American Journal of Biomedical Research, 2016, Vol. 4, No. 3, 46-60
Available online at http://pubs.sciepub.com/ajbr/4/3/1

DOI:10.12691/ajbr-4-3-1

Highly Conserved Epitopes of ZIKA Envelope Glycoprotein May Act as a Novel Peptide Vaccine with High Coverage: Immunoinformatics Approach

Marwan Mustafa Badawi[1,*], Marwa Mohamed Osman[1], Afra AbdElhamid Fadl Alla[1], Ammar Mohammed Ahmedani[1], Mohamed hamed Abdalla[1], Mosab Mohamed Gasemelseed[2], Ahmed Abubakar Elsayed[3], Mohamed Ahmed Salih[1]

[1]Department of Biotechnology, Africa city of Technology- Khartoum, Sudan
[2]Al Neelain University, Faculty of Medical Laboratory Sciences
[3]Department of Microbiology, Soba University Hospital, Khartoum-Sudan
*Corresponding author: mmbadwi44new@gmail.com

Abstract Zika virus (ZIKV) is positive sense single stranded RNA of Flavivirus genus belonging to the Flaviviridae family. It has neither drug nor protective vaccine, and considered to be in relatedness to neurological abnormalities such as Guillain Barre Syndrome and microcephaly of neonates. The aim of this study is to analyze envelope glycoprotein E of all Zika strains using *in silico* approaches looking for conservancy, which is further studied to predict all potential epitopes that can be used after *in vitro* and *in vivo* confirmation as a therapeutic peptide vaccine. A total of 50 Zikavirusvariants' (include 12 from South America) polyproteins retrieved from NCBI database were aligned, and the conserved regions of Envelope Glycoprotein-E were selected for epitopes prediction. IEDB analysis resource was used to predict B and T cell epitopes and to calculate the population coverage. Epitopes with high scores in both B cell and T cell epitopes predicting tools were suggested. Three epitopes were proposed for international therapeutic peptide vaccine for B cell (**AQDKP**, **TPNSPRAE** and **TPHWNNK**) and two other epitopes designed especially for South America strains (**LDKQSDTQYV** and **EVQYAGTDGPCK**). For T cell epitopes, **MMLELDPPF** epitope was highly recommended as therapeutic peptide vaccine to interact with MHC class I along with three other epitopes (**MAVLGDTAW**, **KEWFHDIPL** and **DTAWDFGSV**) which showed very good population coverage against the whole world population. Three epitopes showed high affinity to interact with MHC class II alleles (**FKSLFGGMS**, **LITANPVIT** and **VHTALAGAL**) with excellent population coverage throughout the world and South America region. Herd immunity protocols can be achieved in countries with low population coverage percentage to minimize the active transmission of the virus, especially among pregnant women and other groups at risk.We recommend *in vitro* and *in vivo* proving the effectiveness of these proposed epitopes as a vaccine, as well as to be used as a diagnostic screening test.

Keywords: *Zika virus (ZIKV), arboviruses, peptide vaccine, Immune Epitope Database IEDB, epitopes, Herd immunity and Vaccine*

Cite This Article: Marwan Mustafa Badawi, Marwa Mohamed Osman, Afra AbdElhamid Fadl Alla, Ammar Mohammed Ahmedani, Mohamed hamed Abdalla, Mosab Mohamed Gasemelseed, Ahmed Abubakar Elsayed, and Mohamed Ahmed Salih, "Highly Conserved Epitopes of ZIKA Envelope Glycoprotein May Act as a Novel Peptide Vaccine with High Coverage: Immunoinformatics Approach." *American Journal of Biomedical Research*, vol. 4, no. 3 (2016): 46-60. doi: 10.12691/ajbr-4-3-1.

1. Introduction

Zika virus (ZIKV) is positive sense single stranded RNA, an arbovirus of *Flavivirus* genus belonging to the *Flaviviridae* family, which includes dengue, yellow fever, St. Louis encephalitis, West Nile and Japanese encephalitis viruses, among others [1,2,3].

ZIKA was initially isolated in 1947 from blood of a febrile sentinel rhesus monkey during a yellow fever study in the Zika forest of Uganda .The virus was subsequently isolated from a pool of *Aedes africanus* mosquitoes collected in 1948 from the same region of the Zika forest; a serological survey conducted at that time showed that 6.1% of the residents in nearby regions of Uganda had specific antibodies to ZIKV [3,4]. Up to 2006, only sporadic cases of ZIKV human infections were reported in literature [1,2,4]. In 2007, ZIKV has caused a large epidemic on Yap Island, Federated States of Micronesia and involved in infecting three quarters of local population. This outbreak shows that, ZIKV has been detected outside of Africa and Asia, having the potential as an emerging pathogen [5,6]. By October 2015, a single state in the northeast of Brazil (Bahia), reported 56,318 suspected cases of Zika virus disease. Brazilian national authorities estimate that between 497,593 and 1,048,701 cases of Zika virus infection have occurred since the outbreak

began, after Brazil, Colombia has been the most-affected country so far, with 20,297 cases reported (up to 23 January 2016) since the country's first cases were detected in October 2015. There has been a rapid regional spread of the virus. By 4 February 2016, 31 countries and territories in the Americas reported local transmission of the virus [7,8]. From 1 January 2007 to 25 February 2016, Zika virus transmission was documented in a total of 52 countries and territories. This includes 40 countries that reported local transmission between 2015 and 2016 [7]. It is likely to be transmitted and detected in other countries within the geographical range of competent mosquito vectors, especially *Aedes aegypti* [7]. Humans are get infected by infective mosquito bites, however recent report suggests that there is a possibility of secondary sexual transmission [9].

ZIKA is classified as Bio Safety Level 2 pathogen in the EU (with exception of the UK) and USA, and thought to be in relatedness to neurological diseases such as *Guillain Barre* Syndrome and microcephaly in neonates of infected mothers. Although ZIKV is a BSL2 pathogen, laboratories should assess the risks for pregnant laboratory personnel [5,10].

Currently, there is neither medicine against ZIKA fever nor specific antiviral treatment for clinical ZIKV infection, someone protects him/herself by preventing mosquito bites. Therefore, the development of a vaccine against ZIKV is very much important and this development has not been achieved yet. Surface or envelope proteins of the virus are the most antigenic ones and often considered as good candidates for immunization. They are important for vaccine development as it mediates the viral entry, and they are the primary targets of adaptive immune response [11,12]. Several institutions now are encouraging varieties of approaches to Zika virus vaccine development, using different strategies include DNA-based vaccine, a live-attenuated vaccine, a recombinant VSV-virus expressing Zika virus E glycoprotein and other approaches related to vector control [13].

The aim of this study is to analyze envelope glycoprotein E of all Zika strains using *in silico* approaches looking for conservancy, which is further studied to predict all potential B and T cell epitopes that can be used after in vitro and in vivo confirmation as a therapeutic peptide vaccine or diagnostic screening protocol, and regarding the current endemic in South America region; Zika isolated from these countries are separately further tested for identification of unique conserved epitopes.

2. Materials and Methods

2.1. Protein Sequence Retrieval

A total of 50 Zika virus strains' polyproteins from different geographic regions (include 12 strains from South America) were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/protein/?term=zika+virus+polyprotein) database in March 2016. These 50 strains sequences retrieved are from different parts of the world; 18 isolates were collected in South America region, Central America and the Caribbean; particularly from Brazil, Colombia, Guatemala, Martinique, Puerto Rico, Haiti and Suriname. Thirty-two International strains were collected in different countries, Mexico, Thailand, Philippines, China, Italy, Uganda, Senegal, Nigeria, Malaysia, Cambodia, French Polynesia and Central African Republic. Retrieved Strains and their Accession numbers are listed in Table 1.

Table 1. Virus Strains retrieved and their Accession numbers and area of collection

GenBank Protein Accession No.	Country	Region
*YP_002790881	Uganda	East Africa
AMD61711	Philippines	Southeast Asia
AMD61710	Thailand	Southeast Asia
AMO03410	China	Northeast Asia
AMH87239	Brazil	South America
AMA12087	Brazil	South America
AMA12086	Brazil	South America
AMA12085	Brazil	South America
AMA12084	Brazil	South America
ALX35659	Suriname	South America
BAP47441	Uganda	East Africa
AHZ13508	French Polynesia	Oceania
AHL43504	Senegal	East Africa
AHL43503	Senegal	East Africa
AHL43502	Senegal	East Africa
AHL43501	Senegal	East Africa
AHL43500	Central African Republic	Central Africa
ACD75819	Micronesia	Oceania
AMQ34004	Mexico: Chiapas	North America
AMQ34003	Mexico: Chiapas	North America
AMM39805	China	Northeast Asia
ALU33341	Brazil	South America
AMK79468	China	Northeast Asia
AMN14620	Italy: Padua	Europe
AMN14619	Italy: Padua	Europe
AMK49164	Brazil	South America
AMK49165	Brazil	South America
AML82110	China	Northeast Asia
AMK79469	China	Northeast Asia
AMD16557	Brazil	South America
AMC33116	Martinique	Caribe
AMB18850	Brazil: Rio Grande do Norte, Natal	South America
AMC13913	Guatemala	Central America
AMC13912	Guatemala	Central America
AMC13911	Puerto Rico	Central America
AMB37295	Haiti	North America
AHF49785	Central African Republic	Central Africa
AHF49784	Central African Republic	Central Africa
AHF49783	Central African Republic	Central Africa
AAV34151	Uganda	East Africa
ABI54475	Uganda	East Africa
AMM39806	China	Northeast Asia
AMM39804	Colombia: Barranquilla	South America
AMK02027	Uganda	East Africa
AHL37808	Canada	North America
AFD30972	Cambodia	Southeast Asia
AEN75266	Senegal	East Africa
AEN75265	Nigeria	East Africa
AEN75264	Malaysia	Southeast Asia
AEN75263	Uganda	East Africa

* *Ref* sequence.

