

**Original Research Article*****In Silico* Modeling and Immunoinformatics Probing Disclose the Epitope Based Peptide Vaccine Against Zika Virus Envelope Glycoprotein**

Mohammad Mahfuz Ali Khan Shawan<sup>1\*</sup>, Hafij Al Mahmud<sup>1</sup>, Md. Mahmudul Hasan<sup>1</sup>, Afroza Parvin<sup>1</sup>, Md. Nazibur Rahman<sup>1</sup> and S. M. Badier Rahman<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

**ARTICLE INFO:****Article history:**

Received: 30 September 2014

Received in revised form:

15 October 2014

Accepted: 20 October 2014

Available online: 31 December 2014

**Keywords:**

Epitope,  
Vaccine,  
*in silico*,  
Immunoinformatics,  
Conservancy,  
Molecular docking.

**ABSTRACT**

Zika virus (ZIKV) is an aedes mosquito borne pathogen belonging to the member of flaviviridae subgroup is the causative agent of an emerging disease called Zika fever, known as a benign infection usually presenting as influenza like illness with cutaneous rash. Due to recent epidemic outbreaks it is realized as a major health risk which need enhanced surveillance, but no attempt has been made to design an epitope based peptide vaccine against Zika virus. Viral envelope proteins are derived from host cell membrane proteins with some viral glycoproteins and are used to cover their protective protein capsid, help the viruses to enter host cells and help them to avoid the host immune response. In this study, amino acid sequence of ZIKV envelope glycoprotein was obtained from a protein database and examined with *in silico* approaches to determine the most immunogenic epitopes for B cell and T cell which could induce humoral as well as cell mediated immune response. Both the linear and conformational epitopes for B cell were predicted by immunoinformatics tools housed in IEDB resources. The peptide sequence DAHAKRQTVVVVLSQEGAV from position 121 and peptide sequence from 117-137 amino acids were predicted as most potential B cell linear and conformational epitopes respectively. Epitopes for CD4+ and CD8+ T cell were also predicted by using tools within IEDB resource and peptide sequence MMLELDPPF from position 250-258 amino acids was predicted as most immunogenic CD8+ T cell epitope with immune response evoking ability prediction score (I pMHC) of 0.09139 and conservancy of 52.17%. The innate immune response for ZIKV envelope glycoprotein was determined by interferon (IFN)-gamma effectuation and mimicking capacity by immunoinformatics and molecular docking study respectively. However, this is an introductory approach to design an epitope based peptide vaccine against Zika virus; we hope this model will be very much helpful in designing and predicting novel vaccine candidate.

**Introduction**

Zika virus (ZIKV) a group IV, positive sense single stranded RNA (11 kilobases) arbovirus of *Flavivirus* genus belonging to Flaviviridae family is related to yellow fever, dengue, West Nile, St. Louis encephalitis and Japanese encephalitis viruses causing dengue fever like syndrome including mild headache, maculopapular rash, fever, malaise, conjunctivitis, and arthralgia [1,2]. Nowadays a lot of research efforts have enlightened on many of these viruses but other members of the

mosquito borne flaviviruses, such as ZIKV, have received far less attention. In 1947 the virus was first isolated from a febrile sentinel rhesus monkey (Rhesus 766) placed in a cage in the Zika forest of Uganda under the observation of Rockefeller Foundation's program for research on jungle yellow fever [2] while a second isolation was done from the mosquito *Aedes africanus* followed at the same site in January 1948 [1]. Most likely it is thought that, ZIKV maintained a sylvatic cycle involving non-human primates and several *Aedes* species (*Ae.africanus*, *Ae.aegypti* and others) as

\*Corresponding Author: Mohammad Mahfuz Ali Khan Shawan, Department of Biochemistry and Molecular Biology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh E-Mail: [mahfuz\\_026shawan@yahoo.com](mailto:mahfuz_026shawan@yahoo.com)

mosquito vectors [2-4], with cyclic epizootics in monkeys reported in Uganda [5-8]. Humans are get infected by infective mosquito bites, however recent report suggests that there is a possibility of secondary sexual transmission [9]. ZIKV is endemic in Africa and south-east Asia [2] and phylogenetic analysis revealed that African and Asian strains emerged as two distinct lineages [10-11]. In 2007, ZIKV has caused a large epidemic on Yap Island, Federated States of Micronesia and involved in infecting three quarter of local populations [12]. This outbreak shows that, ZIKV has been detected outside of Africa and Asia, having the potential as an emerging pathogen [9]. The viral illness (Zika fever) is an emerging disease due to the expanding distribution area of ZIKV, confirmed by the recent epidemic affecting French Polynesia, New Caledonian since October 2013 [10] and Cook Island in march, 2014.

Zika virus genome contains 59 and 39 untranslated regions flanking a single open reading frame (ORF) that encodes a polyprotein that is cleaved into three structural proteins: the capsid (C), premembrane/membrane (prM), and envelope (E); and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5) [13]. The 5' end of positive-strand genomic RNA is modified with a cap-1 structure (m<sup>7</sup>-GpppA-me<sub>2</sub>) formed by an RNA triphosphatase, with guanylyltransferase, N7-methyltransferase and 2'-O methyltransferase. The non structural proteins are responsible for these activities. The NS3 protein encodes a RNA triphosphatase within its helicase domain. The N-terminal domain of the non-structural protein 5 (NS5) has both the N7-methyltransferase and guanylyltransferase activities necessary for forming mature RNA cap structures. RNA binding affinity is reduced by the presence of ATP or GTP and enhanced by S-adenosyl methionine [14]. This protein also encodes a 2'-O methyltransferase.

Currently there is no medicine against Zika fever or specific antiviral treatment for clinical ZIKV infection is available. Someone protect him/herself by preventing mosquito bites [15]. So the development of a new vaccine against ZIKV is very much important and this development has not yet been achieved. Surface or envelope proteins of virus are most antigenic one and often considered as good candidates for immunization. It is particularly important for vaccine development as it mediates the viral entry and also the primary target of adaptive immune response [16, 17]. In vaccine development it would be very much helpful if the epitopes associated with envelope glycoprotein are well known, and that could facilitate their synthetic production along with consistent cost and quality advantages over the current treatment [18]. Development of immunity against viral infection is mediated by a variety of specific and non-specific immune mechanisms involves the stimulation of antigen specific CD8<sup>+</sup> cytotoxic T lymphocytes (CTL) and B cells [19]. Induction and development of effective immune response highly depends on specific and potent antigen presentation by human leukocyte antigen (HLA) class I alleles

(MHC class I molecule) which present antigenic peptides on infected cells to CTLs and class II alleles (MHC class II molecule) which presents peptide antigen to CD4<sup>+</sup> helper T (Th) cells and ultimately converge to B cell response [20-22]. During viral infection a non-specific immune response is also mediated by interferon (IFN) gamma, which increase CD8<sup>+</sup> cytotoxic T lymphocytes along with the enhancement of MHC class I and II dependent antigen presentation [23].

To design an effective synthetic peptide vaccine candidate, *in silico* modeling and immunoinformatics strategies have been exploited which uses a variety of statistical and machine learning approaches by the help of bioinformatics software and machine learning programs. Such active vaccine candidate must have to contain minimum two antigenic epitopes; one to induce specific B cell or CTL responses while other induce specific Th cell response. This study has important implications on computational tools to screen B and T cell epitopes as well as to assess the IFN-gamma inducing effect of ZIKV envelope glycoprotein, a critical step in the development of vaccines.

## Materials and methods

### Retrieval of Zika virus envelope glycoprotein sequence

Zika virus envelope (outer membrane) glycoprotein sequences were obtained from UniProtKB ([www.uniprot.org/](http://www.uniprot.org/)) in FASTA format [24].

### Protein antigenicity determination

To predict the protective antigens as vaccines, the sequence was then analyzed with VaxiJen [25] with default parameters to find out antigenicity. 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.

### Primary and secondary structure prediction

A proteomics server, ExPASy ProtParam ([www.expasy.org](http://www.expasy.org)) was used to analyze the primary structure of the target protein. Several parameters given by ProtParam tool for example molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) were examined. The secondary structure (alpha helix, beta plated sheets, turns and coils) of the protein was checked by using different servers like Jpred3 [26], GOR IV [27] and SOPMA [28] which aims to predict solvent accessibility, transmembrane helices, globular regions, and coiled-coil regions and finally determines the protein's stability and function.

