ISSN: 2319-6505

Available Online at http://journalijcar.org

*International Journal of Current Advanced Research*
*Vol 5, Issue 9, pp 1229-1235, September 2016*

RESEARCH ARTICLE

*International Journal of Current Advanced Research*

ISSN: 2319 - 6475

# IMMUNOINFORMATICS SCREENING OF PROSPECTIVE MHC CLASS I RESTRICTED CYTOTOXIC T-CELL BASED EPITOPES IN ZIKA VIRUS

**Kishore Sarma., Sudipta Sankar Bora., Aniruddha Jakharia., Pradyumna K. Mohapatra and Biswajyoti Borkakoty***

Regional Medical Research Centre NE, Indian Council of Medical Research,
Post Box. 105, Dibrugarh, 786 001, Assam, India

**A R T I C L E I N F O**

***Article History:***

Received 29th June, 2016
Received in revised form 4th July, 2016 Accepted 18th August, 2016 Published online 23rd September, 2016

***Key words:***

Zika virus, Epitope, MHC class I, Docking & Population coverage

**A B S T R A C T**

In a recent development, Zika virus (ZIKV) a zoonotic arbovirus responsible for nonspecific clinical presentation has cautiously been linked to the latest cases of neonatal microcephaly. Several reports of ZIKV epidemic outbreaks certainly necessitate new regimen of preventive measures. Therefore the present study aims to screen novel promiscuous epitopes which can efficiently evoke CTL response against ZIKV infections. Using immunoinformatics approach, ZIKV polyprotein was screened to identify prospective epitopes preferentially binding to MHC class I molecules. Out of 1,798 hits, five candidate epitopes from each protein were selected based on IEDB and SYFPEITHI score. These candidate epitopes were further tested for their population coverage, antigenicityd conservancy within the human proteome and existing ZIKV lineages. Five most promiscuous epitopes (covering more than 60% worldwide population) from various structural and nonstructural proteins (E1: MMLELDPPF from envelope, E2: FAAGAWYVY and E3: MTICGMNPI from NS2B, E4: YAWDFGVPL and E5: MAMATQAGV from NS4B) were selected. E1, E2 and E5 exhibited 100% conservancy among the representative strains. Molecular docking study revealed their efficient binding affinity with representative HLA allele A*0201. This study proposes the possible usage of these epitopes towards candidate vaccine development against ZIKV.



## INTRODUCTION

Zika virus (ZIKV) belonging to the genus *Flavivirus* (family flaviviridae) is a zoonotic arbovirus (Faye *et al*., 2014), responsible for nonspecific clinical presentation like rash, prolong arthralgia, headaches and mild fever (dengue- or chikungunya-like syndromes) in human (Zanluca *et al*., 2015; Campos *et al*., 2015). A close relative of Spondweni virus, ZIKV was originally transmitted in Africa (and latter in Southeast Asia) through a sylvatic cycle involving mainly nonhuman primates and several *Aedes* mosquitoes (Faye *et al*., 2014). In recent years, its Asian lineage is reported to be spreading epidemically in a dengue-like *Aedes aegypti*-human-*Aedes aegypti* cycle in the Pacific Islands and in South America (Musso *et al*., 2014). Additionally, sexual and perinatal infection route may also aid in viral transmission (Besnard *et al*., 2014).

Till date, a few epidemic outbreaks of this lineage have been reported; e.g. outbreaks in Yap Island, Federated States of Micronesia in 2007 (Duffy *et al*., 2009) and, in French Polynesia (Cao-Lormeau *et al*., 2014) in 2013 and 2014. The virus was confirmed to reach America (at Easter Island) via Oceania in 2014 (Musso *et al*., 2014). More recently, alarming viral outbreaks have been reported in the Northeastern region of Brazil (Zanluca *et al*., 2015; Campos *et al*., 2015).

Outbreaks of ZIKV infections positively correlate to incidences of Guillain-Barre syndrome, a rapidly progressive motor disorder associated with absent reflexes and a raised CSF protein that could lead to ascending paralysis, respiratory failure and death (Haymaker, 1949). A 20-fold increase in such cases was observed in French Polynesia (Musso *et al*., 2015). Disturbingly, latest outbreak of ZIKV in Brazil has recently been associated with sharp increases in a birth defect known as congenital microcephaly (Ministério da Saúde, 2015). According to WHO, microcephaly is defined as a head circumference equal to or lower than two standard deviations below the mean (≤ -2SD) for age and sex or about less than the second percentile (PAHO/WHO 2015). There is currently only ecological evidence of an association between the two events. A possible materno-foetal transmission (intrapartum/ transplacental/vertical transmission) cannot be ruled out with the evidence available (Calvet *et al*., 2016). This kind of transmission has already been demonstrated for other flaviviruses (Dengue, Japanese Encephalitis) (Chaturvedi *et al*., 1980; Kerdpanich *et al*., 2001).

ZIKV contains a positive-sense, single-stranded RNA genome of about 11 kilobases in length containing 59 bp and 39 bp untranslated regions flanking a single open reading frame (ORF) (Kuno and Chang, 2007). Whole genome sequencing of ZIKV (strain H/PF/2013) infecting a 51-year-old woman returning from French Polynesia revealed that the ORF encodes a polyprotein with three structural proteins, capsid

(105 amino acids (aa), premembrane/membrane (187 aa), and envelope (505 aa), and seven non-structural proteins, NS1 (352 aa), NS2A (217 aa), NS2B (139 aa), NS3 (619 aa), NS4A (127 aa), NS4B (255 aa), and NS5 (904 aa) (Baronti *et al*., 2014). The envelope protein includes a 154 glycosylation motif associated with virulence (Faye *et al*., 2014). The 5′ end of positive strand genomic RNA is modified with a cap-1 structure (me7-GpppA-me2) formed by an RNA triphosphatase (encoded by NS3), with guanylyl transferase, N7-methyl transferase and 2′-O methyl transferase (encoded by NS5) (Henderson *et al*., 2011).