From structural polyproteins, Envelope Glycoprotein-E was selected for epitopes prediction using NCBI Reference Sequence: (YP_002790881), Table 2.

Table 2. Structural proteins of Zika Virus obtained from NCBI

Structural Protein	Size	Start	End
Capsid Protein	122	1	122
Glycoprotein Precursor M	168	123	290
*Envelope Glycoprotein-E	500	291	790

*Selected protein.

2.2. Determination of Conserved Regions

The retrieved sequences were used as platform to obtain conserved regions using multiple sequence alignment (MSA) [14]. Sequences aligned with the aid of ClustalW as implemented in the BioEdit program, version 7.0.9.0 (Hall, 1999) for finding the conserved regions among international strains and for South America strains [15]. Later on, the candidate epitopes were analyzed by different prediction tools from Immune Epitope Database IEDB analysis resource (http://www.iedb.org/), Figure 1.

2.3. B-cell Epitope Prediction

B cell epitope is the portion of an immunogen, which interacts with B-lymphocytes. As a result, the B-lymphocyte is differentiated into antibody-secreting plasma cell and memory cell. B cell epitope is characterized by being hydrophilic, accessible and in a beta turn region [16]. Thus, the classical propensity scale methods and hidden Markov model programmed softwares from IEDB analysis resource were used for the following aspects:

Prediction of linear B-cell Epitopes: BepiPred from immune epitope database (http://toolsiedb.ofg/bcell/) [17] was used as linear B-cell epitopes prediction from the conserved region with a default threshold value of 0.023.

Prediction of surface accessibility: by using Emini surface accessibility prediction tool of the immune epitope database (IEDB) (http://tools.immuneepitope.org/tools/bcell/iedb) [18] the surface accessible epitopes were predicted from the conserved region holding the default threshold value 1.480.

Prediction of Epitopes antigenicity sites: (http://tools.immuneepitope.org/bcell/) [19] the kolaskar and tongaonker antigenicity method was used to determine the antigenic sites with a default threshold value of 0.860.

Prediction of epitopes hydrophilicity: parker hydrophilicity prediction tool of the IEBD database (http://tools.immuneepitope.org/bcell/) [20] was used to determine the hydrophilicity of the conserved regions; the threshold default value was 1.480.

Prediction of beta turns sites: Chou and Fasman beta turn prediction method (http://tools.immuneepitope.org/bcell/) was used with the default threshold 0.988 to determine the sites contain beta turns.

2.4. MHC Class I Binding Predictions

Analysis of peptide binding to MHC class I molecules was assessed by the IEDB MHC I prediction tool at http://tools.iedb.org/mhci/n, MHC-I peptide complex presentation to T lymphocytes undergo several steps. The attachment of cleaved peptides to MHC molecules step was predicted. prediction methods can be achieved by Artificial Neural Network (ANN), Stabilized Matrix Method (SMM), or Scoring Matrices derived from Combinatorial Peptide Libraries (Comblib_Sidney2008), consensus method was used which combines ANN, SMM and comblib different methods [21,22,23,24,25]. Prior to prediction, all epitope lengths were set as 9mers, all internationally conserved epitopes binded to alleles at score equal or less than 1.0 percentile rank was selected for further analysis. [26]

2.5. MHC Class II Binding Predictions

Analysis of peptide binding to MHC class II molecules was assessed by the IEDB MHC II prediction tool at http://tools.immuneepitope.org/mhcii/ [27,28]. For MHC-II binding predication, certain HLA-DR alleles were analyzed including: DRB1*01:01, DRB1*04:01, DRB1*07:01, DRB1*11:01, DRB1*15:01. MHC class II groove has the ability to bind to peptides with different lengths. This variability in binding makes prediction as difficult as less accurate [29]. There are four prediction methods for IEDB MHC II prediction tool: ARB, SMM_align, Sturniolo's method and a consensus method. ARB predict IC (50) values through combination of searches different peptide sizes and alleles into a single global prediction based on ARB matrices. SMM-align is a matrix-based method with extensions incorporating flanking residues outside of binding grooves. It also predict the IC50 values of peptides. The consensus approach is used to combine the outcome of the three methods. Firstly, a random scan set of Swiss-Prot proteins and achieve scores for 2,000,000 random peptides. Thereafter, act as reference to rank new predictions. The consensus method uses the median rank of the three approaches as the final prediction score [30]. All internationally conserved epitopes binded to alleles at score equal or less than 10-percentile rank was selected for further analysis.

2.6. Population Coverage Calculation:

All potential MHC I and MHC II binders from Zika virus envelope glycoprotein E was assessed for population coverage against the whole world population and South America population with the selected MHC-I and MHC-II interacted alleles by the IEDB population coverage calculation tool at http://tools.iedb.org/tools/population/iedb_input [31].

2.7. Homology Modeling

ZIKA envelope glycoprotein 3D structure was obtained by phyre2, (http://www.sbg.bio.ic.ac.uk/phyre2) which uses advanced remote homology detection methods to build 3D models, 99% of residues of envelope glycoprotein modelled at > 90% confidence. UCSF Chimera (version 1.8) was used to visualize the 3D structure, Chimera currently available within the Chimera package and available from the chimera web site (http://www.cgl.ucsf.edu/cimera). Homology modeling was achieved for further verification of the service accessibility and hydrophilicity of B lymphocyte epitopes predicted, as well as to visualize all predicted T cell epitopes in the structural level [32,33].

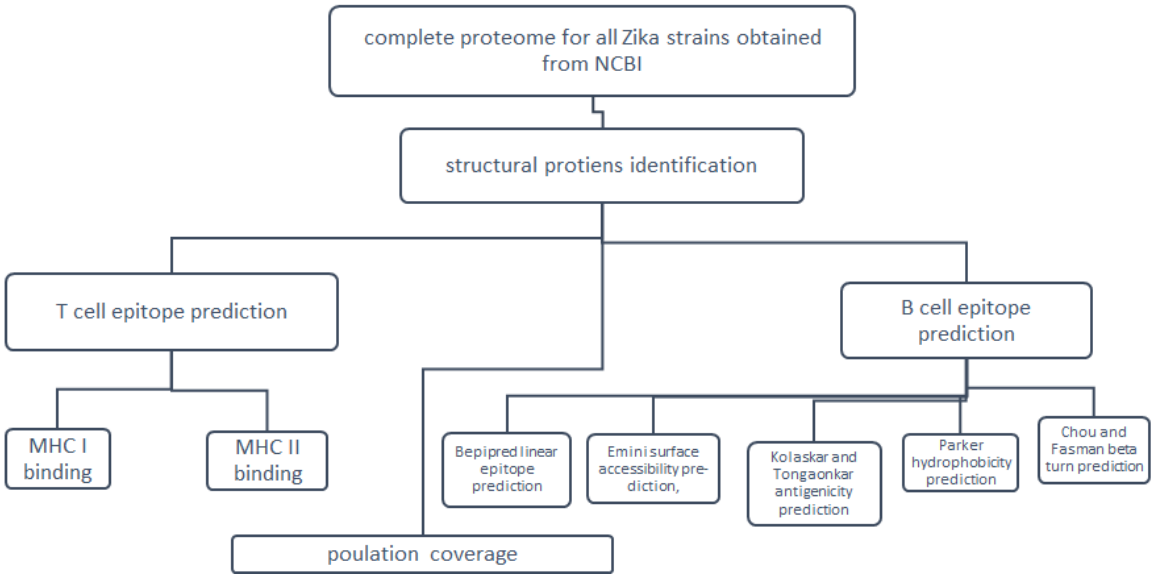

Figure 1. Workflow of prediction of potential B lymphocyte epitopes and T lymphocyte epitopes for the development of peptide vaccine or diagnostic tool for ZIKV infection

3. Results

3.1. Prediction of B-cell Epitope

Envelope protein was subjected to Bepipred linear epitope prediction, Emini surface accessibility, Kolaskar and Tongaonkar antigenicity, Parker hydrophilicity and Chou and Fasman beta turn prediction methods in IEDB, that predict the probability of specific regions in the protein to bind to B cell receptor, being in the surface, being immunogenic, being in a hydrophilic region and being in a beta turn region, respectively, Figure 2.

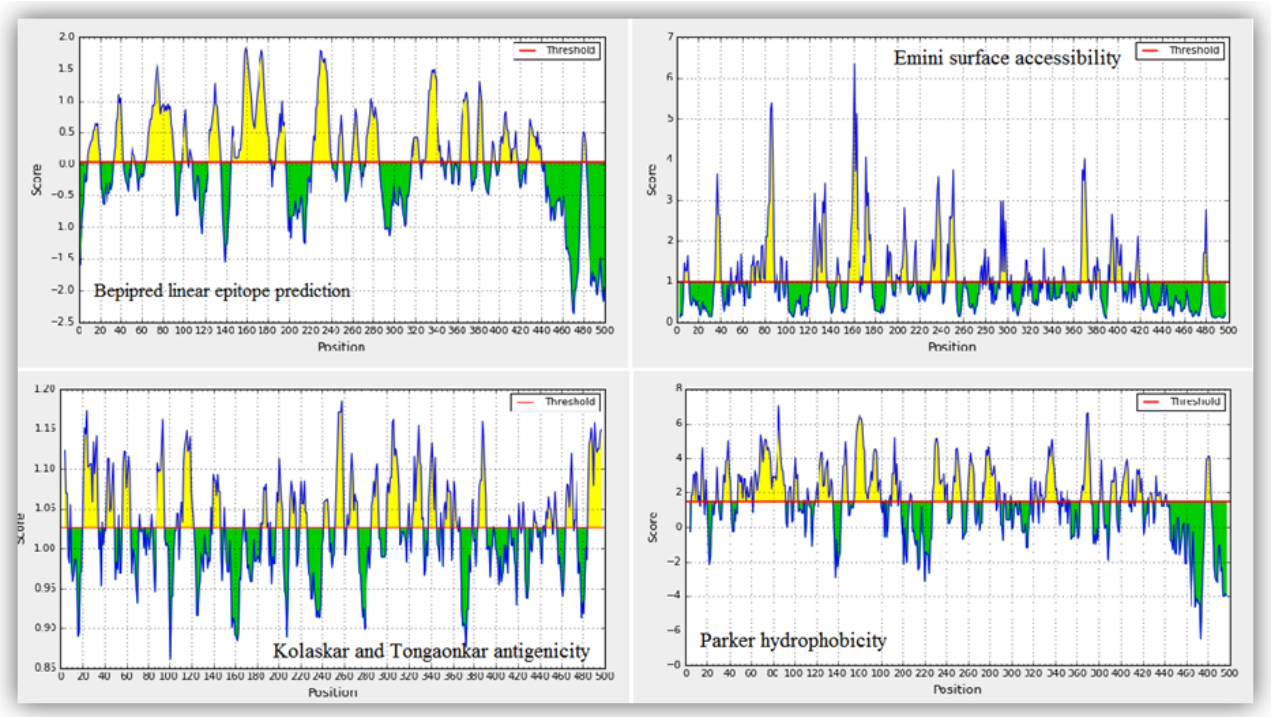

Figure 2. Prediction of B-cell epitopes by different scales

Yellow areas above threshold (red line) are proposed to be a part of B cell epitope. While green areas are not.