### 3D structure modeling (homology modeling) and validation

The sequence homology approach to model 3D structure of a protein uses sequence alignment to identify the best matching 3D structure for different components: conserved portion, loop portion and side chains from the database, and threads them to predict the overall 3D structure [29]. The crystal or NMR structure of Zika virus envelope protein was not available in the Protein Data Bank (PDB), so that the sequence of envelope protein was used to develop the 3D structure through homology modeling.

For homology modeling we did blastp (protein-protein BLAST) by using PDB database and among 67 templates chose two with highest hit (template 1: pdb accession 3P54\_A and template 2: pdb accession 4FG0\_A). Then 3D structure of the best template for Zika virus envelope glycoprotein was generated by using protein structure modeling server SWISS-MODEL ([swissmodel.expasy.org](http://swissmodel.expasy.org)) and HHpred a protein function and protein structure prediction server ([toolkit.tuebingen.mpg.de/hhpred](http://toolkit.tuebingen.mpg.de/hhpred)) along with their confidence score. After generating the 3D models, structure analysis and stereochemical analysis were performed and best 3D model was selected by using evaluation and validation tools QMEAN, ProSA, PROCHECK, RAM-PAGE and SAVES online tool (<http://services.mbi.ucla.edu/SAVES/>) [30-34].

The best 3D model was selected on the basis of QMEAN4 global scores, Z-score and Ramachandran plot. The QMEAN4 global score was obtained by using QMEAN server while the Ramachandran plot was obtained using PROCHECK and RAM-PAGE. Ramachandran plot was very much useful in evaluating backbone conformation as well as in checking non-GLY residues at disallowed regions. The verify 3D and ProSA web tool was used to determine Z-scores while ERRAT was used to predict overall quality for model and was assured using Z-scores [34]. The validated structure was subjected to energy minimization using KoBaMIN, a knowledge-based web server for protein structure refinement [35].

#### Potential B cell epitope identification

##### Continuous (linear) epitope identification

The B cell epitope is the portion of the antigen which interacts with B-lymphocytes. As a result, the B-lymphocyte is differentiated into antibody-secreting plasma cell and memory cells. In other terms, the objective of the B cell epitope is to ultimately stimulate the B cell to synthesize the antibody specific for it (primary humoral response) or to convert the naive B cell into a memory B cell and make it ready to produce specific antibody in later encounters [36].

B cell epitope is characterized by both being hydrophilic, accessible and in a flexible region of an immunogen [37]. Thus, the classical propensity scale methods such as Kolaskar and Tongaonkar antigenicity scale [38], Emini surface accessibility prediction [39], Parker hydrophilicity prediction [40], Karplus and Schulz flexibility prediction [41], and

Bepipred linear epitope prediction which uses a combinatorial algorithm comprising both hidden markov model and propensity scale methods antigenic propensity and thus performs significantly better than any of the other methods and analysis were done computationally from the IEDB (<http://www.iedb.org/>) analysis resource [42]. In several experimental studies, it was found that the antigenic parts of a protein belong to the beta turn regions [43]. Therefore, the Chou and Fasman beta turn prediction tool [44] was used. The results from all these sites were cross-referenced, and apparently common findings were taken as the continuous (linear) B cell epitopes.

##### Discontinuous (Conformational) epitope identification

Conformational epitope for B cell was predicted by ElliPro (<http://tools.immuneepitope.org/tools/ElliPro>) in IEDB analysis resources which implements three algorithm performing approximation of the protein shape as an ellipsoid [45], calculation of residue protrusion index [46] and clustering of neighboring residues based on their PI values.

#### Potential T-cell epitope identification

##### CD8+ T-cell epitopes identification

Linear peptides as T-cell epitopes for ZIKV envelope glycoprotein were identified by using NetCTL 1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>) at a threshold of 0.75 to have sensitivity and specificity of 0.80 and 0.970 respectively [47]. The server expands the MHC class I binding prediction to 12 MHC supertypes and integrates prediction of peptide MHC class I binding, proteasomal C terminal cleavage and TAP transport efficiency using artificial neural network (ANN). Combined algorithm of MHC-1 binding, transporter of antigenic peptides (TAP) transport efficiency and proteasomal cleavage efficiency were used to determine the overall scores. On the basis of these overall score, 4 best epitopes were selected for further evaluation.

For the prediction of peptides binding to MHC-1, we used a different prediction method at Immune Epitope Database (IEDB) and predict IC50 values for peptides binding to specific MHC molecules [48]. The stabilized matrix base method (SMM) was used to calculate IC50 values of peptide binding to MHC-1 molecules. Prior to the prediction, peptide length was set to 9 amino acids. The alleles having binding affinity IC50 less than 200 nm were chosen for further analysis.

The best four epitopes were then fetched to a web based tool T-cell epitope processing prediction at IEDB [49] which utilizes proteasomal cleavage /TAP transport/ MHC-1 combined prediction scores using SMM for each peptide's intrinsic potential of being a T cell epitope. MHC-NP; prediction of peptides naturally processed by the MHC, another tool at IEDB server was used to assess the probability

that a given peptide is naturally processed and binds to a given MHC molecule [50].

The conservancy and relative ability of peptide epitope – MHC I complex to elicit an immune response was predicted by analysis tools and T cell class I pMHC immunogenicity predictor at IEDB which uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a class I peptide MHC (pMHC) complex [51].

#### CD4+ T-cell epitopes identification

CD4+ T-cell responses against ZIKV envelope glycoprotein were done by using Peptide binding to MHC class II molecules program under MHC II binding prediction tool in IEDB analysis resource. For this prediction we chose seven abundant HLA class II alleles DRB1\*01:01, DRB1\*04:01, DRB1\*07:01, DRB1\*11:01 and DRB1\*15:01 from the selection panel [52]. The predicted T-cell epitopes having IC50 value less than 50 were considered as potential T-cell epitopes and their corresponding scores for respective alleles were determined by PREDIVAC (<http://predivac.biosci.uq.edu.au/cgi-bin/binding.py>) [53].

#### IFN induction capacity prediction and molecular simulation study

Both ZIKV envelope glycoprotein and predicted Bcell linear epitope (123-141) were fed to computational tool IFNepitope (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>) to scan INF-gamma induction. This server generates all the possible overlapping peptides at window length of 20 and predicts IFN epitope in these overlapping peptides to rank these peptides/epitopes based on their SVM score. During scanning motif and SVM hybrid algorithm was applied for IFN-gamma versus Non IFN-gamma epitopes prediction. To perform epitopes-receptor docking, 3D structure of the IFN-gamma receptor alpha chain (PDB id: 1fyh) was retrieved from RCSB Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)). The protein-protein docking of IFN-gamma receptor and ZIKV envelope glycoprotein was performed by PatchDock server [54] under

default settings. The outputs of the PatchDock were fed to FireDock for refinement in molecular docking [55].

## Results

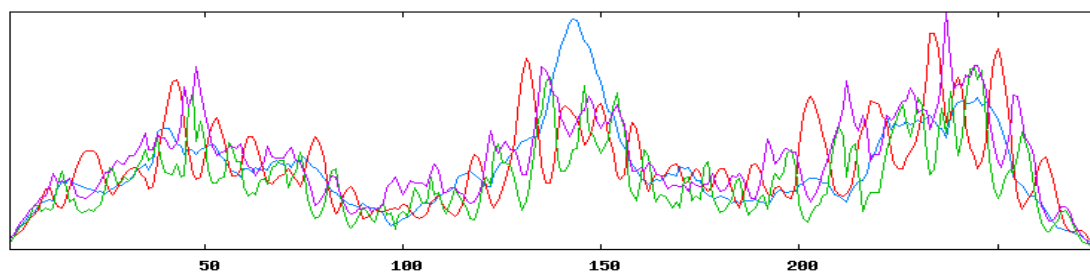
### Retrieval and prediction of antigenic Protein

The current study was originated to perform structure based sequence analysis of envelope glycoprotein isolated from Zika virus. A total of 12 polyproteins were retrieved from UniProtKB in FASTA format and then subjected to VaxiJen to predict most immunogenic protein. UniProtKB primary accession number **Q91KX7** was predicted as the most antigenic one with a total score of 0.6485 at a threshold of 0.4. Finding of this prediction coincides with a previous finding where envelope glycoprotein was presented as an immunogenic protein [56].

