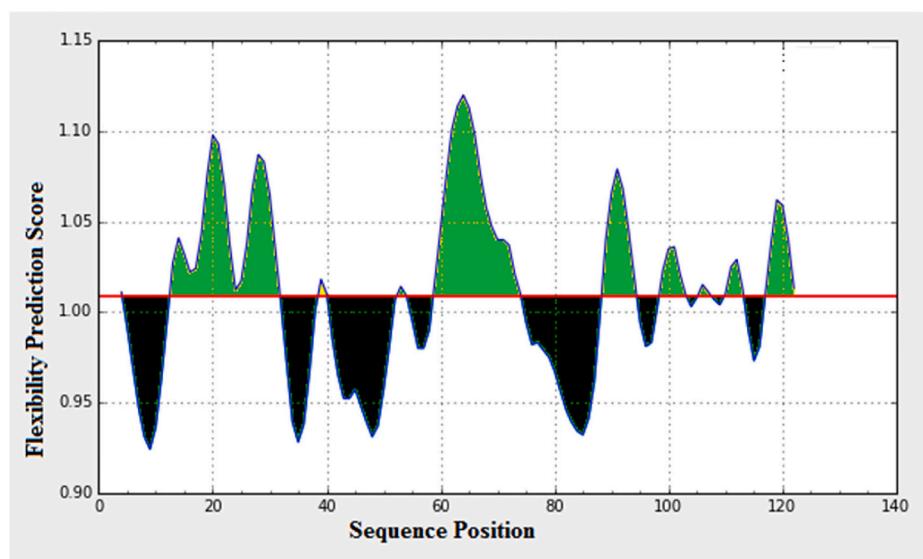
#### Allergenicity and antigenicity assessment

The allergenicity of the conserved sequence identified in GPP was evaluated using AlgPred [49] and the support vector machine (SVM) prediction method, which indicated that this sequence was non-allergenic based on the amino acid composition. The predicted AlgPred score was 1.51, with a default threshold of 0.4. All of the

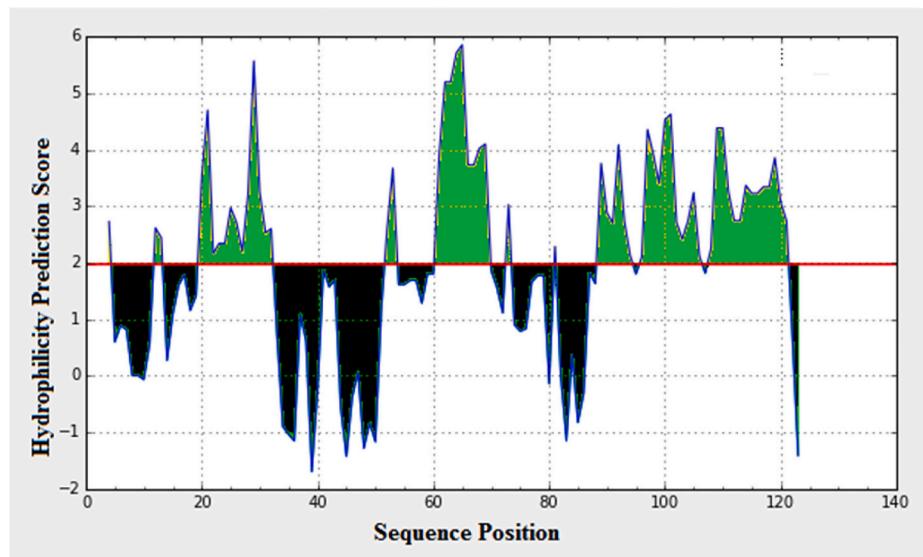
proposed epitopes from all three proteins were evaluated by VexiJen server [50], which revealed the highest antigenicity for the LEFY-SYKKSG epitope in GPP. The overall score for this epitope was 0.7522, indicating that it could serve as a probable antigen.