In Bepipred Linear Epitope Prediction method; the average binders score of Envelope Glycoprotein to B cell was 0.023, with a maximum of 1.831 and a minimum of -2.366, all values equal or greater than the default threshold 0.023 were predicted to be a potential B cell binders.

In Emini surface accessibility prediction; the average surface accessibility areas of the protein was scored as 1.000, with a maximum of 6.354 and a minimum of 0.074, all values equal or greater than the default threshold 1.0 were potentially in the surface.

The default threshold of antigenicity of the protein was 1.026; all values greater than 1.026 are potential antigenic determinants, the average of the antigenicity was 1.026, with a maximum of 1.186 and minimum of 0.861.

In Parker hydrophilicity prediction; the average hydrophilicity score of the protein was 1.480, with a maximum of 7.057and a minimum of -6.471, all values equal or greater than the default threshold 1.480 were potentially hydrophilic.

None of international strains epitopes had succeeded the Kolaskar and Tongaonkar antigenicity prediction method, however six B cell epitopes were found from international strains to satisfy all other scales of predicting B cell epitope. **AQDKP** from 35 to 39 was found to have the highest score, followed by **TPNSPRAE** from 170 to 177. The result is summarized in Table 3 and proposed epitopes are shown in Figure 3 at the structural level.

Table 3. list of B- cell epitopes predicted by different scales from international strains

Bepipred Linear Epitope Prediction							
Epitope	start	End	length	surface accesability[a] score	hydrophilicity score[b]	antigenicity score[c]	beta turn score[d]
AQDKP	35	39	5	3.142	5.18	0.988	1.126
LDKQSDT	82	88	7	3.461	4.886	0.978	1.127
*KSIQPENLE	128	136	9	2.959	2.856	0.989	1.004
*KSIQPEN	128	134	7	3.148	3.871	0.971	1.101
TPNSPRAE	170	177	8	4.609	4.625	0.952	1.167
TPHWNNK	233	239	7	3.71	2.729	0.922	1.217
TENS	369	372	4	2.313	6.625	0.887	1.172

[*] Peptide from 128 to 136 gives high score when it is shorten (128 to 134) in all tools but Kolaskar and Tongaonkar antigenicity.
[a] threshold: 1.000 [b] threshold : 1.480 [c] threshold: 1.026 [d] threshold : 0.988.
Position of peptides is according to position of amino acids in the Envelope glycoprotein.

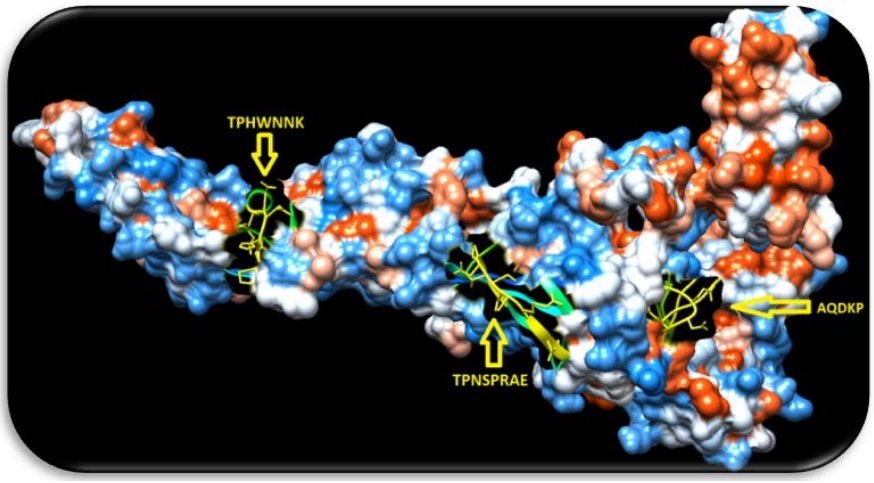

Figure 3. proposed B-Cell Epitopes in international strains

Proposed epitopes of B cell that are conserved in all international strains are shown here in the structural level of envelope glycoprotein of Zika virus.

In South America strains, the number of conserved Epitopes and their lengths were increased compared to international strains. Two epitopes "from South America" were satisfied the threshold values for all predicted scales: **LDKQSDTQYV** from 82 to 91 and **EVQYAGTDGPCK** from 329 to 340, as shown in Figure 4. The result of South America regions B cell epitopes is summarized in Table 4.

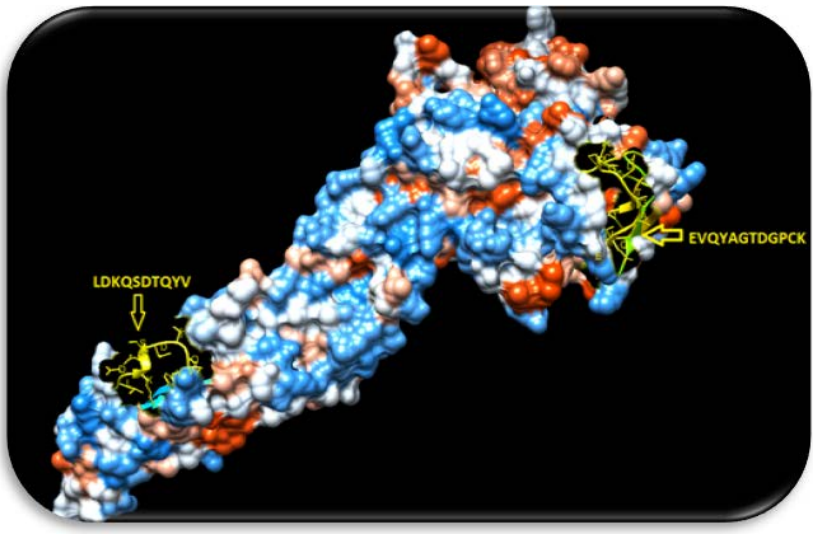

Figure 4. proposed B-Cell epitopes in South America strains

Proposed epitopes of B cell that are conserved in all South America strains are shown here in the structural level of envelope glycoprotein of Zika virus.

The Chou and Fasman beta turn prediction method was used with the default threshold 0.988 with maximum of 1.373 and a minimum of 0.739 for more confirmation for the prediction of the epitope to elicit B cell employed.

Table 4. list of B- cell epitopes predicted by different scales from South America strains

Bepipred Linear Epitope Prediction							
Epitope	Start	End	Length	Surface Accessibility score [a]	Hydrophilicity Score [b]	Antigenicity Score [c]	Beta turn Score[d]
MAQDKPTVD	34	42	9	2.365	3.689	0.991	1.017
[1*]SISDMASDSRCPTQGEA YLDKQSDTQYV	64	91	28	6.201	3.682	1.016	1.092
[1*]LDKQSDTQYV	82	91	10	3.711	3.46	1.041	1.051
DRGWGNG	98	104	7	1.204	4.043	0.861	1.373
KMTGKSIQPENL	124	135	12	2.545	2.525	0.966	1.036
[2*]TPNSPRAEATLG	170	181	12	1.093	3.4	0.976	1.093
[2*]TPNSPRAE	170	177	8	4.609	4.625	0.952	1.167
LPWHAGADTGTPHWNNKE	223	240	18	0.959	2.372	0.959	1.119
EAEMDGAKG	274	282	9	1.465	4.744	0.911	0.999
EVQYAGTDGPCK	329	340	12	1.075	3.842	1.034	1.107
PVITESTENS	363	372	10	1.736	3.64	0.992	1.031
DPPFGD	379	384	6	1.595	3.45	0.971	1.353
KKI	394	396	3	1.479	1.133	1.004	0.83
HRSGSTIGKA	401	410	10	1.146	3.57	0.98	1.098
TKNG	479	482	4	1.972	5.9	0.872	1.272

[1*] peptide from 64 to 91 gives higher score if it is shorten (82 to 91) in the Kolaskar and Tongaonkar antigenicity prediction.
[2*] peptide from 170 to 181 gives higher score if it is shorten (170 to 177) in all tools but Kolaskar and Tongaonkar antigenicity.
[a] threshold :1.000 [b] threshold : 1.480 [c] threshold: 1.026 [d] threshold : 0.988
Position of peptides is according to position of amino acids in the Envelope glycoprotein.