### Primary and secondary structure analysis:

Primary structure analysis revealed that the envelope protein (276 aa) had a molecular weight of 30.3 kD and theoretical isoelectric point (PI) 6.23. The total number of positively charged residues (Arg+Lys) and negatively charged residues (Asp+Glu) were 27 and 32 respectively. This finding along with isoelectric point indicates the protein as a negatively charged one. The estimated half-life (*in vitro*) was 1.1 hours in mammalian reticulocytes. The instability index (II) was found 25.21, thereby categorizes the protein as a stable. The aliphatic index appeared as 77.72 and the N-terminus of the sequence showed the presence of F (Phe). The negative grand average of hydropathicity (GRAVY) of -0.298 denoted that the protein was hydrophilic. The amino acids, Leu (L), Gly (G), Thr (T), Ala (A) and Val (V) were found in high proportion in the protein.

The secondary structure at a threshold of 8.0 from SOPMA was disclosed with the presence of 32 (11.59%)  $\alpha$ -helix, 100 (36.23%)  $\beta$ -sheet, 99 (35.87%) extended strand and 121 (43.84%) coils (Fig 1) while from GOR IV had 43 (15.58%)  $\alpha$ -helix, 79 (28.62%) extended strand and 154 (55.80%) coils.



**Figure 1: Secondary structure plot of Zika virus envelope glycoprotein (UniprotKB id: Q91KX7). Here, helix, extended strands and beta turns are indicated by blue, red and green respectively.**

From primary structure the GRAVY of the protein was found to be negative which predicts that, the most of the residues to

be present in the surface. In addition, the aliphatic and the instability index of the protein described it as an aliphatic and stable protein respectively. The target protein also contained

8.3% of the threonine (T) residue, which prefers to lie in the beta sheet and the secondary structure analysis by SOPMA showed that, 36.23% region of the target protein remains as beta sheet. In several experiments, it was shown that the antigenic part of the protein is more likely to belong to the beta sheet region [57]. Primary and secondary structure analyses indicate an important feature of the target protein as an antigenic one.

### 3D structure modeling (homology modeling) and validation

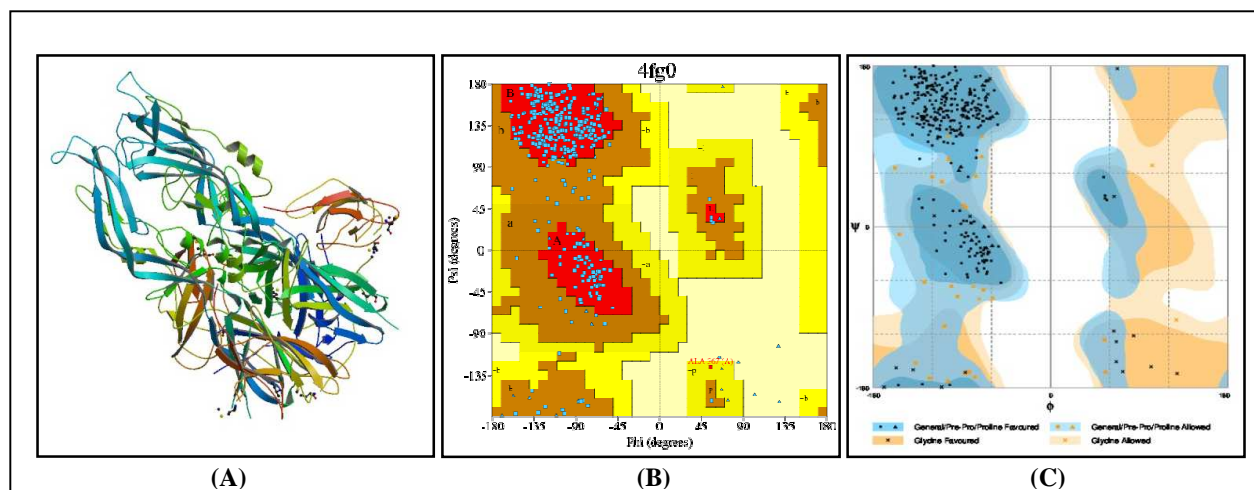
From blastp it was found that template 1 had max & total score of 310, E-value  $7e-103$  and identity 50% while template 2 had max & total score of 295, E-value  $2e-97$  and identity 48%. The characteristic features for 3D structure modeling of desired protein by HHpred and SWISS-MODELLER are summarized in **Table 1**.

**Table 1: Characteristic features of template 1(3P54\_A) and template 2 (4FG0\_A) in HHpred and SWISS-MODELLER**

Templa te	HHpred					SWISS-MODEL					
	Coverag e	Zscor e	Seq id	Confiden ce	Prob	E value	P value	Score	Seq Identity	GMQ E	QMEA N4
<b>3P54_A</b>	0.981	44.29 2	0.513	High	100.0	2E- 130	5E-135	937.2	51.85%	0.79	-2.42
<b>4FG0_ A</b>	0.931	44.30 2	0.506	High	100.0	5E- 129	1E-133	929.5	49.82%	0.78	-2.27

Post modeling analysis of the qualities of two templates revealed a reliable and favorable structure of both template 1 and 2 due to similar QMEAN4 score (template 1: 0.587 and template 2: 0.582) but had difference in Ramachandran plots. Template 1 (3P54\_A) at resolution 2.10 Å and R-factor of 0.181 perceived Ramachandran plot (Psi-Phi) pairs had 305 (87.6%) residues in most favored region [A, B,L], 41 (11.8%) core residues in additional allowed regions [a,b,l,p], 1 (0.3%) residue in generously allowed regions [~a,~b,~l,~p] and 1 (0.3%) residue in disallowed regions [XX]. While template 2 (4FG0\_A) at resolution 3.90 Å and R-factor of 0.226 the perceived Ramachandran plot (Psi-Phi) pairs had 289 (87.0%)

residues in most favored region [A, B,L], 42 (12.7%) core residues in additional allowed regions [a,b,l,p], 1 (0.3%) residue in generously allowed regions [~a,~b,~l,~p] and no residues in disallowed regions [XX]. These values indicated a good quality model of template 2 over template 1. From RAMPAGE it was found that, template 2 had 92.1% and 7.9% residues in favored and allowed region respectively while no residues were found in outlier region.



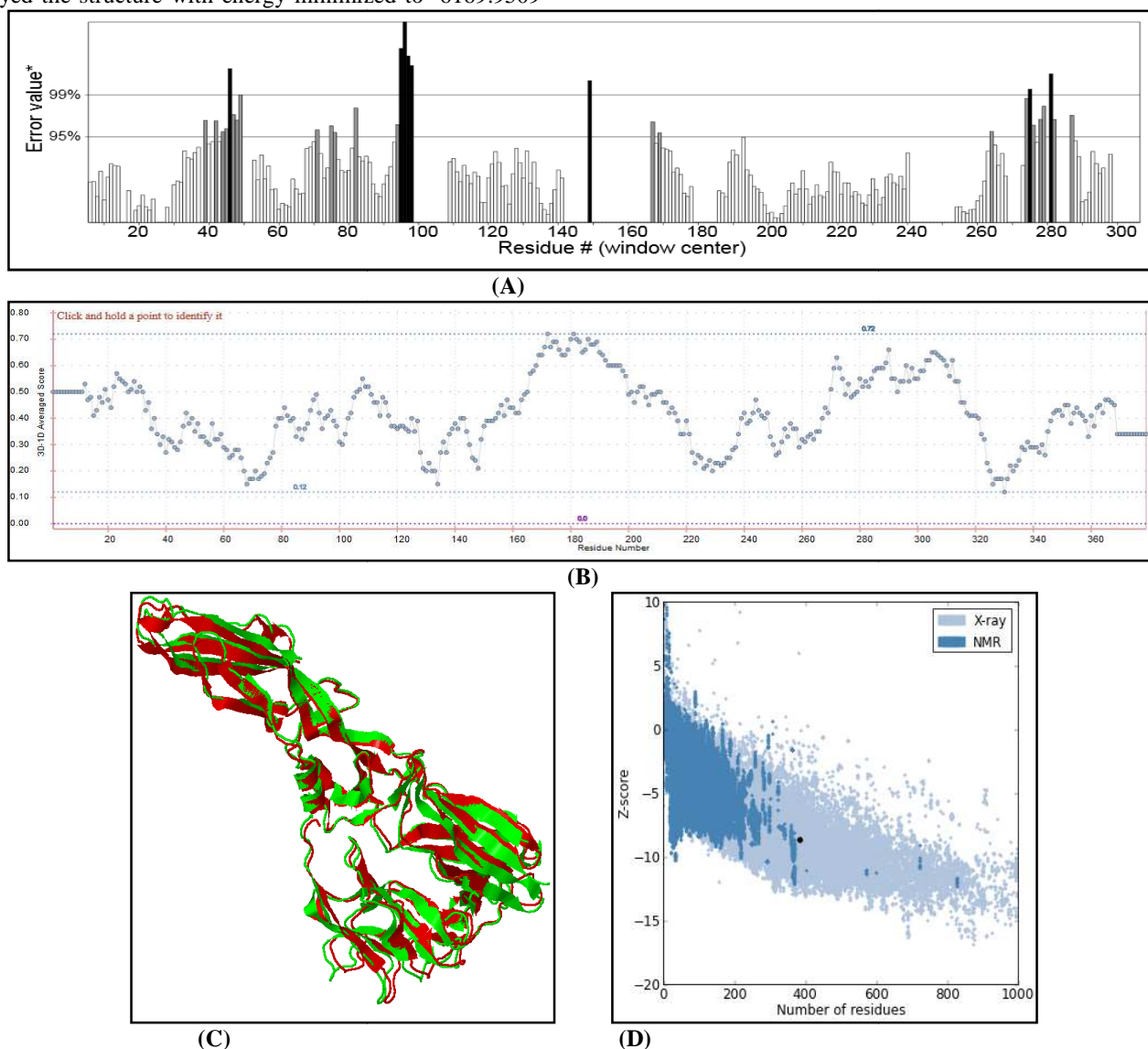
**Figure 2: (A) the 3D structure of Zika virus envelope protein. The image was retrieved from RCBS protein data bank (B) The Ramachandran plot for Model 2 (4FG0\_A) showing residues predicted by PROCHECK and (C) RAMPAGE showing the residues in favored, allowed and outlier regions**