#### Assessment of the epitope affinity toward B-cells

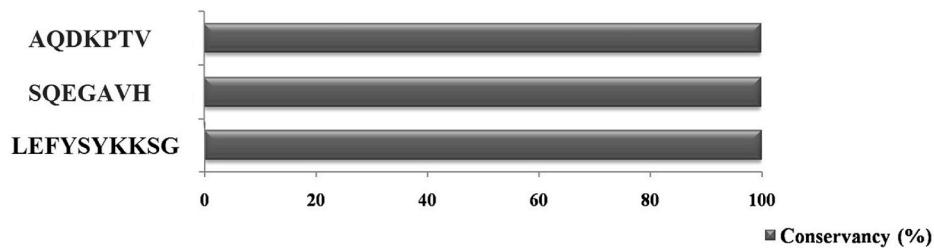
Before conducting the molecular docking analysis using the CABSdock server [51], we first obtained the crystal structure of HLA\*3501 [31] in combination with the influenza NP418 epitope (PDB ID: 3LKO) from the RCSB Protein Data Bank (PDB) database and removed the attached ligand. We submitted the prepared HLA structure and the epitope sequence (LEFY-SYKKSG) to the CABSdock server for molecular docking analysis. The results received from the server are presented in Table S1, Fig. 10, and Fig. S4.



**Fig. 7.** Karplus and Schulz Flexibility prediction of Genome Polyprotein. The flexible epitope residues of the conserved peptide which are above the cut off are located in the Green region. The red horizontal line indicates flexible residues cutoff (1.000).



**Fig. 8.** Parker Hydrophilicity prediction of Genome Polyprotein. The hydrophilic epitope residues of the conserved peptide which are above the cut off are located in the Green region. The red horizontal line indicates hydrophilic residues cutoff (1.000).



**Fig. 9.** Conservancy of the predicted consensus epitopes. All of the predicted epitope from each protein category were found to be 100% conserved. In this case, Y axis indicates the epitopes and X axis indicates the conservancy percentage.

## Discussion

The expeditious spread of the Zika virus in Brazil could potentially

pose a serious threat to public health preparedness both across the continent and around the world. A vaccine represents an ideal solution to addressing this challenge and reducing the risks associated with Zika

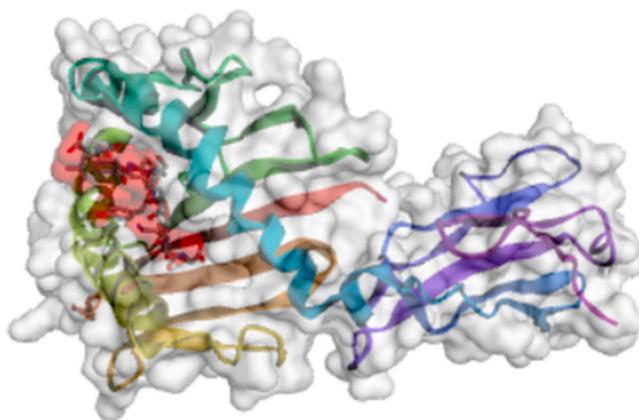
**Table 5**

Most potential T-cell epitopes of the selected protein with interacting mhc-1 alleles.