Table 5. list of epitopes that had binding affinity with the MHC Class I alleles

Epitope	Start	End	Allele	percentile Rank	ANN_ic50*
KEWFHDIPL	215	223	HLA-B*40:02	0.15	7
			HLA-B*40:01	0.2	7
			HLA-B*48:01	0.3	132
			HLA-B*18:01	0.95	238
KSLFGGMSW	454	462	HLA-B*57:01	0.15	13
			HLA-B*58:01	0.2	5
			HLA-A*32:01	0.3	13
GLDFSDLYY	195	203	HLA-A*01:01	0.2	9
			HLA-A*29:02	0.35	17
MAVLGDTAW	421	429	HLA-B*53:01	0.2	11
			HLA-B*58:01	0.2	7
			HLA-B*57:01	0.25	49
			HLA-B*35:01	0.5	13
MMLELDPPF	374	382	HLA-B*15:01	0.2	18
			HLA-B*46:01	0.25	728
			HLA-A*32:01	0.3	11
			HLA-B*35:01	0.3	6
			HLA-B*48:01	0.65	1651
			HLA-A*23:01	0.7	195
			HLA-A*29:02	0.7	34
			HLA-A*02:06	0.8	11
FSDLYYLTM	198	206	HLA-A*01:01	0.25	52
RLKGVSYSL	299	307	HLA-A*32:01	0.3	14
			HLA-B*08:01	0.9	152
DTAWDFGSV	426	434	HLA-A*68:02	0.3	6
			HLA-A*25:01	0.35	419
			HLA-A*26:01	0.6	381
SIQPENLEY	129	137	HLA-A*29:02	0.6	28
RLKMDKLRL	292	300	HLA-B*08:01	0.9	217
VHTALAGAL	265	273	HLA-B*39:01	0.7	31
			HLA-B*38:01	0.9	1437
SLFGGMSWF	455	463	HLA-A*26:01	0.8	130
			HLA-A*25:01	1	2362
WFHDIPLPW	217	225	HLA-A*23:01	0.8	187
			HLA-B*53:01	1	441
LAGALEAEM	269	277	HLA-B*35:01	1	44
TPNSPRAEA	170	178	HLA-B*07:02	0.8	57
SQILIGTLL	464	472	HLA-B*39:01	0.8	86
			HLA-B*38:01	0.7	1472
			HLA-B*48:01	0.4	300
MSWFSQILI	460	468	HLA-B*58:01	0.7	41
GMSWFSQIL	459	467	HLA-B*48:01	0.35	446
DPPFGDSYI	379	387	HLA-B*51:01	0.8	4252

*ANN ic50 is the inhibitory concentration needed for successful binding of peptide to MHC molecule by the Artificial Neural Network method. The lower the number the better is the epitope.
Position of peptides is according to position of amino acids in the Envelope glycoprotein.

3.2. Prediction of Cytotoxic T-lymphocyte Epitopes and Interaction with MHC Class I

Envelope Glycoprotein from international strains (including South America) was analyzed using IEDB MHC-1 binding prediction tool to predict T cell epitope suggested interacting with different types of MHC Class I alleles. Based on Consensus (ann/smm/comblib_sidney2008) with percentile rank ≤1; 19 peptides were predicted to interact with different MHC-1 alleles. The peptide **MMLELDPPF** from 374 to

382 had higher affinity to interact with 8 alleles (HLA-B*15:01, HLA-B*46:01, HLA-A*32:01, HLA-B*35:01, HLA-B*48:01, HLA-A*23:01, HLA-A*29:02 and HLA-A*02:06), followed by **MAVLGDTAW** from 421 to 429, **and KEWFHDIPL** from 215 to 223 that had affinity to interact with 4 alleles for each. The epitopes and their corresponding MHC-1 alleles are shown in Table 5.

Table 6. List of epitopes that had Binding affinity with the Class II alleles

epitope (core)	Allele	Percentile Rank	Peptide	Start	End
FKSLFGGMS	HLA-DRB1*11:01	1.82	IFGAAFKSLFGGMSW	448	462
	HLA-DRB1*08:02	2.72			
	HLA-DRB1*01:01	3.19			
	HLA-DRB1*04:01	9.08			
	HLA-DRB1*07:01	9.71			
	HLA-DRB1*15:01	9.94	AAFKSLFGGMSWFSQ	451	465
LITANPVIT	HLA-DRB1*08:02	3.06	PVGRLITANPVITES	354	368
	HLA-DRB1*11:01	5.07			
	HLA-DRB1*04:01	5.52			
	HLA-DRB1*01:01	7.88			
	HLA-DQA1*05:01	9.86			
	HLADQB1*03:01	9.86			
VHTALAGAL	HLA-DQB1*03:01	0.93	EGAVHTALAGALEAE	262	276
	HLA-DQA1*05:01	0.93			
	HLA-DRB1*01:01	7.73			
YYLTMNNKH	HLA-DRB1*04:01	0.11	DFSDLYYLTMNNKHW	197	211
	HLA-DRB1*07:01	4.9			
	HLA-DRB1*11:01	6.47			
FSDLYYLTM	HLA-DRB1*04:01	0.88	LDFSDLYYLTMNNKH	196	210
	HLA-DRB1*07:01	6.8			
ENLEYRIML	HLA-DRB1*15:01	1.59	QPENLEYRIMLSVHG	131	145
GHLKCRLKM	HLA-DRB1*11:01	2.41	RLSSGHLKCRLKMDK	283	297
LKCRLKMDK	HLA-DRB1*11:01	2.41	LSSGHLKCRLKMDKL	284	298
	HLA-DRB1*08:02	9.49	GHLKCRLKMDKLRLK	287	301
RLKMDKLRL	HLA-DRB1*11:01	2.41	GHLKCRLKMDKLRLK	287	301
LKMDKLRLK	HLA-DRB1*11:01	2.41	LKCRLKMDKLRLKGV	289	303
	HLA-DRB1*08:02	8.6			
TLGGFGSLG	HLA-DRB1*15:01	2.48	PRAEATLGGFGSLGL	174	188
LGGFGSLGL	HLA-DRB1*15:01	2.54	RAEATLGGFGSLGLD	175	189
LVEFKDAHA	HLA-DRB1*04:01	3.52	KEALVEFKDAHAKRQ	239	253
	HLA-DRB1*15:01	3.75			
	HLA-DRB1*08:02	3.8			
	HLA-DRB1*11:01	8.05			
	HLA-DRB1*07:01	9.94			
VGRLITANP	HLA-DRB1*04:01	3.75	LTPVGRLITANPVIT	352	366
	HLA-DRB1*08:02	4.65			
LYYLTMNNK	HLA-DRB1*15:01	3.91	DFSDLYYLTMNNKHW	197	211
LTPVGRLIT	HLA-DRB1*11:01	4.33	MQTLTPVGRLITANP	349	363
	HLA-DRB1*08:02	5.86			
	HLA-DRB1*15:01	6.52	QTLTPVGRLITANPV	350	364
SQILIGTLL	HLA-DRB1*15:01	4.43	WFSQILIGTLLMWLG	462	476
	HLA-DRB1*11:01	6.23			
	HLA-DRB1*01:01	9.25			
GRLITANPV	HLA-DRB1*15:01	5.39	LTPVGRLITANPVIT	352	366
	HLA-DRB1*01:01	7.73			
	HLA-DRB1*07:01	9.25			
ALVEFKDAH	HLA-DRB1*15:01	4.81	WNNKEALVEFKDAHA	236	250
WFHDIPLPW	HLA-DRB1*04:01	5.49	LVHKEWFHDIPLPWH	212	226
	HLA-DRB1*11:01	5.81			
GAAFKSLFG	HLA-DRB1*15:01		IHQIFGAAFKSLFGG	445	459
	HLA-DRB1*04:01				
MMLELDPPF	HLA-DRB1*04:01	6.18	NSKMMLELDPPFGDS	371	385
FSQILIGTL	HLA-DRB1*11:01	6.64	GGMSWFSQILIGTLL	458	472
	HLA-DRB1*01:01	8.09			
	HLA-DRB1*07:01	9.3			
WFSQILIGT	HLA-DRB1*11:01	6.64	FGGMSWFSQILIGTL	457	471
	HLA-DRB1*04:01	8.45			
FKDAHAKRQ	HLA-DRB1*04:01	7.02	ALVEFKDAHAKRQTV	241	255
	HLA-DRB1*11:01	8.05			
	HLA-DRB1*07:01	9.45			
MSWFSQILI	HLA-DRB1*08:02	6.28	FGGMSWFSQILIGTL	457	471
	HLA-DRB1*15:01	7.1			
LFGGMSWFS	HLA-DRB1*15:01	7.51	KSLFGGMSWFSQILI	454	468
FGGMSWFSQ	HLA-DRB1*04:01	8.45	KSLFGGMSWFSQILI	454	468
	HLA-DRB1*11:01	8.62			
SKMMLELDP	HLA-DRB1*04:01	9.97	STENSKMMLELDPPF	368	382
MLELDPPFG	HLA-DRB1*04:01	9.97	KMMLELDPPFGDSYI	373	387
*FEATVRGAK	HLA-DRB1*11:01	9.53	TIGKAFEATVRGAKR	406	420
	HLA-DQA1*05:01	9.93			
	HLA-DQB1*03:01	9.93			
*LHGTVTVEV	HLA-DQA1*04:01	5.8	IPAETLHGTVTVEVQ	317	331
	HLA-DQB1*04:02	5.8			
	HLA-DRB1*07:01	7.08			

*these two epitopes are not conserved in whole world strains but are conserved in South America strains.
Position of peptides is according to position of amino acid in the Envelope glycoprotein.