Currently there is no vaccine available against ZIKV or specific antiviral treatment for clinical ZIKV infection. 'Zikavac', a candidate vaccine produced by Bharat Biotech is ready for pre-clinical trials (Bagla, 2016). Vaccine development generally is a cumbersome process which needs high bio-safety level containment facilities and in some cases faces difficulties with virus yield in cell culture. In this scenario, computational immunology methods may effectively be used for screening out potential epitopes from available viral genome databases (De Groot *et al*., 2001) that could facilitate their synthetic production along with consistent cost and quality (Testa and Philip, 2012). Many web based publicly available immunoinformatics tools and servers may be used in accurate identification of T cell epitopes (Khan *et al*., 2006). Keeping these points in view, this study was envisaged to prospect potential epitopes from ZIKV polyprotein towards the development of T-cell based candidate vaccines.

## MATERIALS AND METHODS

### *Retrieval of polyprotein sequence*

The FASTA formatted amino acid sequence of ZIKV polyprotein (GenBank: AMD16557.1) was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/). This viral strain was isolated from the amniotic fluid sample of a pregnant woman from the state of Paraíba in Brazil whose fetus had been diagnosed with microcephaly (Calvet *et al*., 2016). The woman tested negative for dengue virus, chikungunya virus, *Toxoplasma gondii*, rubella virus, cytomegalovirus, herpes simplex virus, HIV, *Treponema pallidum*, and parvovirus B19; however showed clinical manifestations like fever, myalgia and rash that attributes to ZIKV infection. House

### *T-cell epitope prediction*

The Immune Epitope Database-Analysis Resource (IEDB-AR) was used for the identification of the T cell epitope. Presentation of peptides MHC-I complexes to T lymphocytes is a multistep process which involves peptide- MHC-I binding, proteasomal C terminal cleavage and TAP transport. The database houses several tools which can integrate analysis of this process into one prediction. For the binding analysis, 74 most frequently occurring alleles harboured in IEDB-AR were selected, and the length of the peptide was set to 9. A tool NetMHCpan housed in IEDB-AR was used to calculate the half-maximal inhibitory concentration ($IC_{50}$) values (cut off score ≤ 200 nM) of epitope binding to MHC-I molecules.

The protein sequences were also screened individually using SYFPEITHI with cut off score adjusted above 20 (Rammensee *et al*., 1999). To predict the antigenicity nature of the predicted epitopes, the sequences were then analyzed as per Kolaskar and Tongaonkar Antigenicity method (Kolaskar and Tongaonkar, 1999). All the antigenic proteins with their respective predicted score were then filtered and a single antigenic protein with highest antigenicity score was selected for further evaluation.

### *Screening of potential epitopes*

To screen out efficient epitopes from a pool of predicted consensus sequences from IEDB-AR and SYFPEITHI, the following parameters *viz.,* antigenicity, conservancy and population coverage were undertaken. BLASTP search was performed to check the conservancy of the selected epitopes within the human proteome; epitopes with 100% identity to human proteins were excluded from further evaluation. Epitope conservancy among different strains plays a pivotal role in terms of providing protection across the species (Bui *et al*., 2006). For the analysis of the epitope conservancy, the web based epitope conservancy tool (http://tools.immuneepitope.org/tools/conservancy/iedb_input ) was used considering various African lineage (GenBank accession no.: KF268948, KF268950, KF268949, AY632535, and HQ234501) and Asian lineage (GenBank accession no.: KF993678, JN860885, EU545988, and HQ234499) strains as reference. Population coverage for each individual epitope was selected by the IEDB population coverage calculation tool analysis resource. Here we used the allelic frequency of the interacting HLA alleles for the prediction of the population coverage for the corresponding epitope.

### *Molecular docking study*

PEPstr web server (Kaur *et al*., 2007) was used to predict the three dimensional structure of the selected epitopes. HLA-A*0201 allele was used as a reference allele for docking study as most of the candidate epitopes show their binding potency towards this allele as evident from Table 1. Subsequently the crystal structure of HLA-A*0201 allele was retrieved from PDB (PDB Id: 1I4F). The structure was co-crystallised with a tumor-specific 10mer antigenic peptide. Prior to docking, the water molecules as well as the 10mer peptide was removed and was used as a positive control to compare the binding affinity of the selected epitopes. ZDOCK server (Pierce *et al*., 2014) with default settings was used to predict optimal binding mode of all the selected epitopes with HLA-A*0201 considering mouse H-2Kb-restricted peptide VSV8 (RGYVYQGL) (Tsukahara *et al*., 2009) as negative control. Reproducibility of our ZDOCK based docking strategy was cross-validated by re-docking the co-crystal peptide into the peptide binding groove of HLA-A*0201. PatchDock web server which access the surface flexibility attained by intermolecular interactions was used to assess the binding affinity of the docked complexes.