Epitope	Interacting mhc-1 ALLELE	Protein Region
MSALEFYSY	HLA-B*15:17; HLA-B*35:01; HLA-B*27:20; HLA-A*29:02; HLA-C*12:03; HLA-A*68:23; HLA-C*03:03; HLA-A*30:02; HLA-B*15:01; HLA-A*32:15; HLA-A*32:07; HLA-A*26:02; HLA-B*58:01; HLA-A*32:01;	GPP
YLAGASLIY	HLA-A*80:01; HLA-A*29:02; HLA-B*15:02; HLA-C*03:03; HLA-B*15:01; HLA-B*27:20; HLA-A*32:15; HLA-B*35:01; HLA-A*32:07; HLA-A*68:23; HLA-C*12:03; HLA-C*14:02; HLA-B*15:17;	GPP
ITAATSTLW	HLA-B*58:01; HLA-B*15:17; HLA-A*68:23; HLA-B*57:01; HLA-C*03:03; HLA-A*32:07; HLA-C*12:03; HLA-B*27:20; HLA-A*32:01; HLA-A*02:50; HLA-B*40:13; HLA-A*32:15; HLA-B*53:01	GPP
CREEARRAL	HLA-B*27:20; HLA-C*06:02; HLA-A*68:23; HLA-B*15:02; HLA-A*02:17; HLA-A*02:50; HLA-C*07:01; HLA-C*14:02; HLA-A*32:07; HLA-C*12:03; HLA-C*03:03	GPP
YTVTTRNAGL	HLA-A*68:23; HLA-C*03:03; HLA-B*15:17; HLA-B*15:02; HLA-A*32:07; HLA-B*27:20; HLA-A*02:50; HLA-C*15:02; HLA-C*12:03; HLA-A*26:02; HLA-C*14:02; HLA-A*32:15; HLA-A*68:02; HLA-A*26:03; HLA-A*02:06	GPP
MAEVRSYCY	HLA-C*12:03; HLA-A*68:23; HLA-B*35:01; HLA-C*03:03; HLA-A*32:15; HLA-C*05:01; HLA-A*32:07; HLA-B*15:17; HLA-B*27:20; HLA-C*07:01; HLA-B*15:02; HLA-A*30:02	GP
TTVSNMAEV	HLA-A*68:23; HLA-A*68:02; HLA-C*12:03; HLA-A*02:50; HLA-A*32:07; HLA-A*26:02; HLA-C*15:02; HLA-C*03:03; HLA-A*69:01; HLA-A*02:06; HLA-B*27:20; HLA-C*14:02; HLA-A*02:03	GP
APAYSIRCI	HLA-C*03:03; HLA-B*27:20; HLA-A*02:50; HLA-A*68:23; HLA-A*32:15; HLA-C*12:03; HLA-A*32:07; HLA-B*07:02	GP
LEHGGCVTV	HLA-A*02:50; HLA-A*32:07; HLA-C*12:03; HLA-A*68:23; HLA-B*40:01; HLA-C*03:03; HLA-A*32:15; HLA-B*40:02	GP
CPTQGEAYL	HLA-C*03:03; HLA-A*68:23; HLA-A*32:07; HLA-A*02:50; HLA-B*15:02; HLA-C*12:03; HLA-B*27:20;	GP
SYSLCTAAF	HLA-A*24:03; HLA-A*32:07; HLA-C*14:02; HLA-A*68:23; HLA-B*27:20; HLA-A*32:15; HLA-C*03:03; HLA-A*24:02; HLA-C*12:03; HLA-A*23:01	EP
HAKRQTVVV	HLA-B*27:20; HLA-C*12:03; HLA-A*02:50; HLA-A*68:23; HLA-A*32:07; HLA-A*30:01; HLA-C*15:02	EP
QEAGAVHTAL	HLA-B*27:20; HLA-C*03:03; HLA-B*40:01; HLA-A*68:23; HLA-A*02:50; HLA-C*12:03; HLA-B*15:02; HLA-A*32:07; HLA-A*32:15;	EP
KLRLKGVSY	HLA-B*27:20; HLA-A*32:07; HLA-C*12:03; HLA-B*15:01; HLA-A*32:15; HLA-A*68:23; HLA-A*30:01; HLA-C*14:02	EP
GRLFSGHILK	HLA-B*27:20; HLA-A*32:07; HLA-A*68:23; HLA-B*27:05; HLA-C*12:03; HLA-C*07:01; HLA-C*03:03; HLA-C*14:02;	EP

virus transmission. The potential for the global spread of Zika was speculated due to the mass gathering of global citizens associated with the Olympic games hosted by Brazil [28], although, fortunately, the feared global spread has not been reported following the return of athletes and spectators to their home countries.

Compared with other dangerous mosquito-borne *Flaviviruses*, such as dengue, research on the Zika virus was not prioritized until the recent outbreak in Brazil due to the mild nature of this illness. Existing literature regarding the Zika virus is scarce; therefore, viral pathogenesis is not yet understood in detail. Intriguingly, the Zika virus is capable of entering cells through a variety of receptors and can replicate inside the



**Fig. 10.** Molecular docking study analyzed by CABS-dock server between the HLA\*3501 and the epitope sequence (LEFYSYKKSG).