3.3. Prediction of T Helper Cell Epitopes and Interaction with MHC Class II

By the same way in IEDB MHC-1 binding prediction tool, T-cell epitopes from international strains (including South America), were analyzed using MHC-II binding prediction method; based on Consensus (smm/nn/sturniolo) with percentile rank ≤10.There were 30 predicted epitopes found to interact with MHC-II alleles for which the peptide (core) **FKSLFGGMS** and **LITANPVIT** had high affinity to interact with six alleles (HLA-DRB1*11:01, HLA-DRB1*08:02, HLA-DRB1*01:01, HLA-DRB1*04:0,1 HLA-DRB1*07:01, HLA-DRB1*15:01) and (HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*04:01, HLA-DRB1*01:01, HLA-DQA1*05:01, HLADQB1*03:01) respectively. Moreover, **LVEFKDAHA** epitopes that could interact with five alleles (HLA-DRB1*04:01, HLA-DRB1*15:01, HLA-DRB1*08:02, HLA-DRB1*11:01 and HLA-DRB1*07:01). The result is listed in Table 6.

3.4. Analysis of the Population Coverage

Epitopes that are suggested interacting with MHC-I and II alleles (especially high affinity binding epitopes and that can bind to different set of alleles) were selected for population coverage analysis. The results of population coverage of all epitopes are listed in Table 7.

Table 7. population coverage of all epitopes in both MHC class I and II in the world

Epitope	Coverage Class I	No. of alleles	Epitope	Coverage Class II	No. of alleles
*KEWFHDIPL	19.56%	4	*FKSLFGGMS	61.00%	6
KSLFGGMSW	11.53%	3	*LITANPVIT	70.93%	6
GLDFSDLYY	20.86%	2	*VHTALAGAL	61.47%	3
*MAVLGDTAW	17.68%	4	YYLTMNNKH	37.22%	3
*MMLELDPPF	33.02%	8	FSDLYYLTM	28.34%	2
FSDLYYLTM	17.34%	1	ENLEYRIML	18.41%	1
RLKGVSYSL	14.67%	2	GHLKCRLKM	37.22%	1
*DTAWDFGSV	11.46%	3	LKCRLKMDK	12.74%	2
SIQPENLEY	3.89%	1	RLKMDKLRL	10.54%	1
RLKMDKLRL	10.55%	1	LKMDKLRLK	12.74%	2
VHTALAGAL	5.93%	2	TLGGFGSLG	18.41%	1
SLFGGMSWF	9.07%	1	LGGFGSLGL	18.41%	1
WFHDIPLPW	7.83%	2	LVEFKDAHA	53.22%	5
LAGALEAEM	8.42%	1	VGRLITANP	13.41%	2
TPNSPRAEA	12.78%	1	LYYLTMNNK	18.41%	1
SQILIGTLL	7.58%	3	LTPVGRLIT	29.87%	3
MSWFSQILI	3.42%	1	SQILIGTLL	37.64%	3
GMSWFSQIL	1.70%	1	GRLITANPV	44.03%	3
DPPFGDSYI	7.43%	1	ALVEFKDAH	18.41%	1
Epitope set	**81.59%**		WFHDIPLPW	21.13%	2
			GAAFKSLFG	28.50%	2
			MMLELDPPF	11.21%	1
			FSQILIGTL	37.48%	3
			WFSQILIGT	21.13%	2
			FKDAHAKRQ	37.22%	3
			MSWFSQILI	20.51%	2
			LFGGMSWFS	18.41%	1
			FGGMSWFSQ	21.13%	2
			SKMMLELDP	11.21%	1
			MLELDPPFG	11.21%	1
			Epitope set	**83.01%**	

*proposed Epitopes.

In MHC class I, Four epitopes that interact with most frequent MHC class I alleles **(MMLELDPPF, KEWFHDIPL, MAVLGDTAW** and **DTAWDFGSV)** gave high percentage against the whole world population by IEDB population coverage tool. The maximum population coverage (80.45%) for these proposed epitopes was found in the Guinea-Bissau, while the higher population coverage in South America: (77.93%) was found in American Samoa followed by Venezuela (72.26%) then Chile Amerindian, Brazil Mixed and Chile (54.46%, 52.85% and 51.01% respectively). Table 8 represents the populations for which the coverage is greater than 50 % and Figure 5 shows these proposed Epitopes at the structural level.

Also in MHC class II, three epitopes that interact with most frequent MHC class II alleles (**LITANPVIT, VHTALAGAL** and **FKSLFGGMS**) gave high percentage against the whole world population by IEDB population coverage tool. The maximum population coverage for these proposed epitopes were found in two countries of South America: New Caledonia(95.75%) followed by Fiji (95.07%). Table 9 represents the populations for which the coverage is greater than 50 % and Figure 6 shows these proposed Epitopes at the structural level.

FEATVRGAK and LHGTVTVEV that bind to (HLA-DRB1*11:01, HLA-DQA1*05:01 and HLA-DQB1*03:01) and (HLA-DQA1*04:01 and HLA-DQB1*04:02); respectively were found to increase the population coverage in most South America regions and the surrounding areas. Moreover, they show their best coverage scores in Argentine, Brazil and Brazil Amerindians. The result is shown in Table 10.

Table 8. population coverage for Class I proposed epitopes throughout the world

Population / Area	Class I Epitope set coverage	Epitope coverage			
		MMLELDPPF	**MAVLGDTAW**	**KEWFHDIPL**	**DTAWDFGSV**
World	**58.82%**	**33.02%**	**17.34%**	**19.56%**	**11.46%**
East Asia	**71.36%**	**50.86%**	**17.07%**	**26.79%**	12.73%
Japan	72.10%	51.47%	15.89%	29.49%	13.88%
Korea; South	73.09%	48.25%	21.98%	27.21%	11.63%
Northeast Asia	**60.41%**	**39.87%**	**13.03%**	**19.47%**	**4.43%**
China	58.02%	38.53%	12.60%	17.68%	4.68%
Hong Kong	73.40%	44.26%	17.50%	31.19%	3.37%
South Asia	**68.97%**	**32.70%**	**16.95**	**37.02%**	**5.75%**
Malaysia	55.55%	18.50%	16.91%	37.60%	0.00%
Philippines	73.13%	22.00%	0.00%	70.84%	0.00%
Singapore	70.69%	32.02%	17.89%	28.79%	3.92%
Taiwan	79.88%	37.94	20.12%	48.66%	7.59
Thailand	63.10%	33.81%	19.63%	23.65%	2.03%
Vietnam	61.58%	36.14%	18.41%	15.87%	4.15%
Southwest Asia	**68.97%**	**32.70%**	**16.95**	**37.02%**	**5.75%**
Iran	63.46%	29.91%	39.47%	1370.00%	10.89%
Israel	60.21%	28.57	19.00%	26.04%	10.70%
Oman	66.93%	36.71%	32.43%	9.18%	22.38
Europe	**61.69%**	**32.05**	**18.17%**	**21.64%**	**16.17%**
Austria	68.37%	31.55%	21.15	19.00%	32.09%
Bulgaria	66.82%	36.10%	15.73%	25.35%	19.00%
Croatia	67.79%	36.91%	19.00%	27.75%	15.91%
Czech Republic	60.46%	29.99%	20.43%	21.32%	15.18%
England	57.00%	29.89%	15.17%	21.14%	9.54%
Finland	72.06%	45.18%	24.93%	34.14%	4.35%
France	66.86%	39.09	22.88%	18.46%	14.70%
Georgia	56.18%	33.39%	20.20%	16.47%	13.63%
Germany	63.70%	38.16%	20.09%	20.75%	11.75%
Ireland Northern	55.07%	32.45%	18.40%	16.42%	9.10%
Ireland South	56.44%	29.29%	19.00%	13.88%	13.88%
Italy	72.81%	39.11%	33.05%	34.62%	11.30%
Poland	63.35%	30.53%	18.00%	23.09%	19.70%
Portugal	59.62%	38.39%	17.94%	17.36%	9.18%
Romania	66.42%	37.71%	23.44%	28.09%	14.07%
Russia	70.54%	39.97%	24.73%	31.75%	11.33%
Russia Siberian	77.98%	44.95%	32.31%	36.67%	9.32%
Sweden	69.67%	40.87%	12.52%	35.62%	8.04%
East Africa	**64.16%**	**31.06%**	**29.88%**	**9.47%**	**17.21%**
Kenya	65.70%	29.60%	34.34%	8.92%	18.78%
Uganda	64.61%	31.09%	30.37%	11.31%	15.12%
Zambia	61.58%	34.33%	25.35%	11.08%	11.42%
Zimbabwe	73.98%	42.54%	28.43%	12.76%	21.50%
West Africa	**73.05%**	**50.29%**	**41.86%**	**7.99%**	**11.29%**
Burkina Faso	51.46%	7.94%	39.97%	2.68%	8.66%
Cape Verde	78.97%	55.45%	48.38%	10.89%	13.14%
Guinea-Bissau	80.54%	56.51%	54.03%	8.99%	13.98%
Ivory Coast	55.02%	11.05%	51.89%	4.54%	0.00%
Senegal	73.13%	54.35%	42.74%	5.31%	11.44%
Central Africa	59.49%	38.81%	28.59%	4.30%	10.91%
Cameroon	66.05%	47.00%	34.06%	3.70%	9.04%
Sao Tome and Principe	77.42%	37.28%	51.12%	4.00%	25.52%
North Africa	**62.62%**	**40.67%**	**26.78%**	**8.71%**	**12.54%**
Mali	78.94%	59.16%	52.11%	1.40%	12.02%
Morocco	60.26%	35.72%	19.50%	12.45%	14.69%
Sudan	54.88%	31.50%	21.88%	8.58%	13.14%
Tunisia	62.11%	41.07%	24.67%	12.95%	7.42%
South Africa	**60.60%**	**39.37%**	**17.87%**	**3.80%**	**17.67%**
West Indies	**71.30%**	44.89%	30.39%	15.38%	13.70%
Cuba	70.29%	42.99%	30.39%	15.38%	13.88%
North America	**65.83%**	**38.96%**	**26.99%**	**21.61%**	**11.81%**
Mexico	69.56%	55.09%	25.49%	22.68%	6.78%
United States	66.13%	38.82%	27.13%	21.79%	11.89%
United States Amerindian	76.29%	60.40%	30.38%	42.57%	3.43%
Brazil Mixed	52.85%	28.77%	19.36%	12.58%	12.34%
Chile	51.01%	23.80%	8.31%	22.68%	11.40%
Chile Amerindian	54.46%	26.53%	0.00%	34.39%	4.74%
Venezuela	72.26%	18.70%	1.15%	37.35%	42.15%
American Samoa	77.93%	38.69%	0%	66.36%	11.64%
Australia	53.59%	14.23%	6.81%	38.80%	4.38%