## RESULTS

Initially 1,798 prospective epitopes were computed from ZIKV polyprotein based on IEDB threshold ($IC_{50} \leq 200$ nM). The highest number of epitopes (403 epitopes) was generated from NS4B gene followed by NS5 (339 epitopes), NS3 (262 epitopes), NS3 (262 epitopes), NS2A (189 epitopes), NS1 (109 epitopes), NS2B (103 epitopes), NS4A (82 epitopes), whereas 197 and 70 epitopes generated from envelope and capsid protein, respectively. The lowest number of epitopes (44 epitopes) was generated from the membrane protein.

**Table 1** Various attributes of top five selected peptides from each protein of ZIKV

| Protein | Epitope and Position | Top restricted HLA type | IEDB Score ($IC_{50}$ in nm) | Other HLA alleles recognizing the epitope | No. of HLA recognizing the epitope | Theoretical population coverage of epitope (%) | Theoretical population coverage of epitope (%) |
|---|---|---|---|---|---|---|---|
| **Capsid** | RMVLAILAF (45-53) | HLA-B*15:25 | 5.17 | HLA-B*15:01, HLA-A*32:01, HLA-C*03:02, HLA-B*15:02 | 5 | 18.83 | 79.15 |
| | MVLAILAFL (46-54) | HLA-A*02:06 | 7.75 | HLA-A*68:02, HLA-A*02:01, HLA-A*26 | 4 | 42.53 | |
| | RVSPFGGLK (23-31) | HLA-A*11:01 | 9.42 | HLA-A*30:01, HLA-A*03:01, HLA-A*31:01 | 4 | 38.48 | |
| | AMLRIINAR (90-98) | HLA-A*31:01 | 9.94 | HLA-A*33:03 | 2 | 10.01 | |
| | GFRIVNMLK (10-18) | HLA-A*30:01 | 14.73 | HLA-A*31:01 | 2 | 9.14 | |
| | ISFPTTLGM (21-29) | HLA-C*03:02 | 8.29 | HLA-C*12:03, HLA-C*12:02, HLA-B*58:01, HLA-B*58:02, HLA-B*15:16, HLA-B*15:25, HLA-C*15:02, HLA-C*16:01, HLA-C*03:03, HLA-C*03:04, HLA-C*02:02, HLA-C*02:09, HLA-B*35:01 | 14 | 58.69 | |
| **Propeptide** | HMCDATMSY (41-49) | HLA-B*15:25 | 8.64 | HLA-A*29:02, HLA-B*15:02, HLA-B*35:01, HLA-A*30:02, HLA-B*15:01, HLA-C*03:02, HLA-A*11:01 | 8 | 39.03 | 84.62 |
| | VTRRGSAYY (1-9) | HLA-A*30:02 | 10.13 | HLA-A*30:01, HLA-B*15:25, HLA-A*29:02, HLA-B*15:01, HLA-A*01:01, HLA-B*15:16 | 7 | 34.05 | |
| | ATMSYECPM (45-53) | HLA-C*03:02 | 24.47 | HLA-C*03:03, HLA-C*03:04, HLA-B*15:25, HLA-A*02:06 | 5 | 25.69 | |
| | VVYGTCHHK (74-82) | HLA-A*11:01 | 28.39 | HLA-A*03:01, HLA-A*68:01 | 3 | 35.75 | |
| **Envelope** | KSLFGGMSW (454-462) | HLA-B*58:01 | 2.75 | HLA-B*57:01, HLA-A*32:01, HLA-B*58:02, HLA-B*15:25, HLA-C*03:02 | 6 | 17.78 | 94.68 |
| | TTVSNMAEV (48-56) | HLA-A*68:02 | 3.78 | HLA-A*02:03, HLA-A*02:06, HLA-A*26:01, HLA-B*15:16, HLA-C*12:03, HLA-C*15:02, HLA-C*03:02, HLA-C*12:02, HLA-C*03:03, HLA-C*03:04, HLA-A*26:01, HLA-C*02:02, HLA-C*02:09 | 14 | 53.46 | |
| | CTAAFTFTK (308-316) | HLA-A*68:01 | 4.56 | HLA-A*11:01, HLA-A*03:01, HLA-*31:01, HLA-A*30:01, HLA-A*33: 03 | 6 | 46.63 | |
| | MMLELDPPF (374-382) | HLA-B*15:25 | 4.56 | HLA-B*15:02, HLA-B*35:01, HLA-B*15:01, HLA-A*32:01, HLA-A*02:06, HLA-C*03:02, HLA-B*53:01, HLA-A*02:01, HLA-A*29:02, HLA-C*03:03, HLA-C*03:04, HLA-A*23:01 | 13 | 70.71 | |
| **Membrane** | DTAWDFGSV (426-434) | HLA-A*68:02 | 4.73 | HLA-A*26:01, HLA-B*15:16 | 3 | 8.62 | 76.99 |
| | IYLVMILLI (62-70) | HLA-A*23:01 | 9.81 | HLA-A*24:02 | 2 | 26.18 | |
| | LPSHSTRKL (4-12) | HLA-B*07:02 | 16.11 | HLA-B*35:01, HLA-B*51:01 | 3 | 27.23 | |
| | MILLIAPAY (66-74) | HLA-A*29:02 | 20.49 | HLA-B*35:01, HLA-B*15:25, HLA-B*15:02, HLA-A*30:02, HLA-A*03:01 | 6 | 32.02 | |
| | RSQTWLESR (15-23) | HLA-A*31:01 | 21.14 | - | 1 | 5.36 | |
| **NS1** | VTLPSHSTR (2-10) | HLA-A*31:01 | 24.71 | HLA-A*68:01, HLA-A*33:03, HLA-A*11:01, HLA-B*15:16 | 5 | 30.05 | 68.89 |
| | CTMPPLSFR (315-323) | HLA-A*68:01 | 3.49 | HLA-A*33:03, HLA-A*31:01, HLA-A*11:01, HLA-A*74:01, HLA-A*03:01 | 6 | 45.36 | |
| | SPRRLAAAV (37-45) | HLA-B*07:02 | 3.97 | HLA-B*08:01 | 2 | 22.61 | |
| | KSYFVRAAK (119-127) | HLA-A*30:01 | 4.19 | HLA-A*31:01, HLA-A*11:01, HLA-A*03:01 | 4 | 38.48 | |
| | LAAAVKQAW (41-49) | HLA-B*58:01 | 5.19 | HLA-B*57:01, HLA-B*53:01, HLA-C*03:02 | 4 | 12.40 | |
| **NS2A** | IPKSLAGPL (242-250) | HLA-B*07:02 | 6.82 | HLA-B*08, HLA-B*35:01 | 3 | 20.62 | 85.40 |
| | HLALIAAFK (65-73) | HLA-A*68:01 | 9.07 | HLA-A*03:01, HLA-A*11:01 | 3 | 35.75 | |
| | VMALGLTAV (187-195) | HLA-A*02:01 | 9.29 | HLA-A*02:06 | 2 | 40.60 | |