immune cells of the skin [42]. This finding indicates that the virus might also be capable of replicating inside neuronal cells, indicating the potential viral involvement in neurological disorders. Zika virus has been suggested as a potential causative agent for post-infection microcephaly and other neurological disorders, indicating the potential of Zika virus to become a global burden, resulting in increased infant neurological morbidity. The importance of effective vaccine development has been emphasized by the WHO and other health authorities, which will help stem the spread of infection. The identification of effective vaccine development approaches remains necessary for the Zika virus, and vaccine development becomes increasingly important when a virus successfully spreads across continents. A global vaccine represents an ideal solution to reducing the risks associated with Zika virus spread. A vaccine that elicits immunity in humans against all circulating Zika virus strains can be developed by identifying a well-conserved epitope for vaccine targeting purposes. This study used various bioinformatics tools to predict a 100% conserved and highly immunogenic B-cell epitope for the development of a vaccine to confer immunity against the Zika virus.

Epitopes capable of eliciting humoral immunity are traditionally predicted through experimental approaches that involve tedious, time-consuming, and expensive screening procedures. The current study is an example of computer-aided epitope selection that can be validated by later *in vitro* and *in vivo* experiments performed in a targeted manner. Predicting a conserved epitope in a peptide can effectively reduce the efforts required to achieve the ultimate goal of vaccine development. Similar approaches have been used to identify potential epitopes for many other viruses. For instance, Medina et al. identified conserved epitopes across virulent strains of influenza virus [45], Sakib et al. found epitopes in Nipah virus [46], Ali et al. described conserved epitopes in Ebola virus [43], and Mahmud et al. presented conserved epitopes in Marburg virus [53]. The underlying rationale for epitope-based vaccines suggests that the use of a conserved epitope should provide broad-spectrum immunity against both currently prevalent strains and newly evolved species, unlike vaccines developed against variable peptide regions, which become ineffective when viral mutations change the peptide sequence. The identification of specific and conserved epitopes can also be applied to diagnostic and disease monitoring purposes.

To develop an epitope-based vaccine against the global distribution of the Zika virus, we analyzed the genome sequences obtained from 18 unique countries. Three conserved regions were identified by multiple sequence alignment (Table 1, Supp Figs. S1–S3). These regions were analyzed through multiple analyses to identify potential B-cell epitopes. We applied all of the available IEDB B-cell analysis tools to each conserved region; however, none of the epitopes identified in GP or EP satisfied all of the necessary criteria to represent a suitable vaccine epitope. By contrast, a single epitope in GPP was identified as a potential B-cell epitope. The conservancies of the predicted epitopes were

determined, revealing that all of the predicted epitopes were 100% preserved across all of the sequences examined from 18 different countries.

In this study, we identified a conserved peptide region within the Zika virus GPP protein. Based on sequence conservancy, antigenicity, hydrophilicity, and surface accessibility, which we analyzed using various computational tools, our approach to vaccine design predicts that the identified LEFYSYKKSG epitope will be able to induce B-cell-mediated immune responses. However, because this study was performed completely as an *in silico* study, these predictions must be confirmed in a suitable disease model.

The use of a suitable experimental animal model remains a prerequisite for validating the candidacy of the identified epitope for peptide vaccine development. If confirmed, the identified highly immunogenic epitope can be used to develop chimeric or recombinant vaccines in the near future. This conserved and immunogenic epitope can also be tested for the development of a monoclonal antibody (mAb), for use as a therapeutic antibody and for disease diagnostic approaches, which can be applied to all global strains of Zika virus.

Upon the identification of B-cell epitopes, the identification of T-cell epitopes was also expected. The epitopes YTVTRNAGL in GPP, TTVSNMAEV in GP, and SYSLCTAAF in EP showed the best T-cell immunity activation outcomes, with potential antigenicity based on the MHC-I binding analysis. The best T-cell epitopes were selected based on the interaction with the greatest number of HLA loci, indicating that MHC molecules could easily display these epitopes on the T-cell surface.