Table 9. population coverage for Class II proposed epitopes throughout the world

Population / Area	Class II Epitope Set coverage	Epitope coverage		
		FKSLFGGMS	**LITANPVIT**	**VHTALAGAL**
World	**83.01%**	**61.00%**	**70.93%**	**61.47%**
East Asia	**62.35%**	**41.91%**	**51.71%**	**41.80%**
Japan	56.40%	32.53%	50.06%	41.54%
Korea; South	62.54%	47.86%	46.14%	37.16%
Mongolia	83.13%	52.90%	75.88%	66.25%
Northeast Asia	71.93%	32.21%	63.99%	59.73%
China	71.93%	32.21%	63.99%	59.73%
South Asia	**77.91%**	**59.96%**	**57.95%**	**49.93%**
India	77.91%	59.96%	57.95%	49.93%
Southeast Asia	**77.65%**	**59.31%**	**57.97%**	**50.09%**
Indonesia	65.32%	27.34%	57.97%	52.65%
Malaysia	51.97%	24.50%	40.02%	37.08%
Singapore	77.65%	46.27%	63.55%	59.14%
Taiwan	78.61%	31.96%	74.38%	68.77%
Thailand	62.17%	32.46%	47.88%	44.29%
Vietnam	61.47%	20.32%	53.69%	51.64%
Southwest Asia	**71.30%**	**29.57%**	**65.45%**	**60.33%**
Iran	81.33%	46.96%	74.74%	68.40%
Israel	83.78%	45.40%	75.30%	70.97%
Jordan	71.87%	49.34%	65.97%	47.31%
Lebanon	81.41%	57.14%	75.56%	58.86%
Saudi Arabia	63.14%	55.82%	27.21%	17.59%
Europe	**86.03%**	**68.43%**	**72.10%**	**62.46%**
Austria	86.47%	79.02%	66.67%	50.18%
Belgium	70.90%	54.98%	60.48%	46.01%
Croatia	84.62%	45.28%	81.87%	76.97%
Czech Republic	90.40%	71.78%	78.11%	71.29%
Denmark	85.33%	74.09%	73.44%	57.15%
England	89.89%	79.12%	73.32%	59.75%
France	89.32%	70.52%	79.03%	69.24%
Georgia	90.28%	62.53%	84.95%	75.99%
Germany	92.08%	76.89%	80.87%	72.12%
Greece	84.23%	49.03%	78.43%	72.77%
Ireland Northern	84.07%	77.74%	53.04%	38.10%
Ireland South	88.40%	80.03%	64.06%	50.66%
Italy	58.34%	19.79%	51.82%	48.81%
Macedonia	88.83%	50.66%	83.94%	80.10%
Netherlands	79.14%	62.44%	65.37%	51.57%
Norway	90.86%	79.75%	75.28%	61.79%
Poland	88.03%	70.45%	74.92%	66.61%
Portugal	79.13%	57.44%	62.20%	55.42%
Russia	85.17%	52.79%	79.14%	70.74%
Scotland	83.72%	75.50%	55.34%	40.43%
Slovenia	89.80%	67.69%	81.32%	75.26%
Spain	82.17%	62.55%	64.36%	57.36%
Sweden	83.00%	63.71%	74.66%	60.67%
Turkey	88.87%	55.64%	83.13%	77.16%
Ukraine	50.64%	50.64%	50.64%	0.00%
East Africa	**69.25%**	**30.44%**	**56.29%**	**57.54%**
Kenya	68.69%	0.00%	48.41%	68.69%
Zimbabwe	65.65%	30.44%	53.35%	52.57%
West Africa	89.08%	24.50%	71.97%	85.78%
Gambia	85.12%	0.00%	64.12%	85.12%
Central Africa	55.06%	18.13%	38.06%	45.35%
Congo	58.62%	19.90%	42.05%	49.35%
Gabon	59.95%	26.73%	50.35%	46.32%
North Africa	**65.33%**	**23.16%**	**45.02%**	**54.05%**
Algeria	74.11%	29.09%	54.45%	64.20%
Ethiopia	55.44%	8.64%	40.29%	49.51%
Morocco	62.72%	24.61%	40.30%	50.85%
Sudan	54.19%	21.85%	16.83%	40.71%
Tunisia	68.10%	24.70%	47.31%	55.79%
West Indies	71.94%	32.27%	54.75%	60.38%
Cuba	60.69%	29.44%	15.91%	49.07%
Jamaica	59.04%	4.55%	39.80%	53.94%
Martinique	58.44%	35.06%	35.06%	38.90%
North America	92.37%	68.79%	85.04%	78.86%
Canada	93.52%	21.36%	93.09%	91.94%
Mexico	81.24%	39.91%	78.13%	69.47%
United States	93.99%	69.04%	88.10%	83.22%
United States Amerindian	98.86%	33.91%	98.79%	98.30%
Central America	**89.59%**	**41.86%**	**87.29%**	**82.33%**
Costa Rica	94.18%	17.19%	93.39%	92.98%
South America	**82.03%**	**41.97%**	**78.37%**	**70.29%**
Argentina	89.37%	46.42%	85.98%	81.26%
Bolivia	70.90%	51.56%	70.49%	39.94%
Brazil	87.31%	47.02%	84.66%	76.98%

Brazil Amerindian	91.62%	42.53%	91.52%	85.49%
Brazil Caucasoid	65.81%	56.15%	43.39%	30.42%
Brazil Mixed	77.96%	49.25%	68.81%	60.67%
Chile	77.90%	46.37%	71.47%	60.03%
Colombia	77.46%	38.94%	73.40%	64.88%
Ecuador	66.13%	19.68%	66.13%	57.84%
Peru	79.41%	33.00%	79.18%	69.62%
Oceania	80.77%	36.78%	76.71%	69.83%
Cook Islands	90.39%	42.45%	89.85%	83.54%
Fiji	95.07%	64.33%	94.15%	86.66%
Kiribati	68.40%	9.37%	66.24%	65.13%
Nauru	77.49%	33.71%	72.74%	66.05%
New Caledonia	95.75%	77.24%	92.57%	81.62%
New Zealand	77.73%	52.87%	70.63%	55.45%
Niue	89.45%	26.56%	88.93%	85.64%
Papua New Guinea	86.98%	48.07%	82.23%	74.92%
Samoa	79.80%	30.89%	77.41%	70.97%
Tonga	61.10%	34.39%	58.16%	40.71%

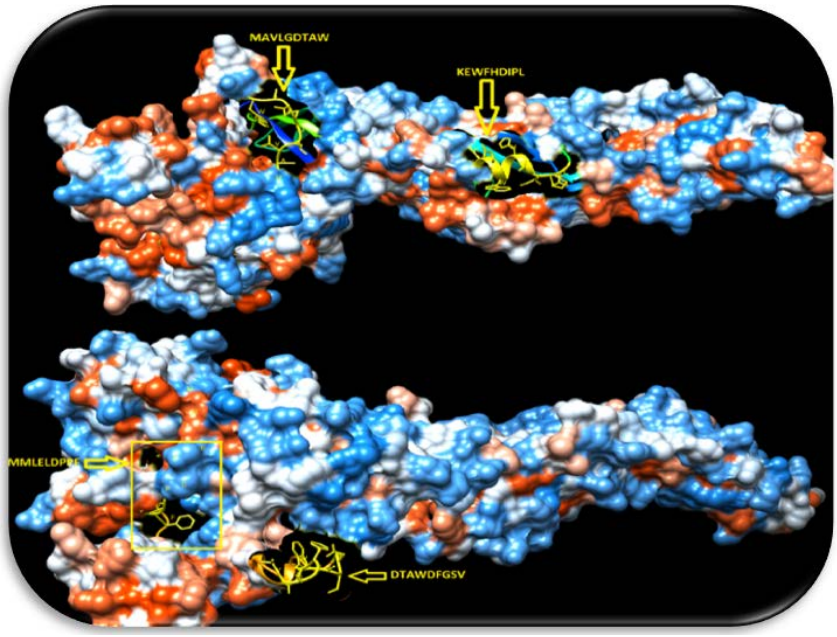

Figure 5. proposed T-Cell epitopes that interact with MHC Class I

Proposed epitopes of MHC Ithat are conserved in all international strains are shown here in the structural level of envelope glycoprotein of Zika virus. Both upper and lower structures are envelope glycoprotein of ZIKV, shown twice for better visualization purposes.