| | VSFIFRANW (80-88) | HLA-B*58:01 | 10.07 | HLA-B*57:01, HLA-B*58:02, HLA-A*32:01 | | 13.82 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | VSFIFRANW (80-88) | HLA-B*58:01 | 10.07 | HLA-B*57:01, HLA-B*58:02, HLA-A*32:01 | | 13.82 | |
| | LAILMGATF (46-54) | HLA-C*03:02 | 10.25 | HLA-B*15:25, HLA-C*03:03, HLA-C*03:04, HLA-B*15:01, HLA-B*35:01, HLA-B*15:02, HLA-B*58:01, HLA-B*58:02, HLA-C*12:03 | 10 | 49.28 | |
| | LLVSFIFRA (78-86) | HLA-A*02:01 | 11.11 | HLA-A*02:06 | 2 | 40.60 | |
| | FAAGAWYVY (113-121) | HLA-B*35:01 | 1.79 | HLA-C*03:02, HLA-A*29:02, HLA-B*15:02, HLA-B*15:25, HLA-C*12:03, HLA-C*16:01, HLA-C*12:02, HLA-B*53:01, HLA-C*03:03, HLA-C*03:04, HLA-B*15:01, HLA-A*30:02, HLA-C*14:02, HLA-B*58:01, HLA-A*68:01, HLA-C*02:02, HLA-C*02:09, HLA-B*46:01, HLA-A*26:01 | 20 | 70.23 | |
| **NS2B** | MTICGMNPI (101-109) | HLA-A*68:02 | 2.49 | HLA-A*02:06, HLA-C*03:02, HLA-C*03:03, HLA-C*03:04, HLA-C*12:03, HLA-A*02:01, HLA-C*15:02, HLA-C*12:02, HLA-C*16:01, HLA-A*32:01, HLA-A*25:01, HLA-A*26:01, HLA-B*15:25, HLA-B*15:16, HLA-C*14:02 | 16 | 74.21 | 89.54 |
| | EMAGPMAAV (24-32) | HLA-A*68:02 | 5.43 | HLA-A*02:06, HLA-A*02:01 | 3 | 42.53 | |
| | IPFAAGAWY (111-119) | HLA-B*35:01 | 6.26 | HLA-B*53:01, HLA-B*15:02 | 3 | 13.16 | |
| | REIILKVVL (92-100) | HLA-B*40:01 | 9.89 | HLA-B*40:02, HLA-B*49:01, HLA-B*27:05, HLA-B*18, HLA-B*37 | 6 | 17.81 | |
| | RAWSSGFDW (326-334 | HLA-B*58:01 | 2.77 | HLA-B*57:01, HLA-B*53:01, HLA-A*32:01, HLA-B*58:02 | 5 | 16.11 | |
| | EAAAIFMTA (292-300) | HLA-A*68:02 | 3.25 | - | 1 | 2.50 | |
| **NS3** | MTATPPGTR (298-306) | HLA-A*68:01 | 4.51 | HLA-A*33:03, HLA-A*31:01, HLA-A*11:01 | 4 | 29.76 | 83.44 |
| | GVFHTMWHV (27-35) | HLA-A*02:06 | 4.74 | HLA-A*02:01, HLA-A*68:02 | 3 | 42.53 | |
| | FPDSNSPIM (309-317) | HLA-B*35:01 | 5.49 | HLA-B*53:01, HLA-B*35:03, HLA-C*03:02, HLA-C*03:03, HLA-C*03:04, HLA-B*07:02, HLA-C*08:02, HLA-C*16:01 | 9 | 48.65 | |
| | NQMAIIIMV (129-137) | HLA-A*02:06 | 3.74 | HLA-A*02:01, HLA-A*68:02, HLA-B*39:01, HLA-B*13 | 5 | 4.60 | |
| | MAIIIMVAV (131-139) | HLA-A*68:02 | 5.13 | HLA-A*02:06, HLA-C*03:03, HLA-C*03:04, HLA-C*03:02, HLA-C*12:03, HLA-C*12:02, HLA-B*51:01 | 8 | 42.41 | |
| **NS4A** | TVSLGIFFV (60-68) | HLA-A*68:02 | 8.25 | HLA-A*02:06, HLA-A*02:01 | 3 | 42.53 | 67.98 |
| | MLLGLLGTV (53-61) | HLA-A*02:01 | 9.09 | HLA-A*02:06 | 2 | 40.60 | |
| | LGASAWLMW (85-93) | HLA-B*58:01 | 9.84 | HLA-B*57:01 | 2 | 7.26 | |
| | ITAATSTLW (199-207) | HLA-B*58:01 | 1.93 | HLA-B*57:01, HLA-A*32:01, HLA-B*58:02, HLA-B*53:01 | 5 | 16.11 | |
| **NS4B** | YAWDFGVPL (87-95) | HLA-C*03:03 | 3.3 | HLA-C*03:04, HLA-C*03:02, HLA-C*12:03, HLA-C*16:01, HLA-A*02:06, HLA-C*12:02, HLA-B*35:01, HLA-B*39:01, HLA-A*02:01, HLA-C*08:01, HLA-C*14:02, HLA-C*15:02, HLA-A*68:02, HLA-B*15:25, HLA-B*15:02, HLA-C*02:02, HLA-C*02:09, HLA-B*35:03, HLA-A*51:01 | 20 | 79.33 | 83.16 |
| | ITAATSTLW (199-207) | HLA-B*58:01 | 1.93 | HLA-B*57:01, HLA-A*32:01, HLA-B*58:02, HLA-B*53:01 | 5 | 16.11 | |
| | YLIPGLQAA (122-130) | HLA-A*02:06 | 4.19 | HLA-A*02:01, HLA-C*03:02, HLA-C*12:03 | 4 | 48.41 | |
| | MAMATQAGV (68-76) | HLA-A*68:02 | 4.39 | HLA-A*02:06, HLA-C*03:02, HLA-C*03:03, HLA-C*03:04, HLA-C*12:03, HLA-A*02:01, HLA-C*12:02, HLA-C*15:02, HLA-C*16:01, HLA-A*51:01 | 11 | 66.76 | |
| | MSMVSSWLW (124-132) | HLA-B*58:01 | 1.39 | HLA-B*57:01, HLA-B*53:01, HLA-B*58:02, HLA-A*32:01, HLA-A*23:01 | 6 | 20.78 | |
| | IAMTDTTPY (91-99) | HLA-B*35:01 | 2.61 | HLA-C*03:02, HLA-B*15:25, HLA-B*15:02, HLA-C*03:03, HLA-C*03:04, HLA-C*12:03, HLA-A*29:02, HLA-B*15:01, HLA-C*12:02, HLA-C*16:01, HLA-B*53:01, HLA-A*30:02, HLA-B*58:01, HLA-C*14:02 | 15 | 58.83 | |
| **NS5** | YAQMWQLLY (509-517) | HLA-A*29:02 | 3.64 | HLA-C*03:02, HLA-B*35:01, HLA-A*01:01, HLA-C*12:03, HLA-B*15:25, HLA-A*30:02, HLA-C*16:01, HLA-B*15:02, HLA-C*12:02, HLA-B*58:01, HLA-C*14:02, HLA-B*15:01, HLA-B*53:01 | 14 | 56.06 | 88.52 |
| | YMDYLSTQV (631-639) | HLA-A*02:01 | 4.1 | HLA-A*02:06, HLA-C*05:01, HLA-C*03:02, HLA-C*08:01, HLA-C*08:02 | 6 | 51.89 | |
| | WFFDENHPY (42-50) | HLA-A*29:02 | 4.31 | HLA-B*35:01, HLA-B*15:02, HLA-C*03:02, HLA-C*14:02, HLA-A*30:02, HLA-B*15:25 | 7 | 22.68 | |