Verify3D and ERRAT were also used to further assess the quality of the model i.e. template 2. Verify3D analyzes the compatibility of the model against its own amino acid sequence and revealed that 95.65% of residues had an average

3D/1D score  $\geq 0.2$  while ProSA gave the Z-score of -8.62 for the model. Both Verify3D and ProSA gave good scores for overall model quality of template 2. However the ERRAT validation of the model indicated the regions where the calculated errors were higher than expected with an overall

quality factor of 86.48. Subsequently, the selected model displayed the structure with energy minimized to -6169.9509

KB energy (kcal/mol).



**Figure 3: (A) the graphical representation of ERRAT result. On the error axis two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error values; (B) the Varify3D result showing 95.65% residues had an average 3D/1D score  $\geq 0.2$ ; (C) energy minimized structure and (D) Screenshot of ProSA web Z-score of template 2 (4FG0)**

#### Potential B cell epitope identification

##### Continuous (Linear) epitope identification

Characteristic features of the B cell epitope include hydrophilicity, surface accessibility, and beta-turn prediction. Thus, several prediction methods were used to determine the epitope considering all these criteria.

Kolaskar and Tongaonkar antigenicity prediction method functions on the basis of physiochemical properties of amino acids and abundances in experimentally known epitopes. The

antigenic propensity for the protein was 1.00; all the values greater than 1.00 are potential antigenic determinants. Sixteen epitopes were found to satisfy the threshold value set prior to the analysis, and they have the potential to evoke the B cell response. The peptide from RQTVVVLGS from 128 to 136 was found to have the highest antigenic propensity score (1.12). The results are summarized in **Table 2** and **Figure 4**. Kolaskar and Tongaonkar antigenicity scale predicted four different peptides (Each seven peptide long) sequences in position 131-137 with maximum antigenicity score 1.186 and other three in positions between 129-135, 128-134 and 130-136 had second highest antigenicity score of 1.171. BcePred

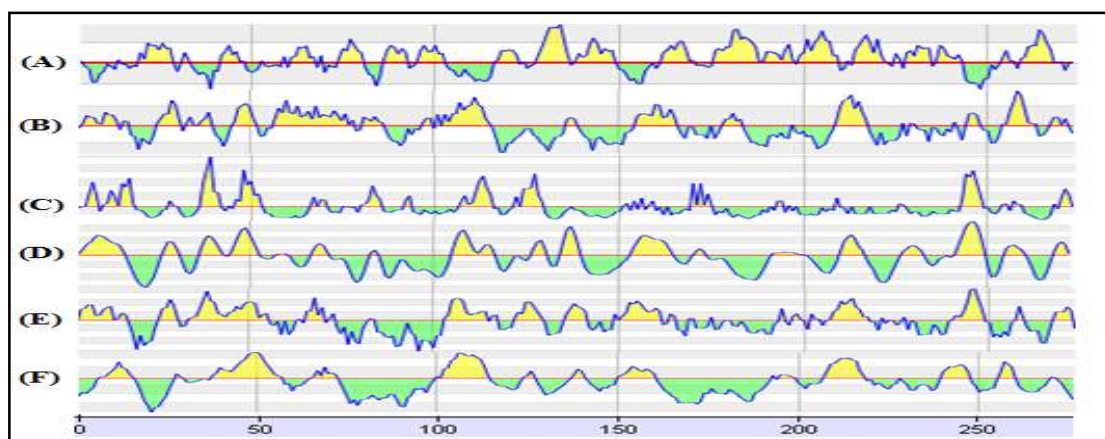
Prediction Server predicts two possible antigenic peptide sequences RQTVVVLGSQ and VSYSLCT in position between 128-137 and 179-185 respectively at threshold 2.5.

When all the scales from IEDB server for B cell linear epitope prediction were considered, sequence spanning in between 123-141 showed highest antigenic propensity, the combined result illustrated in **Figure 5**. Analysis of the predicted peptide sequence with VaxiJen showed the sequence as probable

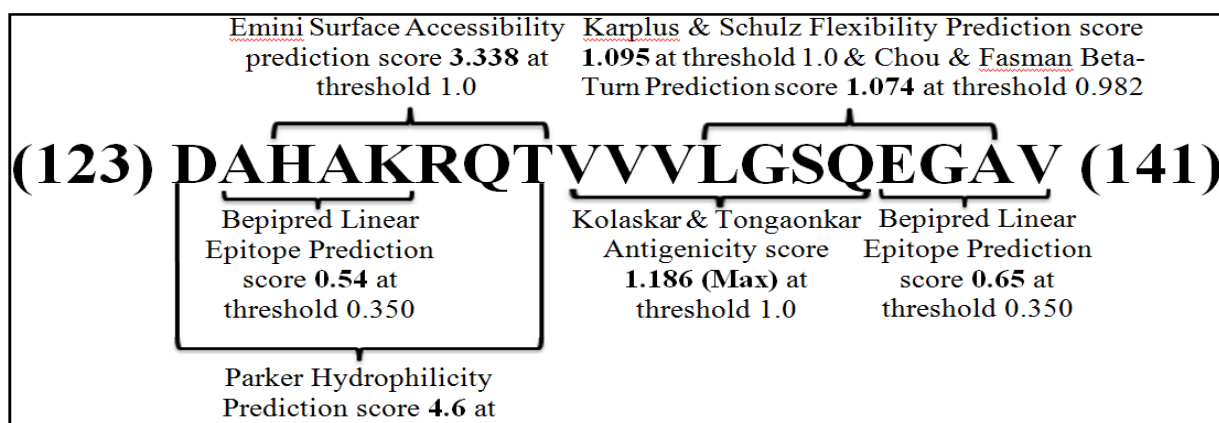
antigen with a score of 0.73 at threshold of 0.4. Further, Bepipred analysis (B cell epitope prediction) revealed two continuous predicted epitope with 4 amino acid residues each in sequence positions between 124-127 and 138-141 at a threshold of 0.350.