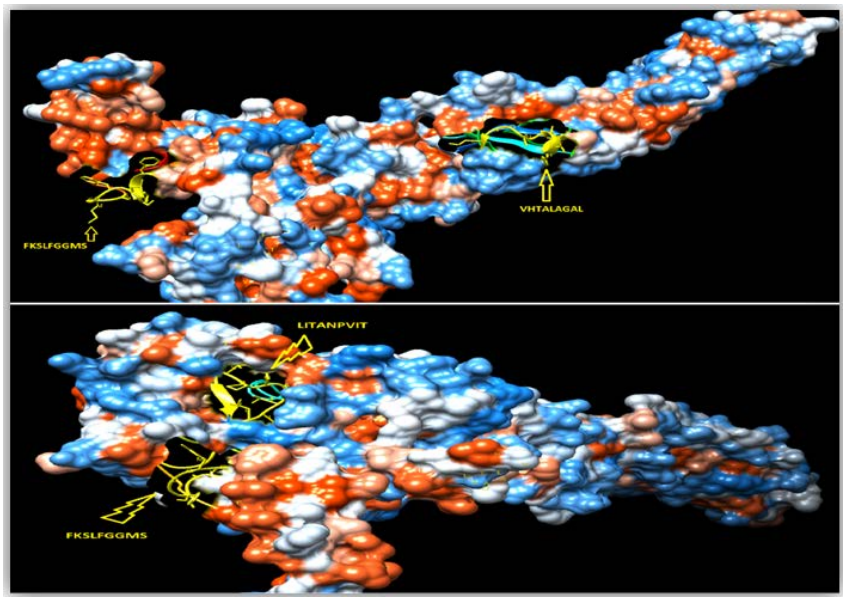

Figure 6. proposed T-Cell epitopes that interact with MHC Class II

Proposed epitopes of MHC II that are conserved in all international strains are shown here in the structural level of envelope glycoprotein of Zika virus. Above and below pictures show two of proposed epitopes for each,both pictures show envelope glycoprotein of ZIKV, shown twice for better visualization purposes.

Table (10.A). MHC class II proposed epitopes' population coverage in South America regions

epitope/population	South America	Guatemala Amerindian	Argentina	Argentina Amerindian	Costa Rica*	Guatemala*
FKSLFGGMS	41.97%	41.32%	46.42%	37.27%	17.19%	41.32%
LITANPVIT	78.37%	52.68%	85.98%	91.69%	93.39%	52.68%
VHTALAGAL	70.29%	25.04%	81.26%	88.42%	92.98%	25.04%
FEATVRGAK	70.60%	26.25%	81.47%	88.05%	92.98%	26.25%
LHGTVTVEV	52.08%	34.36%	41.01%	51.58%	16.86%	34.36%
Epitope set	**92.56%**	**72.70%**	**94.12%**	**97.77%**	**94.82%**	**72.70%**

* Central America

Table (10.B). MHC class II proposed epitopes' population coverage in South America regions

epitope/population	Bolivia	Bolivia Amerindian	Brazil	Brazil Amerindian	Chile	Colombia
FKSLFGGMS	51.56%	51.56%	47.02%	42.53%	46.37%	38.94%
LITANPVIT	70.49%	70.49%	84.66%	91.52%	71.47%	73.40%
VHTALAGAL	39.94%	39.94%	76.98%	85.49%	60.03%	64.88%
FEATVRGAK	39.94%	39.94%	77.57%	85.74%	63.49%	64.24%
LHGTVTVEV	51.35%	51.35%	64.23%	72.80%	38.49%	54.30%
Epitope set	**89.02%**	**89.02%**	**96.82%**	**99.24%**	**86.87%**	**91.11%**

Table (10.C). MHC class II proposed epitopes' population coverage in South America regions

epitope/population	Colombia Amerindian	Ecuador	Ecuador Amerindian	Paraguay	Peru	Peru Amerindian	Venezuela
FKSLFGGMS	37.40%	19.68%	19.68%	4.90%	33.00%	33.00%	0.00%
LITANPVIT	75.79%	66.13%	71.69%	47.49%	79.18%	79.72%	42.85%
VHTALAGAL	66.99%	57.84%	64.76%	44.79%	69.62%	70.39%	42.85%
FEATVRGAK	66.47%	57.84%	64.76%	44.79%	69.47%	70.25%	42.85%
LHGTVTVEV	57.58%	40.82%	48.25%	38.66%	54.26%	56.10%	41.38%
Epitope set	**92.80%**	**82.72%**	**88.25%**	**70.39%**	**92.92%**	**93.56%**	**69.28%**

4. Discussion

Vaccination has proven to be the mainstay in prevention of various deadly infectious diseases [34,35,36]. Historically, live-attenuated or inactivated forms of microbial pathogens (viruses, bacteria, etc.) have been used for induction of antigen-specific responses that protect the host against subsequent infections. Based on the pathogen being used, such vaccine formulations can contain anywhere between tens of to a few hundred proteins. However, protective immunity is usually dependent upon a few select proteins within such formulations, whereas the majority of proteins are unnecessary for the induction of protective immunity. Furthermore, these additional proteins may induce allergenic responses, thus emphasizing the need to eliminate them from vaccine formulations. This rationale led to an interest in subunit vaccines using single, or a select few, proteins of the microbes in vaccine formulations for induction of protective immunity [37,38]. An extension of this logic would be that even single proteins contain many hundreds of antigenic epitopes, all of which are not necessary; whereas some may even be detrimental to the induction of protective immunity. This has created an interest in peptide vaccines containing only epitopes capable of inducing positive, desirable T cell and B cell mediated immune responses [39]. There are many peptide vaccines under development, such as vaccine for human immunodeficiency virus (HIV) [40], hepatitis C virus (HCV) [41], malaria [42], foot and mouth disease [43], swine fever [44], influenza [45], anthrax [46], human papilloma virus (HPV) [47], therapeutic anti-cancer vaccines [48-53] for pancreatic cancer, melanoma, non-small cell lung cancer, advanced hepatocellular carcinoma, coetaneous T-cell lymphoma and B-Cell chronic lymphocytic leukemia. A database of publicly and privately conducted clinical studies is maintained on ClinicalTrials.gov, which is a service of the U.S. National Institute of Health. Around 562 clinical studies of peptide vaccines for preventive or therapeutic purposes on multiple disease conditions are registered with this database until 30th -March 2016.

In this study, we aimed to determine the highly potential immunogenic epitopes for B and T cells, the prime molecules of cell mediated and humoral immunity, as vaccine candidate for ZIKA virus from envelope glycoprotein.

Among the retrieved strains, the earliest strain (YP_002790881) was collected at 1947 from Uganda, and several strains were collected in the sixties, seventies, eighties and the first decade of 21st century, 2013, 2014 and 2015. Last strains collected among retrieved were (AMQ34003 and AMQ34004), as they are collected at 25 February 2016. As a result, we think that the conserved regions among international strains accompanied with good level of confidence.

To determine a potential and effective peptide antigen for B cell, epitopes should get above threshold scores in Bepipred linear epitope prediction, Emini surface accessibility, Parker hydrophilicity, Kolaskar and Tongaonkar antigenicity and Chou and Fasman beta turn prediction methods in IEDB. According to epitopes illustrated in Table 3, these epitopes are the only conserved regions from international strains of zika virus envelope glycoprotein that have high probability of activating B lymphocyte, however, these epitopes predicted were satisfied most of scales of predicting B cell epitope but none of them succeeded the Kolaskar and Tongaonkar antigenicity prediction method. The linear epitope 35 **AQDKP** 39 showed the highest score in Emimi surface and Parker hydrophilicity followed by 128 **KSIQPEN** 134 epitope. Shawan *et al.* (2014) suggested different epitope 123 **DAHAKRQTVVVLGSQEGAV** 141 with 19 a.a length as elicit B lymphocyte response,

but this epitope was not located in any conserved regions of envelope glycoprotein in zika virus strains [54].

According to Table 4; South America strains showed a very great similarity, so the number and length of the conserved Epitopes were increased. The linear epitope 64 **SISDMASDSRCPTQGEAYLDKQSDTQYV** 91, 28 amino acid length in South America strains showed high score in Emini surface and Parker hydrophilicity, but not the antigenicity scale, however, when we decreased the epitope length to 10 amino acid **LDKQSDTQYV** the antigenicity score jumped above the default threshold reaching 1.041. In addition, the linear epitope 329 **EVQYAGTDGPCK** 340, 12 amino acid gave high antigenicity score equal 1.034. Therefore epitopes that are suggested to activate B cell in South America region had more affinity to induce an immune response than the universal epitopes regarding activation of B lymphocyte independently of T lymphocyte.

Since the immune response of T cell is long lasting response comparing with B cell, where the antigen can easily escape the antibody memory response [54] and CD8+ T and CD4+ T cell responses play a major role in antiviral immunity [55], designing of vaccine against T cell epitope is much more promising. For MHC Class I Alleles prediction, we chose the most common HLA-A and HLA-B alleles [56]. So, according to Table 5; among 19 T cell internationally conserved epitopes predicted to interact with MHC Class I, we found the epitope 374 **MMLELDPPF** 382 had higher affinity to interact with 8 alleles (HLA-B*15:01, HLA-B*46:01, HLA-A*32:01, HLA-B*35:01, HLA-B*48:01, HLA-A*23:01, HLA-A*29:02 and HLA-A*02:06). The same epitope was predicted by Shawan *et al.* (2014) as the highest immunogenicity epitope as 0.9139 I PMHC score [54]. In addition, we found epitopes 421 **MAVLGDTAW** 429 and 215 **KEWFHDIPL** 223, had the affinity to interact with four alleles of each, while 426 **DTAWDFGSV** 434 epitope had the affinity to interact with three alleles Table 5. These four epitopes had very good population coverage for class I alleles throughout the world with high coverage percentage in American Samoa and Venezuela. As shown in Table 8, we found **MMLELDPPF** epitope covering 33.02% of the world population and it gave higher coverage percentage in West and North Africa while **KEWFHDIPL** epitope showed higher population coverage in American Samoa (66.36%).

A. Arnaiz- villena *et. al* (2006) reported that there were four HLA-A alleles and four HLA-B alleles (A*02, A*24, A*31, A*68, B*35, B*39, B*40 and B*48) with frequencies higher than 5% found in all Amerindian populations which is distributed differently in south America and the surrounding regions, reaching their highest in Bolivia, Guatemala, Peru and Honduras as 55%, 47%, 46% and 44% of the country population, respectively [57,58]. So, according to our result in Table 5; we found that noticeable, and we found that the 215 **KEWFHDIPL** 223 epitope interact with 2 of HLA-B alleles that are reported in Amerindian population (B*40 and B*48) while **MMLELDPPF** interact with one HLA-A allele (A*02) and two HAL-B allele (B*35 and B*48), Therefore these epitopes may show high affinity in Amerindian populations.