To obtain the most promiscuous MHC-I restricted epitopes, the top five $IC_{50}$ value candidates from each protein were selected for further analysis (Table 1). These data suggested that the selected epitopes are efficient and can be presented by multiple HLA alleles to the $CD8^+$. Population coverage study revealed that the total population coverage for the combination of all 5 promiscuous epitopes from envelope protein was 94.68%. The total population coverage for the combination of all promiscuous epitopes of NS2B, NS5, NS2A, propeptide, NS3, NS4B, membrane, NS1 and NS4A was 89.54%, 88.52%, 85.40% , 84.62%, 83.44%, 83.16%, 76.99%, 68.89%, 67.98% respectively (Table 1). Although the top five epitopes from each component proteins exhibited high IEDB and SYFPEITHI score, their individual global population coverage was low to further process them as ideal vaccine candidate. From this study, it was prominent that the following epitopes *viz.,* E1: MMLELDPPF from envelope, E2: FAAGAWYVY and E3: MTICGMNPI from NS2B, E4: YAWDFGVPL and E5: MAMATQAGV from NS4B had the highest individual population coverage (70.71%, 70.23%, 74.21%, 79.33%, and 66.76% respectively). Consequently, these epitopes were further carried forward for epitope conservancy, antigenicity and molecular docking study.

Results from BLASTP analysis concluded that none of the selected epitopes showed 100% conservancy within the human proteome. However, epitope conservancy analysis within the reported ZIKV lineages showed that only E1, E2 and E5 are 100% conserved with African and Asian lineage strains (Table 2). E4 showed 100% conservancy within the Asian lineage, while 80% with the African lineage. In contrast, E3 did not share any identity with the African lineage strains (75% conservancy within the Asian lineage). Further results from IEDB antigenicity prediction analysis suggested that all the selected epitopes were antigenic in nature (lower percentile rank < 1.00).