**Table 2: Predicted epitopes of ZIKV envelope glycoprotein via Kolaskar and Tongaonkar antigenicity scale (No. 1-16) and Bepipred continuous epitope prediction (No. 17-18)**

No.	Start Position	End Position	Peptide	Peptide Length	Antigenic Propensity Score
1	20	28	RIMLSVHGS	9	1.05
2	42	48	KVEVTPN	7	1.04
3	61	67	SLGLDCE	7	1.07
4	74	80	FSDLYYL	7	1.11
5	86	92	HWLVHKE	7	1.07
6	95	102	HDIPLWH	8	1.06
7	116	122	EALVEFK	7	1.06
8	128	136	RQTVVVLGS	9	1.12
9	138	149	EGAVHTALAGAL	12	1.06
10	161	170	SSGHLKCRK	10	1.06
11	176	197	LEGVSYSLCTAAFTFTKVP AET	22	1.07
12	199	209	HGTVTVEVQYA	11	1.09
13	214	222	PCKVPAQMA	9	1.09
14	228	234	LTPVGRL	7	1.09
15	236	243	TANPVITE	8	1.01
16	259	270	GDSYIVIGVGDK	12	1.04
17	124	127	AHAK	4	0.54
18	138	141	EGAV	4	0.65



**Figure 4: Epitope prediction of ZIKV envelope glycoprotein using (A) Kolaskar andTongaonkar antigenicity prediction, (B) Chou & Fasman beta turn prediction, (C) Emini surface accessibility prediction, (D) Karplus and Schulz flexibility prediction, (E) Parker hydrophilicity prediction and (F) Bepipred linear epitope prediction**



**Figure 5: Combined B cell linear epitope prediction for ZIKV envelope glycoprotein using IEDB server showing all the scores for different scales and correspondent threshold value**

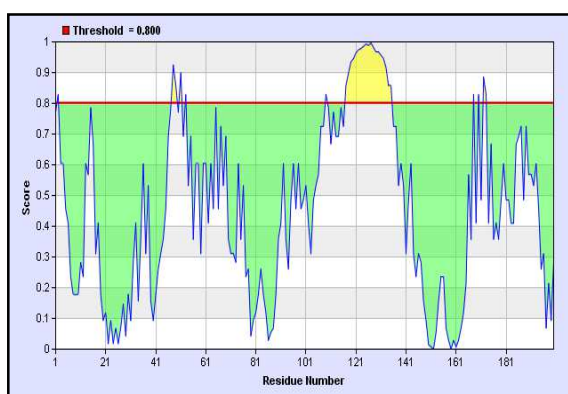
**Discontinuous (Conformational) epitope identification**

3D ZIKV envelope glycoprotein was feed to ElliPro tool to analyze conformational epitopes for B cell. ElliPro predicted four distinct epitopes with residual specification and corresponding score which have been summarized in **Table 3**

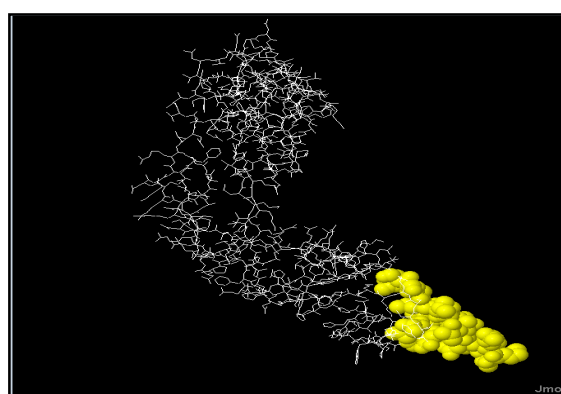
and **Figure 6**. The result revealed that, without any intervention amino acid residues for predicted linear epitope (123-141) overlap 78.9% of the amino acid residues for predicted conformational epitope with a highest score of 0.925 at threshold of 0.80.

**Table 3: Predicted discontinuous epitopes of ZIKV envelope glycoprotein by ElliPro tools**

No.	Residues	Number of Residues	Score
1	_:A117, _:L118, _:V119, _:E120, _:F121, _:K122, _:D123, _:A124, _:H125, _:A126, _:K127, _:R128, _:Q129, _:T130, _:V131, _:V132, _:V133, _:L134, _:G135, _:S136, _:Q137	21	0.925
2	_:V223, _:D224, _:M225, _:Q226, _:T227, _:L228, _:T229, _:P230, _:P257, _:F258, _:G259, _:D260, _:S261, _:Y262, _:T273, _:H274, _:H275, _:W276	18	0.854
3	_:P47, _:N48, _:S49, _:P50, _:R51, _:D65, _:R168, _:K170	8	0.836
4	_:V267, _:G268, _:D269, _:K270, _:K271	5	0.821



(A)



(B)

**Figure 6: B cell discontinuous epitope predicted by ElliPro tool for ZIKV envelope glycoprotein. (A) X axis represents the residues number and Y axis represents the corresponding score at a threshold of 0.80. Yellow colors in the plot represent the polypeptide region having a score above the threshold and are likely to be potential B cell epitopes while green color reflects the polypeptide regions that could not satisfy the threshold margin; (B) Jmol visualization of the first of the four predicted epitopes. The epitope residues are in yellow while the antibody chains are white in color**



**Potential T-cell epitope identification****CD8+ T-cell epitope identification**

The NetCTL prediction tool predicted 72 different epitopes from ZIKV envelope glycoprotein sequence according to all MHC (A1-B62) supertypes. But only 4 most potential epitopes were chosen based on their high combinatorial scores.

To define the binding of peptides to MHC 1, SMM based MHC 1 binding prediction tool retrieved 196 possible MHC-1 allele interactions with four best epitopes for T-cells. Those MHC-1 alleles who had higher binding affinity for each epitope at IC50 < 200 were selected for further analysis.

Proteasome involves in cleaving the peptide bonds, converting the proteins into peptide. For that transporter of antigenic peptides (TAP) transfer the peptides into endoplasmic reticulum. After proteasomal cleavage the peptide molecules associated with MHC-1 and transported to cell membrane where they are presented to T helper cell. For T-cell epitope processing prediction, the total score of each epitope – HLA

interaction was considered and higher the score meant higher the processing efficiency.

Among the 4 best CD8+ T cell epitopes, a 9 mer epitope MMLELDPPF was found to interact with most MHC-1 alleles including HLA-A\*02:01; HLA-A\*02:06; HLA-A\*23:01; HLA-A\*29:02; HLA-A\*32:01; HLA-B\*15:01; HLA-B\*15:02; HLA-B\*35:01; HLA-B\*53:01; HLA-C\*03:03 and HLA-C\*14:02. Then epitope MMLELDPPF was subjected to MHC-NP prediction tool and found that, HLA-B\*53:01, HLA-A\*02:01 and HLA-B\*35:01 had highest probable score of 0.7516, 0.0442, and 0.0300 respectively. High probable score meant more likely it is naturally be processed.

The epitope conservancy and I pMHC immunogenicity prediction analysis of the best epitopes showed that, all the three epitopes except 1 had epitope conservancy value of 52.17% but epitope MMLELDPPF had highest I pMHC immunogenicity score of 0.09139. The NetCTL combined scores, predicted total scores with IC50 values, epitope conservancy and I pMHC immunogenicity prediction scores are summarized in **Table 4**.

**Table 4: Predicted total scores of all the processes for most potential 4 CD8+ T-cell epitopes with interacting MHC-1 alleles**

Epitope	NetCTL combined score	Interacting MHC-1 allele with an affinity of IC50<200 (Total score fo proteasome, TAP , MHC processing and MHC-1 binding)	Epitope conservancy	I pMHC immunogenicity prediction score
<b>GLDFSDLYY</b>	3.0511; A1 0.9143; A3	HLA-A*01:01; 110.8 (0.98) HLA-A*29:02; 35.54 (0.84) HLA-C*05:01; 4.25 (-0.63) HLA-B*15:02; 84.77 (-1.34) HLA-C*14:02; 57.76 (-1.6) HLA-C*12:03; 10.67 (-1.79)	52.17%	-0.02674
<b>MMLELDPPF</b>	1.4149; B62 0.8398; B58	HLA-B*15:02; 7.76 (1.17) HLA-B*35:01; 8.57 (1.13) HLA-B*15:01; 15.96 (0.86) HLA-A*32:01; 20.74 (0.75) HLA-A*02:06; 22.53 (0.71) HLA-B*53:01; 66.79 (0.24) HLA-A*02:01; 78.68 (0.17) HLA-A*29:02; 94.21 (0.09) HLA-C*03:03; 160.11 (-0.14) HLA-A*23:01; 167.23 (-0.16) HLA-C*14:02; 287.82 (-0.39)	52.17%	0.09139
<b>SYSLCTAAF</b>	1.8055; A24 0.7502; B62	HLA-C*14:02; 7.76 (1.5) HLA-A*24:02; 85.83 (1.1) HLA-A*23:01; 104.58 (0.92) HLA-C*07:02; 92.29 (0.09) HLA-C*12:03; 93.16 (-0.69) HLA-C*03:03; 59.1 (-1.39)	52.17%	-0.02494
<b>KLRLEGVSY</b>	1.4556; B62 1.1810; A3	HLA-B*15:01; 62.66 (1.05) HLA-A*30:01; 137.39 (0.45) HLA-C*14:02; 168.9 (-0.81) HLA-C*12:03; 28.72 (-0.99)	4.35%	0.07322

### CD4+ T-cell epitope identification

MHC class II binding prediction tool retrieved a total of 28 epitopes for HLA-DRB1\*01:01; 5 epitopes for HLA-DRB1\*04:01; 5 epitopes for HLA-DRB1\*07:01, 12 epitopes for HLA-DRB1\*11:01 and 5 epitopes for HLA-DRB1\*15:01 within IC50 less than 63. After analysis an epitope YRIMLSVHG was found to interact with all of the above HLA-DRB1 and can act as potential CD4+ T-cell epitope for ZIKV envelope glycoprotein and elicit an immune response.