The allergenicity of the conserved GPP sequence was evaluated using AlgPred, which predicted the identified sequence to be non-allergenic. Antigenicity was also evaluated using the VexiJen server, which found that the sequences showed antigenic properties. Finally, the affinity of the B-cell epitope was verified by molecular docking between the MHC-I allele (HLA\*3501) and the selected epitope, which revealed significant interactions among receptor amino acids and the identified epitope (Fig. 10 and Supplementary Fig. S4, and Table S1).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The Author's reports that no acknowledgments for this work. Any organization did not fund the work. The Author's confirmed that there is no conflict of interest regarding this paper's work and publication.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j imu.2021.100613>.

## References

- [1] European Centre for Disease Prevention and Control (ECDC). Countries and territories with local Zika transmission. April 2016. [http://ecdc.europa.eu/en/healthtopics/zika\\_virus\\_infection/zika-outbreak/Pages/Zika-countries-with-transmission.aspx](http://ecdc.europa.eu/en/healthtopics/zika_virus_infection/zika-outbreak/Pages/Zika-countries-with-transmission.aspx).
- [2] World Health Organization. Zika virus microcephaly and guillain-barré syndrome situation report 19 february 2016. [http://apps.who.int/iris/bitstream/10665/204454/1/zikasitrep\\_19Feb2016\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/204454/1/zikasitrep_19Feb2016_eng.pdf); February 2016.
- [3] Duffy MR, Chen TH, Hancock WT, et al. Zika virus outbreak on Yap Island, Federated States Of Micronesia. *N Engl J Med* 2009;360(24):2536–43.
- [4] Haddow AD, Schuh AJ, Yasuda CY, et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Neglected Trop Dis* 2012;6:2.
- [5] Musso D, Cao-Lormeau VM, Gubler DJ. Zika virus: following the path of dengue and Chikungunya. *Lancet* 2015;386:243–4.
- [6] Rodriguez-Morales AJ. Zika: the new arbovirus threat for Latin America. *J Infect Dev. Ctries.* 2015;9(6):684–5.
- [7] Basarab M, Bowman C, Aarons EJ, Cropley I. Zika virus: clinical review. *BMJ* 2016;352:i1049.
- [8] Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* 1952;46(5):509–20.
- [9] Kuno G, Chang GJ. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch Virol* 2007;152(4):687–96.
- [10] Faye O, Freire CC, Iamarino A, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Neglected Trop Dis* 2014;8(1):e2636.
- [11] Haddow AJ, Williams MC, Woodall JP, et al. Twelve isolations of Zika virus from *Aedes (Stegomyia) Africanus* (Theobald) taken in and above a Uganda forest. *Bull World Health Organ* 1964;31(1):57–69.
- [12] McCrae AW, Kirby BG. Yellow fever and Zika virus epizootics and enzootics in Uganda. *Trans R Soc Trop Med Hyg* 1982;76(4):552–62.
- [13] Tamura K, Stecher G, Peterson D, et al. Molecular evolutionary Genetics analysis version 6.0. *Mol Biol Evol* 2013;30(12):2725–9.
- [14] Sirohi D, Kuhn RJ. Zika virus structure, maturation, and receptors. *J Infect Dis* 2017;216(S10):S935–44.
- [15] Cox BD, Stanton RA, Schinazi RF. Predicting Zika virus structural biology: challenges and opportunities for intervention. *Antiviral Chem Chemother* 2015;24(3–4):118–26.
- [16] Calvet G, Aguiar RS, Melo AS, Sampaio SA, De Filippis I, Fabri A, de Filippis AM. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis* 2016;16(6):653–60.
- [17] Foy BD, Kobylinski KC, Chilson FJL, et al. Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg. Inf Disp* 2011;17:880–2.
- [18] Musso D, Roche C, Robin E, et al. Potential sexual transmission of Zika virus. *Emerg Infect Dis* 2015;21:359–61.
- [19] Atkinson B, Hearn P, Afrouch B, et al. Detection of Zika virus in semen. *Emerg Infect Dis* 2016;22:5.
- [20] Musso D, Nhan T, Robin E, et al. Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia November 2013 - February 2014. *Euro Surveill* 2014;19(14):2.
- [21] Besnard M, Lastère S, Teissier A, et al. Evidence of perinatal transmission of Zika virus. *French Polynesia December 2013 - February 2014. Euro. Surveill.* 2014;19(13):4.
- [22] Stanczyk NM, Behrens RH, Chen-Hussey V, et al. Mosquito repellents for travelers. *BMJ* 2015;350:h99.
- [23] Ahmad SSY, Amin TN, Ustianowski A. Zika virus: management of infection and risk. *BMJ* 2016;352:i1062.
- [24] Mlakar J, Korva M, Tul N, et al. Zika virus associated with microcephaly. *N Engl J Med* 2016;374:951–8.
- [25] European Centre for Disease Prevention and Control. Rapid risk assessment: Zika virus epidemic in the Americas: potential association with microcephaly and Guillain-Barré syndrome. December 2015;10. <http://ecdc.europa.eu/en/publications/Publications/zika-virus-americas-association-with-microcephaly-rapid-risk-assessment.pdf>.
- [26] Loos S, Mallet HP, Leparc GI, et al. Current Zika virus epidemiology and recent epidemics. *Med Maladies Infect* 2014;44(7):302–7.
- [27] Petersen E, Wilson ME, Touch S, et al. Brazil gatherings rapid spread of Zika virus in the Americas - implications for public health preparedness for mass gatherings at the 2016 Brazil olympic. *Int J Infect Dis* 2016;44:11–5.
- [28] The UniProt Consortium. Activities at the universal protein Resource (UniProt). *Nucleic Acids Res* 2014;42(D1):D191–8.
- [29] Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
- [30] Khan MA, Hossain MU, Zaman SMRU, Morshed MN. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *Scand J Immunol* 2015;82(1):25–34.
- [31] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990;276(1):172–4. 2.
- [32] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1995;55(3):836–9.
- [33] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften* 1985;72(4):212–3.
- [34] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;2:2.
- [35] Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem* 1978;47:251–76.
- [36] Parker JM, Guo D, Hodges RS. 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.
- [37] Hasan MA, Khan MA, Datta A, et al. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. *Mol Immunol* 2015;65(1):189–204.
- [38] Bui HH, Sidney J, Li W, et al. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinf* 2007;8:361.
- [39] Sali A, Potterton L, Yuan F, et al. Evaluation of comparative protein modelling by MODELLER. *Proteins Struct. Funct. Bioinf.* 1995;23(3):318–26.
- [40] Söding J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* 2005;21(7):951–60.