**Table 2** Conservancy analysis of top five promiscuous epitopes within African and Asian lineage strains. GenBank accession numbers of the representative strains for both the lineages are shown.

| Epitope id | EPITOPE | Percentage of similarity | |
|---|---|---|---|
| | | African lineage (KF268948, KF268950, KF268949, AY632535, HQ234501) | Asian lineage (KF993678, JN860885, EU545988, HQ234499) |
| **E1** | MMLELDPPF | 100 | 100 |
| **E2** | FAAGAWYVY | 100 | 100 |
| **E3** | MTICGMNPI | 0 | 75 |
| **E4** | YAWDFGVPL | 80 | 100 |
| **E5** | MAMATQAGV | 100 | 100 |

Structural superimposition of the docked pose of co-crystal peptide obtained from ZDOCK server and the co-crystal conformation showed a RMSD of 0.001 Å. The poses having an RMSD < 2 Å are considered as success, and the poses having RMSDs between 2 and 3 Å are considered as partial successes (Cole *et al*., 2005). Furthermore, all the interactions and interacting residues in the docked pose are identical with the co-crystal conformation. This low RMSD difference and similar interactions of participating residues between the docked pose and co-crystal conformation validated our docking protocol. Docking study revealed that all the selected epitopes bind between the α1 and α2 helices of HLA-A*0201 as in the case of control peptide (Figure 1).

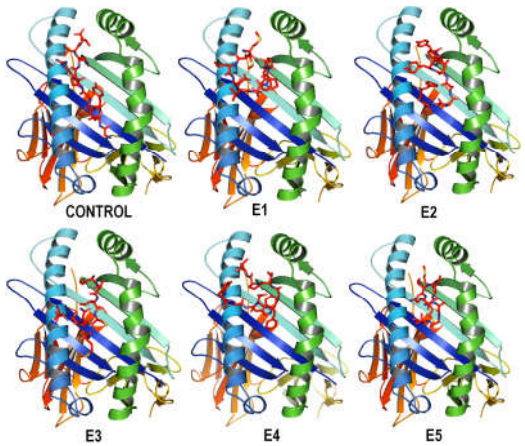


**Figure 1** HLA-A*0201 binding mode of all the selected promiscuous epitopes along with the positive control. The protein is represented with cartoon and epitopes are represented with stick in Pymol.

The negative control VSV8 did not bind to the anticipated peptide binding groove of HLA-A*0201. This finding was further supported by the results obtained from PatchDock web server. The Geometric shape complementarity score for E1, E2, E3, E4 and E5 epitopes were 7,196, 7,102, 7,460, 7,212 and 6,478 respectively which was very close to the score of positive control peptide (7,998). The Geometric shape complementarity score of negative control was 1,256 which showed very low binding affinity with HLA-A*0201 molecule.

## DISCUSSION

Development of immunity against viral infection is mediated by a variety of specific and non-specific immune response mechanisms (Pulendran and Ahmed, 2011). In case of specific immune mechanism, generation of $CD8^+$ cytotoxic T lymphocytes (CTL) immune response depends on specific antigen presentation by class I Major Histocompatibility Complex (MHC) molecules. Both structural and non-structural proteins have been demonstrated to induce CTL immune response against viral infections (Warfield *et al*., 2011). Most of the nonstructural proteins of flavivirus are involved in the formation of replication complexes essential for viral replication and progression and therefore can be useful candidate for vaccine development (Akey *et al*., 2014; Kapoor *et al*., 1995; Motolla *et al*., 2002). Epitopes derived from multiple proteins *i.e.,* structural and nonstructural elicit cellular and humoral immune response and have been investigated in preclinical and clinical trials (De Groot *et al*., 2001; Ferrantelli *et al*., 2004). At present, reports on epitopes of ZIKV that stimulate the CTL response are scanty. With the advent of immunoinformatics strategies, mining of promiscuous epitopes from multiple proteins is in routine use (Khan *et al*., 2006). Therefore, the present study was instrumental in identifying potential epitopes that preferentially bind to MHC class I molecules towards CTL response. For this, polyprotein derived from a circulating ZIKV strain from Brazil was considered. Initially a pool of 1,798 consensus epitopes was derived from the ZIKV polyprotein using IEDB and SYFPEITHI server. HLA alleles are highly polymorphic and are expressed at different frequencies in different ethnicities (Janeway *et al*., 2004). On that account, binding to a diverse array of HLA alleles should be considered as major criteria for screening of T-cell based epitopes (Maenaka and Jones, 1999; Stern and Wiley *et al*.,

1994). Based on the above criteria, five most promiscuous epitope *viz.,* E1: MMLELDPPF, E2: FAAGAWYVY, E3: MTICGMNPI, E4: YAWDFGVPL and E5: MAMATQAGV exhibiting more than 60% individual population coverage across the globe were selected. Out of 74 HLA alleles included in the analysis, these five epitopes (E1-E5) were found to bind at least 10-19 different types accounting for 89.54% cumulative global population coverage. These epitopes were found to be highly antigenic in IEDB analysis. This affirms their potential as possible candidates for vaccine design. They did not show any conservancy with the human proteome suggesting them to be immunogenic. They were further tested for their conservancy within the African and Asian lineages in IEDB conservancy analysis tool. Result from IEDB indicated that E1, E2 and E5 are fully conserved within both the lineages. We assume these three epitopes can cover a wide geographical and demographical range upon ZIKV infection. Previously, epitope MMLELDPPF (E1) derived from ZIKV envelope protein has been reported to be the most immunogenic $CD8^+$ T cell epitope (Shawan *et al*., 2014). This epitope was also identified in the present study. The other four epitopes, *i.e.,* E2, E3, E4 and E5 derived from nonstructural proteins were also included in our study as these were able to generate substantial *in silico* MHC-I binding efficiency. Results from docking study confirmed that all the selected epitopes showed good binding affinity into the peptide binding groove of HLA-A*0201. Geometric shape complementarity score for all the epitopes obtained from Patchdock server was at par with that of the control peptide. This signifies the HLA-A*0201 binding potential of all the selected epitopes which may generate substantial CTL response as in the case of control epitope.