### IFN epitope prediction and docking analysis

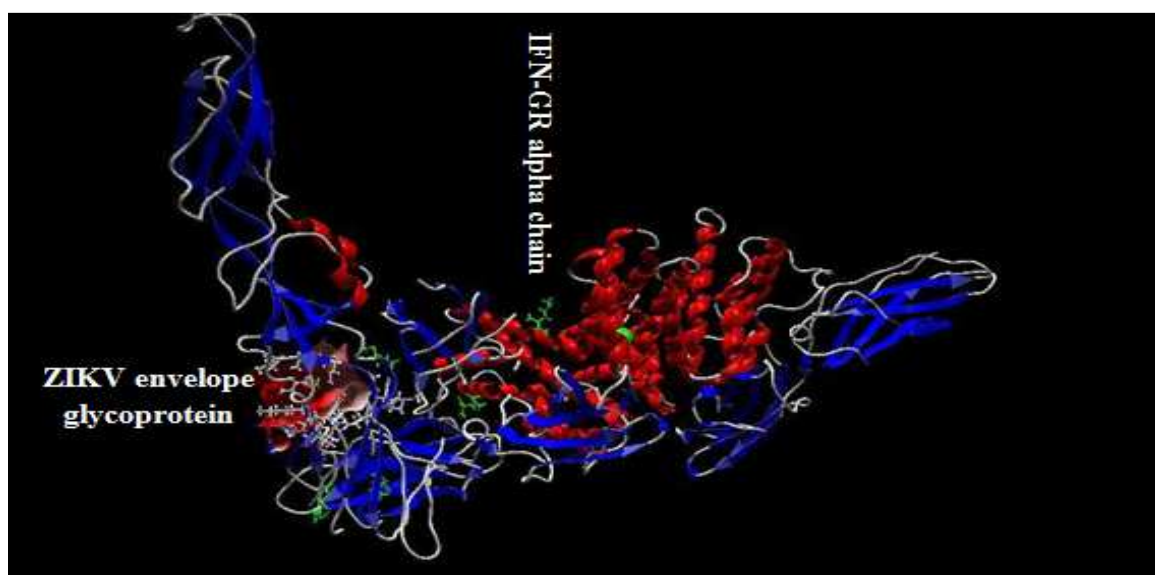
The result from IFNepitope tool for ZIKV envelope glycoprotein and predicted B cell linear epitope revealed that, both of them had the probability to release of IFN-gamma with the positive score. Within entire glycoprotein position in between 124-143 (AHAKRQTVVVLGSQEGAVHT) showed the maximum SVM score of 1.81 whereas the predicted B cell linear epitope had SVM score of 1.29. The rigid and symmetric docking of ZIKV envelope glycoprotein with respect to the IFN-gamma was evolved by PatchDock algorithm and post docking refinement was done by FireDock tool. FireDock server addresses the refinement problem of protein-protein docking solutions and simultaneously targets the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithms. The docking and post docking refinement results have been summarized in **Table 5**. The best docking pose ranked on global energy showed an energetically favorable interaction between ZIKV envelope glycoprotein and IFN-gamma receptor alpha chain (**Figuer 7**).

**Table 5: Docking and post docking analysis results for ZIKV envelope glycoprotein-IFN gamma receptor alpha chain interaction**

Rank	Score	Area	Transformation	GE	Vdw Attractive	Vdw Repulsive	ACE	HB
1	12248	1748.20	1.66 0.33 -0.16 21.84 96.41 16.81	-44.97	-25.23	10.44	6.13	-3.41
2	12164	2090.50	1.65 -0.98 -1.55 36.43 56.36 1.97	-44.79	-42.41	33.77	-3.17	-2.69
3	11578	1465.30	-0.65 -0.26 0.40 2.42 14.18 -75.09	-44.55	-31.27	17.13	-5.56	-2.14
4	14168	1836.00	2.74 -0.33 -1.21 54.04 44.81 -52.40	-43.30	-21.70	7.31	-5.45	-0.78
5	12246	1764.60	-1.31 -0.30 -1.49 - 38.60 5.11 -48.42	-42.58	-19.55	4.91	-13.12	-0.43
6	13382	1829.80	3.10 0.38 -1.43 82.55 72.33 68.62	-38.68	-33.07	18.87	2.00	-4.94
7	11404	1424.70	-0.96 -0.39 0.37 10.85 38.56 13.54	-38.27	-27.27	10.94	0.86	-2.11
8	12688	1658.00	-2.16 -0.89 1.36 76.92 112.84 -3.73	-38.02	-25.46	22.65	-11.98	-1.64
9	12068	1914.00	0.99 -0.66 -2.13 37.38 43.98 -11.11	-37.65	-27.81	9.86	1.15	-5.54
10	12458	1979.60	1.55 -0.67 0.04 - 47.63 104.98 -79.05	-36.97	-24.30	13.94	-7.77	-4.39

The table is sorted by the global energy (GE) value while GE is the binding energy of a solution. Transformation refers to 3D transformation with 3 rotational angles and 3 translational parameters and applied on the ligand molecule. Here Score means geometric shape complementary score; Area is

approximate interface area of the complex; Vdw is Van der Waals; ACE means the contribution of the atomic contact energy (ACE) to the global binding energy and HB is the contribution of hydrogen bonds to global binding energy.



**Figure 7: Best docking pose of ZIKV envelope glycoprotein with IFN-gamma receptor alpha chain. After post docking refinement by FireDock the best interaction was visualized by Molegro Molecula Viewer (v.2.5)**

## Discussion

The present study is a first attempt which aimed to screen new and highly potential immunogenic epitopes for B cell, T cell and IFN gamma as vaccine candidate for ZIKV envelope glycoprotein. Both cell mediated and humoral immunities are facilitated by T cell and B cell epitopes and determination of these epitopes are at the core of vaccine development. IFN gamma induced innate immune response is another important factor for vaccine design. To design an effective peptide antigen it is recommended to keep the hydrophobic amino acid content below 50% and peptide sequences within 8-22 amino acids. The continuous epitope 123 DAHAKRQTVVVLGSQEGAV 141 for B cell was 19 amino acids residue long and had 42% hydrophobic amino acids while the discontinuous epitope 117 ALVEFKDAHAKRQTVVVLGSQ 137 was 21 amino acids residue long and had 43% hydrophobic amino acids. The interesting finding is most of the regions for continuous epitope overlaps with the region of discontinuous epitope which discloses the residual similarities and ultimately suggest cross reactivity.

Design and development of vaccine against T cell epitope is much more promising due to the evoke of long lasting immune response and antigenic drift where antigen can easily escape the antibody memory response [58]. Both the CD8+ and CD4+ T cells have vital role in antiviral immune response as well as antigen mediated clonal expansion of B cell. This study showed a range of potential epitopes for MHC class I and MHC class II molecules. Among the 4 most potential CD8+ T cell epitopes, one epitope MMLELDPPF had an epitope conservancy of 52.17% along with highest

immunogenicity prediction score of 0.9139. The high level of epitope conservancy much more important because ZIKV has a higher tendency towards mutation due to the lack of proofreading activity of RNA polymerase [59] and considering a portion present in all the strains of virus is also recommended. Moreover the epitope MMLELDPPF showed interaction with HLA-A\*02 which is more prevalent in Africa and Asia where outbreak of ZIKV has been occurred. On the other hand it was found that, all the CD4+ T cell epitopes were interact with HLA-DRB1 within IC50 less than 63 which is predominant in southern African population.

Interferon Gamma (IFN-g), also called immune or type II interferon is a pleiotropic cytokine involved in the regulation of nearly all phases of immune and inflammatory responses, including the activation, growth and differentiation of T-cells, B-cells, macrophages, NK cells and other cell types such as endothelial cells and fibroblasts. It enhances MHC expression on antigen-presenting cells, and IFN- production is characteristic of Th1 differentiation [60]. IFN gamma production can be induced by ZIKV envelope glycoprotein and the sequences 124 AHAKRQTVVVLGSQEGAVHT 143 showed the maximum SVM score of 1.81 and obviously induce the antiviral defense.