- [42] Hamel R, Dejarnac O, Wichit S, et al. Biology of Zika virus infection in human skin cells. *J Virol* 2015;89(17):8880–96.
- [43] Sharmin R, Islam ABMMK. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. *BMC Bioinf* 2014;15:61.
- [44] Ali MT, Islam MO. A highly conserved GEQYQQQLR epitope has been identified in the nucleoprotein of Ebola virus by using an in silico approach. *Advances in Bioinformatics* 2015;278197.
- [45] Medina JEM, Vallejo CJS, Tenorio AM, et al. In silico identification of highly conserved epitopes of influenza A H1N1, H2N2, H3N2, and H5N1 with diagnostic and vaccination potential. *BioMed Res Int* 2015;813047.
- [46] Sakib MS, Islam MR, Hasan AKMM, Nabi AHMN. Prediction of epitope-based peptides for the utility of vaccine development from fusion and glycoprotein of Nipah virus using in silico approach. *Advances in Bioinformatics* 2014;402492.
- [47] Larsen MV, Lundsgaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf* 2007;8:424.
- [48] Larsen MV, Lundsgaard C, Lamberth Kasper, Buus S, Brunak S, Lund O, Nielsen M. An integrative approach to CTL epitope prediction. A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 2005;35(8):2295–303.
- [49] Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006;34(Web Server). <https://doi.org/10.1093/nar/gkl343>. W202–W209.
- [50] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8:4.
- [51] 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 Res* 2015;43(W1):W419–24. <https://doi.org/10.1093/nar/gkv456>.
- [52] 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. <https://doi.org/10.1016/j.ymeth.2015.07.004>.
- [53] Mahmud SMN, Rahman M, Kar A, Jahan N, Khan A. Designing of an epitope-based universal peptide vaccine against highly conserved regions in RNA dependent RNA polymerase protein of human Marburg virus: a computational Assay. *Anti-Infective Agents* 2020;18:294–305.