Based on the current immunoinformatics study, we propose E1, E2 and E5 as possible global candidate epitopes for vaccine development against ZIKV infection. E3 and E4 may also be targeted for alternate formulation; however, geographical and demographical features have to be taken into consideration. All the results obtained in the present study are based on various *in silico* tools and servers which warrant further *in vitro* and *in vivo* evaluations of the selected epitopes to confirm their efficacy.

## CONCLUSION

We screened out 5 promiscuous peptides unique to Zika virus (ZIKV) which can competently be presented by MHC class I alleles. The epitopes exhibited wider population coverage and a higher degree of conservancy among the representative lineage strains. Antigen presentation by MHC class I alleles is crucial to invoke cell mediated immune response; therefore in this line we propose that these peptides may further be subjected for candidate vaccine development. However, experimental studies have to be conducted to validate this claim.

**Acknowledgement:** The authors gratefully acknowledge the extra-mural funding of Indian Council of Medical Research, Government of India.

**Compliance with Ethical Standards**

**Disclosure of potential conflicts of interest:** All the authors declare that they have no conflict of interest.

**Research involving human participants and/or animals:** This article does not contain studies with human or animal subjects performed by any of the authors that should be approved by Ethics Committee.

**Informed consent:** The article does not contain any studies in patients by any of the authors.

## References

1. Faye, O. Freire, C.C., Iamarino, A., Faye, O., de Oliveira, J.V., Diallo, M., Zanotto, P.M.A. and Sall, A.A. (2014) Molecular evolution of Zika virus during its emergence in the 20th century. PLoS. Negl. Trop. Dis., 8:e2636.
2. Zanluca, C., de Melo, V.C., Mosimann, A.L., Dos Santos, G.I., Dos Santos, C.N., and Luz, K. (2015) First report of autochthonous transmission of Zika virus in Brazil. Mem. Inst. Oswaldo. Cruz., 110:569-572.
3. Campos, G.S., Bandeira, A.C., and Sardi, S.I. (2015) Zika Virus Outbreak, Bahia, Brazil. Emerg. Infect. Dis., 21:1885-1886.
4. Musso, D., Nilles, E.J., and Cao-Lormeau, V.M. (2014) Rapid spread of emerging Zika virus in the Pacific area. Clin. Microbiol. Infect., 20:595-596.
5. Besnard, M., Lastère, S., Teissier, A., Cao-Lormeau, V.M., and Musso, D. (2014) Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. Euro. Surveill., 19:1-4.
6. Duffy, M.R., Chen, T.H., Hancock, W.T., Powers, A.M., Kool, J.L., Lanciotti, R.S., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C., Guillaumot, L., Griggs, A., Bel, M., Lambert, AJ., Laven, J., Kosoy, O., Panella, A., Biggerstaff, BJ., Fischer, M., and Hayes, E.B. (2009) Zika Virus Outbreak on Yap Island., Federated States of Micronesia. N. *Engl. J. Med*., 360:2536-2543.
7. Cao-Lormeau, V.M., Roche, C., Teissier, A., Robin, E., Berry, AL., Mallet, HP., Sall, A.A., and Musso, D. (2014) Zika virus, French Polynesia, South Pacific, 2013. Emerg. Infect. Dis., 20:1085-1086.
8. Haymaker, W.K.J. (1949) The Landry-Guillain-Barré syndrome: a clinicopathologicic report of fifty fatal cases and a critique of the literature. Medicine., 28:59-141.
9. Musso, D., Cao-Lormeau, V., and Gubler, D. (2015) Zika virus: following the path of dengue and chikungunya? Lancet., 386:243-244.
10. Ministério da Saúde (Brazil) Microcefalia - Ministério da Saúde divulga boletim epidemiológico [Internet] Brasília: Ministério da Saúde; 2015 [accessed on 2016 Feb 12]
11. Pan American Health Organization (PAHO)/World Health Organization (WHO) Epidemiological Alert: Increase of microcephaly in the northeast of Brazil., 17 November 2015 [accessed on 2016 Feb 12]
12. Calvet, G., Aguiar, R.S., Melo, A.S., Sampaio, S.A., de Filippis, I., Fabri, A., Araujo, E.S., de Sequeira, P.C., de Mendonça, M.C., de Oliveira, L., Tschoeke, D.A., Schrago, C.G., Thompson, F.L., Brasil, P., Dos Santos, F.B., Nogueira, R.M., Tanuri, A., and de Filippis AM (2016) Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a