The above findings are the result of analyzing the deposited data on various immune databases. The results suggest that, these epitopes may play a highly informative role in antidote production against Zika virus that can trigger an effective immune response *in vivo*. Along with *in silico* study, both *in vivo* and *in vitro* experiments are required to prove the effectiveness of mounting an immune response.

## Conflict of interest

The authors declares that they have no conflict of interest

## Supplementary data

Table 1: Supplementary data showing the predicted CD4+ T-cell epitopes of ZIKV envelope glycoprotein having IC50&lt;63

Allele	Start	End	Peptide	SMM align. Core	IC50	PREDIVAC Score
HLA-DRB1*01:01	16	30	NLEYRIMLSVHGSQH	YRIMLSVHG	14	94.08
HLA-DRB1*01:01	17	31	LEYRIMLSVHGSQHS	YRIMLSVHG	14	94.08
HLA-DRB1*01:01	187	201	AFTFTKVP AETLHGT	FTKVP AETL	14	84.12
HLA-DRB1*01:01	13	27	QPENLEYRIMLSVHG	NLEYRIMLS	15	-
HLA-DRB1*01:01	14	28	PENLEYRIMLSVHGS	YRIMLSVHG	15	94.08
HLA-DRB1*01:01	15	29	ENLEYRIMLSVHGSQ	YRIMLSVHG	15	94.08
HLA-DRB1*01:01	185	199	TAAFTFTKVP AETLH	FTKVP AETL	15	84.12
HLA-DRB1*01:01	186	200	AAFTFTKVP AETLHGT	FTKVP AETL	15	84.12
HLA-DRB1*01:01	188	202	FTFTKVP AETLHGTV	FTKVP AETL	15	84.12
HLA-DRB1*01:01	184	198	CTAAFTFTKVP AETL	CTAAFTFTK	15	-
HLA-DRB1*11:01	13	27	QPENLEYRIMLSVHG	QPENLEYRI	20	-
HLA-DRB1*11:01	14	28	PENLEYRIMLSVHGS	YRIMLSVHG	20	80.35
HLA-DRB1*11:01	15	29	ENLEYRIMLSVHGSQ	YRIMLSVHG	20	80.35
HLA-DRB1*11:01	16	30	NLEYRIMLSVHGSQH	YRIMLSVHG	20	80.35
HLA-DRB1*11:01	17	31	LEYRIMLSVHGSQHS	YRIMLSVHG	20	80.35
HLA-DRB1*07:01	184	198	CTAAFTFTKVP AETL	CTAAFTFTK	20	-
HLA-DRB1*07:01	185	199	TAAFTFTKVP AETLH	FTKVP AETL	20	80.80
HLA-DRB1*07:01	187	201	AFTFTKVP AETLHGT	FTKVP AETL	21	80.80
HLA-DRB1*07:01	186	200	AAFTFTKVP AETLHGT	FTKVP AETL	22	80.80
HLA-DRB1*07:01	188	202	FTFTKVP AETLHGTV	FTKVP AETL	22	80.80
HLA-DRB1*01:01	178	192	GVSYSLCTAAFTFTK	YSLCTAAFT	32	83.93
HLA-DRB1*01:01	177	191	EGVSYSLCTAAFTFT	YSLCTAAFT	33	83.93
HLA-DRB1*01:01	176	190	LEGVSYSLCTAAFTF	YSLCTAAFT	34	83.93
HLA-DRB1*01:01	175	189	RLEGVSYSLCTAAFT	YSYSLCTAAF	35	-
HLA-DRB1*01:01	179	193	VSYSLCTAAFTFTKV	YSLCTAAFT	36	83.93
HLA-DRB1*01:01	228	242	LTPVGRLITANPVIT	GRLITANPV	39	-
HLA-DRB1*15:01	17	31	LEYRIMLSVHGSQHS	YRIMLSVHG	40	81.01
HLA-DRB1*01:01	229	243	TPVGRLITANPVITE	LITANPVIT	40	81.64
HLA-DRB1*15:01	13	27	QPENLEYRIMLSVHG	ENLEYRIML	41	-
HLA-DRB1*15:01	14	28	PENLEYRIMLSVHGS	YRIMLSVHG	41	81.01
HLA-DRB1*15:01	15	29	ENLEYRIMLSVHGSQ	YRIMLSVHG	41	81.01
HLA-DRB1*15:01	16	30	NLEYRIMLSVHGSQH	YRIMLSVHG	41	81.01
HLA-DRB1*01:01	230	244	PVGRLITANPVITES	LITANPVIT	41	81.64
HLA-DRB1*01:01	18	32	EYRIMLSVHGSQHS	YRIMLSVHG	42	94.08
HLA-DRB1*01:01	19	33	YRIMLSVHGSQHSGM	YRIMLSVHG	43	94.08
HLA-DRB1*01:01	189	203	TFTKVP AETLHGTVT	FTKVP AETL	43	82.12
HLA-DRB1*01:01	190	204	FTKVP AETLHGTVTV	FTKVP AETL	44	82.12
HLA-DRB1*01:01	231	245	VGRLITANPVITEST	LITANPVIT	49	81.64
HLA-DRB1*01:01	138	152	EGAVHTALAGALEAE	VHTALAGAL	51	-
HLA-DRB1*01:01	137	151	QEGAVHTALAGALEA	VHTALAGAL	51	-
HLA-DRB1*01:01	139	153	GAVHTALAGALEAEM	VHTALAGAL	51	-
HLA-DRB1*01:01	136	150	SQEGAVHTALAGALE	VHTALAGAL	53	-
HLA-DRB1*01:01	232	246	GRLITANPVITESTE	LITANPVIT	53	81.64
HLA-DRB1*11:01	75	89	SDLYYLTMNNKHWLV	YLTMNNKHW	56	82.65
HLA-DRB1*11:01	76	90	DLYYLTMNNKHWLVH	YLTMNNKHW	57	82.65
HLA-DRB1*11:01	74	88	FSDLYYLTMNNKHWL	YLTMNNKHW	57	82.65
HLA-DRB1*04:01	15	29	ENLEYRIMLSVHGSQ	YRIMLSVHG	60	87.88
HLA-DRB1*04:01	16	30	NLEYRIMLSVHGSQH	YRIMLSVHG	60	87.88
HLA-DRB1*04:01	17	31	LEYRIMLSVHGSQHS	YRIMLSVHG	61	87.88
HLA-DRB1*11:01	73	87	DFSDLYYLTMNNKHW	YYLTMNNKH	81	73.46
HLA-DRB1*04:01	14	28	PENLEYRIMLSVHGS	YRIMLSVHG	62	87.88
HLA-DRB1*04:01	13	27	QPENLEYRIMLSVHG	LEYRIMLSV	63	-
HLA-DRB1*11:01	18	32	EYRIMLSVHGSQHS	YRIMLSVHG	63	87.88
HLA-DRB1*11:01	19	33	YRIMLSVHGSQHSGM	YRIMLSVHG	63	87.88
HLA-DRB1*11:01	77	91	LYYLTMNNKHWLVHK	YLTMNNKHW	63	73.46