case study. Lancet. Infect. Dis., doi: 101016/S1473-3099(16)00095-5.
13. Chaturvedi, U.C., Mathur, A., Chandra, A., Das, S.K., Tandon, H.O., and Singh, U.K. (1980) Transplacental infection with Japanese encephalitits virus. *J. Infect. Dis*., 141:712-715.
14. Kerdpanich, A., Watanaveeradej, V., Samakoses, R., Chumnanvanakij, S., Chulyamitporn, T., Sumeksri, P., Vuthiwong, C., Kounruang, C., Nisalak, A., and Endy, T. (2001) Perinatal dengue infection Southeast Asian. J. Trop. Med. Public., Health 32:488-493.
15. Kuno, G., and Chang, G.J. (2007) Full-length sequencing and genomic characterization of Bagaza., Kedougou., and Zika viruses. Arch. Virol., 152:687-696.
16. Baronti, C., Piorkowski, G., Charrel, R.N., Boubis, L., Leparc-Goffart, I., and de Lamballeriea, X. (2014) Complete Coding Sequence of Zika Virus from a French Polynesia Outbreak in 2013. Genome. Announc., 2:e00500-14.
17. Henderson, B.R., Saeedi, B.J., Campagnola, G., and Geiss, B.J. (2011) Analysis of RNA binding by the Dengue virus NS5 RNA capping enzyme. PLoS. ONE., 6:e25795.
18. Bagla, P. How Bharat Biotech Made Its Breakthrough In Developing A Vaccine For Zika Virus Huffington Post (New Delhi) PTI (accessed on 11 February 2016)
19. De Groot, A.S., Bosma, A., and Chinai, N. (2001) From genome to vaccine: In silico predictions., ex vivo verification. Vaccine., 19:4385-4395.
20. Testa, J.S., and Philip, R. (2012) Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. Future. Virol., 7:1077-1088.
21. Khan, A.M., Heiny, A.T., Lee, K.X., Srinivasan, K.N., Tan, T.W., August, J.T., and Brusic, V. (2006) Large-scale analysis of antigenic diversity of T cell epitopes in dengue virus. BMC. Bioinformatics., 7:54.
22. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics., 50:213-219.
23. Kolaskar, A.S., and Tongaonkar, C.P. (1999) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS. Lett., 276:172-174.
24. Bui, H.H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., and Sette, A. (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC. Bioinforma., 7:153.
25. Kaur, H., Garg, A., and Raghava, G.P. (2007) PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. Protein. Pept. Lett., 14:626-631.
26. Pierce, B.G., Wiehe, K., Hwang, H., Kim, BH., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics., 30:1771-1773.
27. Tsukahara, T., Awaguchi, S., Torigoe, T., Takahashi, A., Murase, M., Kano, M., Wada, T., Kaya, M., Nagoya, S., Yamashita, T., and Sato, N. (2009) HLA-A*0201-restricted CTL epitope of a novel osteosarcoma antigen, papillomavirus binding factor. *J. Transl. Med*., 7:44.
28. Cole, J.C., Murray, C.W., Nissink, J.W.M., Taylor, R.D., and Taylor, R. (2005) Comparing protein-ligand docking programs is difficult. Proteins., 60:325-332.
29. Pulendran, B., and Ahmed, R. (2011) Immunological mechanisms of vaccination. Nat. Immunol., 12:509-517.
30. Warfield, K.L., and Olinger, G.G. (2011) Protective role of cytotoxic T lymphocytes in filovirus hemorrhagic fever. Biomed. Res. Int., 984241 doi: 101155/2011/984241.
31. Akey, D.L., Brown, W.C., Dutta, S., Konwerski, J., Jose, J., Jurkiw, T.J., DelProposto, J., Ogata, C.M., Skiniotis, G., Kuhn, R.J., and Smith, J.L (2014) Flavivirus NS1 Structures Reveal Surfaces for Associations with Membranes and the Immune System. Science., 343:881-885.
32. Kapoor, M., Zhang, L., Ramachandra, M., Kusukawa, J., Ebner, K.E., and Padmanabhan, R. (1995) Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J. Biol. Chem*., 270:19100-19106,
33. Mottola, G., Cardinali, G., Ceccacci, A., Trozzi, C., Bartholomew, L., Torrisi, M.R., Pedrazzini, E., Bonatti, S., and Migliaccio, G. (2002) Hepatitis C virus nonstructural proteins are localized in a modified endoplasmic reticulum of cells expressing viral subgenomic replicons. Virology., 293:31-43,
34. Ferrantelli, F., Rasmussen, R.A., Buckley, K.A., Li, P.L., Wang, T., Montefiori, D.C., Katinger, H., Stiegler, G., Anderson, D.C., McClure, H.M., and Ruprecht, R.M. (2004) Complete protection of neonatal rhesus macaques against oral exposure to pathogenic simian-human immunodeficiency virus by human anti-HIV monoclonal antibodies. *J. Infect. Dis*., 189:2167-2173.
35. Janeway, C.A. Jr., Shlomchik, M.J., Travers, P., and Walport, M. (2004) The major histocompatibility complex and its functions In: Immunobiology 6th ed London: Garland Publishing167-184.
36. Maenaka, K., and Jones, E.Y. (1999) MHC superfamily structure and the immune system. Curr. Opin. Struct. Biol. 9:745-53.
37. Stern, L.J., and Wiley, D.C. (1994) Antigenic peptide binding by class I and class II histocompatibility proteins. Structure. 2:245-251.
38. Shawan, Md. M.A.K., Mahmud, H.A., Hasan, Md. M., Parvin, A., Rahman, Md. N., and Rahman, S.M.B. (2014) In Silico Modeling and Immunoinformatics Probing Disclose the Epitope Based Peptide Vaccine Against Zika Virus Envelope Glycoprotein. *Indian. J. Pharm. Biol. Res.*, 2:44-57.

*******