## References

- Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg.* 1952; 46(5):509-20.
- Hayes EB. Zika virus outside Africa. *Emerg Infect Dis.* Vol 15, No. 9, September 2009.
- Haddow AJ, Williams MC, Woodall JP, Simpson DI, Goma LK. Twelve isolations of Zika virus from *Aedes (Stegomyia) africanus* (Theobald) taken in and above Uganda forest. *Bull World Health Organ.* 1964; 31:57-69.
- Marchette NJ, Garcia R, Rudnick A. Isolation of Zika virus from *Aedes aegypti* mosquitoes in Malaysia. *Am J Trop Med Hyg.* 1969; 18(3):411-5.
- Henderson BE, Hewitt LE, Lule M (1968) Serology of wild mammals. In: *Virus Research Institute Annual Report*, East African Printer, Nairobi, Kenya. Pp 48-51.
- Kirya BG, Okia NO. A yellow fever epizootic in Zika Forest, Uganda, during 1972: Part 2: Monkey serology. *Trans R Soc Trop Med Hyg.* 1977;71:300-303.
- McCrae AW, Kirya BG. Yellow fever and Zika virus epizootics and enzootics in Uganda. *Trans R Soc Trop Med Hyg.* 1982;76: 552-562.
- McCrae AW, Kirya BG and Tukei PM (1970) Summary of an apparent epizootic of Zika virus: Pattern of incidence from *Aedes africanus* collected from the Zika Forest, 1969-1970. In: *Virus Research Institute Annual Report*, East African Printer, Nairobi. pp 20-21.
- Foy BD, Kobylinski KC, Chilson Foy JL, Blitvich BJ, Travassos da Rosa A, Haddow AD, *et al.* Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg Infect Dis.* 2011; 17(5):880-2.
- Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, *et al.* Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl Trop Dis.* 2012; 6(2).
- Kwong JC, Druce JD, Leder K. Zika virus infection acquired during brief travel to Indonesia. *Am J Trop Med Hyg.* 2013; 89(3):516-7.
- Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, *et al.* Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med.* 2009;360: 2536-2543.
- Kuno G, Chang GJ. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch Virol.* 2007;152: 687-696.
- Henderson BR, Saeedi BJ, Campagnola G, Geiss BJ. "Analysis of RNA binding by the Dengue virus NS5 RNA capping enzyme". In Jeang, K.T. *PLoS ONE* 2011;6 (10).
- Services YSDoH. Zika virus. Information for clinicians and other health professionals. Federated States of Micronesia: Yap State Department of Health; 2007. p. 2
- Cerdeño-Tárraga A. M., Efstratiou A., Dover L. G., Holden M. T., Pallen M., Bentley S. D., Besra G. S., Churcher C., James K. D., De Zoysa A., Chillingworth T., Cronin A., Dowd L., Feltwell T., Hamlin N., Holroyd S., Jagels K., Moule S., Quail M. A., Rabinowitsch E., Rutherford K. M., Thomson N. R., Unwin L., Whitehead S., Barrell B. G., Parkhill J., The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129, *Nucleic Acids Research* 2003; 31(22):6516-23.
- Trent DW, Qureshi AA. Structural and nonstructural proteins of Saint Louis encephalitis virus. *Journal of Virology.* 1971; 7(3):379-88.
- Testa JS, Philip R. Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. *Future virol.* 2012; 7(11):1077-1088.
- Pulendran B., Ahmed R., Immunological mechanisms of vaccination, *Nature Immunology* 2011; 12(6):509-17.
- Barh D., Misra A. N., Kumar A., Azevedo V., A novel strategy of epitope design in *Neisseria gonorrhoeae*, *Bioinformation* 2010; 5(2): 77-82.
- Whitmire J. K., Tan J. T., Whitton J. L., Interferon- $\gamma$  acts directly on CD8+ T cells to increase their abundance during virus infection, *The Journal of Experimental Medicine* 2005; 201(7):1053-1059.
- Zajac AJ and Harrington LJ: *Encyclopedia of Virology.* Elsevier Ltd, Birmingham, Edition 3, VOL.3, 2008:70-77
- Fensterl V., Sen G. C., Interferons and viral infections, *BioFactors* 2009; 35(1):14-20.
- Gaunt M.W., Sall A.A., de Lamballerie X., Falconar A.K.I., Dzhivianian T.I., Gould E.A. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J. Gen. Virol.* 2001;82:1867-1876.
- Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* 2007; 8(4).
- Christian Cole, Jonathan D. Barber and Geoffrey J. Barton. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 2008 Jul 1; 36(Web Server issue):W197-201.
- GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, *Methods in Enzymology*, R.F. Doolittle Ed., vol 266, 540-553, (1996).
- Geourjon C, Deleage G: SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* 1995, 11:681-684.
- Review Bioinformatics in microbial biotechnology – a mini review Arvind K Bansal.
- Benkert P., Biasini M., Schwede T., Toward the estimation of the absolute quality of individual protein structure models, *Bioinformatics* 2011; 27(3):343-50.
- Wiederstein, Sippl, ProSA-web: interactive web service for the recognition of errors in threedimensional structures of proteins, *Nucleic Acids Research* 2007; 35: W407-W410.
- Laskowski R. A., MacArthur M. W., Moss D. S., Thornton J. M., PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 1993; 26: 283-291.
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC: Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins Struct Funct Genet* 2003, 50:437-450.
- Arun G Ingale and Susumu Goto. Prediction of CTL epitope, in silico modeling and functional analysis of

- cytolethal distending toxin (CDT) protein of *Campylobacter jejuni*. BMC Res Notes. 2014; 7: 92.
35. João P. G. L. M. Rodrigues, Michael Levitt and Gaurav Chopra. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. Nucleic Acids Res. Jul 2012; 40(Web Server issue): W323–W328.
  36. Deepak N, Singh K, Siddiqui Z, Nayak B, *et al.* Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. J Immunol. 2002; 168(5):2371–82.
  37. Fieser TM, John A, Tainer H, *et al.* Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein  $\alpha$ -helix. Proc Natl Acad Sci. 1987; 84(23):8568–72.
  38. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990; 276(1–2):172–4.
  39. Emini EA, Hughes JV, Perlow DS, *et al.* Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985; 55(3):836–9.
  40. 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.
  41. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985; 72:212–3.
  42. Andersen PH, Nielsen M and Lund O. Prediction of residues in discontinuous B cell epitopes using protein 3D structures. Protein Sci. 2006; 15 (11):2558–67.
  43. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 1992; 255(5047):959–65.
  44. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978; 47:45–148.
  45. Taylor WR, Thornton JM, Turnell WG: An ellipsoidal approximation of protein shape. Journal of Molecular Graphics 1983, 1:30-38.
  46. Thornton JM, Edwards MS, Taylor WR, Barlow DJ: Location of 'continuous' antigenic determinants in the protruding regions of proteins. EMBO J 1986, 5(2):409-413.
  47. Larsen MV, Lundegaard C, Lamberth K, *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007; 8:424.
  48. Buus S, Lauemøller SL, Wornig P, *et al.* Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens. 2003; 62(5):378–84.
  49. Lundegaard C, Nielsen M, Lund O. The validity of predicted T-cell epitopes. Trends Biotechnol. 2006; 24(12):537–8.
  50. Sébastien Giguère Alexandre Drouin, Alexandre Lacoste, Mario Marchand, Jacques Corbeil and François Lavolette, 2013. A Tool for the Prediction of Peptides Naturally Processed by the MHC. Paper under review in the Journal Of Immunological Methods.
  51. Moutaftsi M, Peters B, Pasquetto V, *et al.* A consensus epitope prediction approach identifies the breadth of murine T (CD8+)-cell responses to vaccinia virus. Nat Biotechnol. 2006; 24(7):817–9.
  52. Gowthaman U., Agrewala J. N., In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. Journal of proteome research 2007; 7(01): 154-163.
  53. Oyarzun P, Ellis JJ, Boden M and Kobe B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II protein diversity. BMC Bioinformatics, 2013;14:52.
  54. Schneidman-Duhovny D., Inbar Y., Nussinov R., Wolfson H. J., Patch Dock and Symm Dock: servers for rigid and symmetric docking. Nucleic Acids Research 2005; 33: W363-367.
  55. Andrusier N., Nussinov R., Wolfson H. J., FireDock: Fast Interaction Refinement in Molecular Docking, Proteins 2007; 69(1):139-159.
  56. Roehrig JT, Mathews JH, Trent DW. Identification of epitopes on the E glycoprotein of Saint Louis encephalitis virus using monoclonal antibodies. Virology. 1983; 128(1):118–26.
  57. Robson B, Pain RS. Analysis of the code relating sequence to conformation in globular proteins, the distribution of residue pairs in turns and kinks in the backbone chain. Biochem J, 1974; 141(3):899–904.
  58. Trainor NB, Crill WD, Roberson JA, *et al.* Mutation analysis of the fusion domain region of St. Louis encephalitis virus envelope protein. Virology. 2007; 360(2):398–406.
  59. Klavinskis LS, Whitton JL, Oldstone MB. Molecularly engineered vaccine which expresses an immunodominant T-cell epitope induces cytotoxic T lymphocytes that confer protection from lethal virus infection. J Virol. 1989;63(10):4311–6.
  60. Romagnani S, The Th1/Th2 paradigm. Immunology Today .1997, 18, 263-266.

**Cite this article as: Mohammad Mahfuz Ali Khan Shawan, Hafij Al Mahmud, Md. Mahmudul Hasan, Afroza Parvin, Md. Nazibur Rahman and S. M. Badier Rahman. In Silico Modeling and Immunoinformatics Probing Disclose the Epitope Based Peptide Vaccine Against Zika Virus Envelope Glycoprotein. Indian J. Pharm. Biol. Res.2014; 2(4):44-57.**