We found that 30 predicted internationally conserved epitopes interacting with MHC-II alleles and we represented this result in Table 6. Epitopes **FKSLFGGMS** and **LITANPVIT** had high affinity to interact with six alleles while **VHTALAGAL** epitope could interact with three alleles.

Related to our result represented in Table 9, we found these three epitopes had excellent population coverage for Class II alleles throughout the world and high coverage percentage in South America region (82.03%). Epitopes **FKSLFGGMS** and **LHGTVTVEV** interact with two of class II haplotype found in high frequency in several Meso and South American groups (HLA-DRB1*08:02 and HLA-DQB1*04:02) alleles that reported in A. Arnaiz-Villena *et.a.l* (2006). However, in our study **FKSLFGGMS** epitope showed slightly low coverage percentage in South America comparing to other epitopes [57]. **FEATVRGAK** epitope interact with HLA-DQB1*03:01 allele which found in different Amerindian populations [57]. In our result listed in Table 6; we found both **FEATVRGAK** and **LHGTVTVEV** epitopes are conserved in South America region. **LITANPVIT** alone showed high coverage percentage in population of most of South America countries reached up to 91.52% in Brazil Amerindians and 84.66% of Brazil. We suggested special set of epitopes in Table (10A, 10B, 10C) for South America region, and we showed that all these epitopes had very high coverage percentage in Brazil, one of the epidemic countries with zika virus according to WHO report for 2015-2016 [59].

It is to be noted that for efficient induction of either B-cell or cytotoxic T cell responses, the induction of a robust helper T cell response is crucial [60,61].

5. Conclusion

Conventional peptide vaccine development methods are costly, and time consuming, the role *in silico* prediction tools do is highly appreciated as they select specific peptides in protein, which then tested *in vitro* and *in vivo* to verify and prove the effectiveness of the proposed epitopes to induce an immune response, as well as to be used as a diagnostic screening test. Herd immunity protocols can be achieved in countries with low percentage of population coverage to minimize the active transmission of the virus, especially among pregnant women and other groups at risk.

Acknowledgments

Authors would like to thank African City of Technology members for their cooperation.

Competing Interest

The authors declare that they have no competing interests.

References

[1] Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg*. 1952; 46(5): 509-20.

[2] Hayes EB. Zika virus outside Africa. Emerg Infect Dis. 2009; 15(9): 1347-50.

[3] Robert S. Lanciotti, Olga L. Kosoy, Janeen J. Laven, Jason O. Velez, Amy J. Lambert, Alison J. Johnson, Stephanie M. Stanfi eld, and Mark R. Duffy. Genetic and Serologic Properties of Zika Virus Associated with an Epidemic, Yap State,Micronesia, 2007.*Emerging Infectious Diseases*. 2008; 14 (8): 1232-39.

[4] Dick GW. Zika virus pathogenicity and physical properties. Trans R Soc Trop Med Hyg. 1952; 46: 521-34.

[5] 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.

[6] Foy BD, Kobylinski KC, Chilson Foy JL, Blitvich BJ, Travassos da Rosa A, Haddow AD, et al. Probable nonvector-borne transmission of Zika virus, Colorado, USA. Emerg Infect Dis. 2011; 17(5): 880-2.

[7] World Health Organization. ZIKA VIRUS MICROCEPHALY AND GUILLAIN-BARRÉ SYNDROME. 2016. Available at: http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/22/microcefalia-protocolo-de-vigilancia-e-resposta-v1-3-22jan2016.pdf

[8] World Health Organization. *ZIKA SITUATION REPORT*. 2016. Print.

[9] Charrel RN, Leparc-Goffart I, Pas S, de Lamballerie X, Koopmans M & Reusken C. State of knowledge on Zika virus for an adequate laboratory response [Submitted]. Bull World Health Organ E-pub: 10 Feb 2016.

[10] Services YSDoH. Zika virus. Information for clinicians and other health professionals. Federated States of Micronesia: Yap State Department of Health; 2007. p. 2.

[11] Cerdeño-Tárraga AM, Efstratiou A, Dover LG, et al. The complete genome sequence and analysis of *Corynebacterium diphtheriae*NCTC13129. *Nucleic Acids Research*. 2003; 31(22): 6516-6523.

[12] Trent DW, Qureshi AA. Structural and nonstructural proteins of Saint Louis encephalitis virus. *Journal of Virology*. 1971; 7(3):379-88.

[13] Global Research Collaboration for Infectious Disease Preparedness,. Zika Virus (ZIKV) Outbreak. 2015. Print.

[14] Rezaul Islam, M Sadman Sakib & Aubhishek Zaman. A computational assay to design an epitope- based peptide vaccine against chokungunya virus. *Future virol*.2012 7(10): 1029-1042.

[15] Arafat Rahman Oany, Tahmina Sharmin,Afrin Sultana Chowdhury,Tahmina Pervin Jyot and Md Anayet Hasan. Highly Conserved Reagions in Ebola Virus RNA dependent RNA polymerase may be act as a universal novel peptide vaccine target: a computational approach.Oany etal, In Silico Pharmacology(2015) 3:7.

[16] Anayet Hasan, Mehjabeen Hossain and Md. Jibran Alam. A Computational Assay to Design an Epitope-Based Peptide Vaccine Against Saint Louis Encephalitis Virus. *Bioinformatics and Biology Insights* 2013:7 347-355.

[17] Jens Erik Pontoppidan Larsen, Ole Lund and Morten Nielsen. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006; 2: 2.

[18] Emini EA, Hughes JV, Perlow DS, Boger J. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55:836-839.

[19] Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*276:172-174.

[20] Parker JM, Guo D, Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25:5425-5432.

[21] Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. 2012. Immune epitope database analysis resource. NAR.

[22] Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O. 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12:1007-1017.

[23] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, and Nielsen M. 2008. NetMHC-3.0: Accurate web accessible predictions of Human, Mouse, and Monkey MHC class I affinities for peptides of length 8-11. NAR 36:W509-512.

[24] Peters B, Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6:132.

[25] Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B. 2008. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res 4:2.

[26] Kim Y, Ponomarenko J, Zhu Z, et al. Immune epitope database analysis resource. Nucleic Acids Research. 2012;40(Web Server issue):W525-W530.

[27] Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol. 4(4):e1000048.

[28] Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. 2010. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics. 11:568.

[29] Pratik Narain Srivastava, Richa Jain, Shyam Dhar Dubey, Sharad Bhatnagar, Nabeel Ahmad. Prediction of Epitope-Based Peptides for Vaccine Development from Coat Proteins GP2 and VP24 of Ebola Virus Using Immunoinformatics, International Journal of Peptide Research and Therapeutics (2016) 22:119-133.

[30] Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, P. E. *et al*. (2008). Immune epitope database analysis resource (IEDB-AR).*Nucleic Acids Research*, *36*(Web Server issue), W513-W518.

[31] Bui HH,Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006 Mar 17;7:153.

[32] The Phyre2 web portal for protein modeling, prediction and analysis Kelley LA et al. Nature Protocols 10, 845-858 (2015).

[33] UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12.

[34] Bachler, B.C.; Humbert, M.; Palikuqi, B.; Siddappa, N.B.; Lakhashe, S.K.; Rasmussen, R.A.; Ruprecht, R.M. Novel biopanning strategy to identify epitopes associated with vaccine protection. J. Virol. 2013, 87, 4403-4416.

[35] Perrie, Y.; Kirby, D.; Bramwell, V.W.; Mohammed, A.R. Recent developments in particulate-based vaccines. Recent Pat. Drug Deliv. Formul. 2007, 1, 117-129.

[36] Black, M.; Trent, A.; Tirrell, M.; Olive, C. Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists. Expert Rev. Vaccines 2010, 9, 157-173.

[37] Thompson, A.L.; Staats, H.F. Cytokines: The future of intranasal vaccine adjuvants. Clin. Dev. Immunol. 2011.

[38] Petrovsky, N.; Aguilar, J.C. Vaccine adjuvants: Current state and future trends. Immunol. Cell Biol. 2004, 82, 488-496.

[39] Sesardic, D. Synthetic peptide vaccines. J. Med. Microbiol. 1993, 39, 241-242.

[40] Liu, Y.; McNevin, J.; Zhao, H.; Tebit, D.M.; Troyer, R.M.; McSweyn, M.; Ghosh, A.K.; Shriner, D.; Arts, E.J.; McElrath, M.J.; et al. Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: Fitness-balanced escape. J. Virol. 2007, 81, 12179-12188.

[41] Kolesanova, E.F.; Sanzhakov, M.A.; Kharybin, O.N. Development of the schedule for multiple parallel ―difficult Peptide synthesis on pins. Int. J. Pept. 2013.

[42] Epstein, J.E.; Giersing, B.; Mullen, G.; Moorthy, V.; Richie, T.L. Malaria vaccines: Are we getting closer? Curr. Opin. Mol. Ther. 2007, 9, 12-24.

[43] Volpina, O.M.; Gelfanov, V.M.; Yarov, A.V.; Surovoy, A.Y.; Chepurkin, A.V.; Ivanov, V.T. New virus-specific T-helper epitopes of foot-and-mouth disease viral VP1 protein. FEBS Lett. 1993, 333, 175-178.

[44] Tarradas, J.; Monso, M.; Munoz, M.; Rosell, R.; Fraile, L.; Frías, M.T.; Domingo, M.; Andreu, D.; Sobrino, F.; Ganges, L. Partial protection against classical swine fever virus elicited by dendrimeric vaccine-candidate peptides in domestic pigs. Vaccine 2011, 29, 4422-4429.

[45] Stanekova, Z.; Kiraly, J.; Stropkovska, A.; Mikušková, T.; Mucha, V.; Kostolanský, F.; Varečková, E. Heterosubtypic protective immunity against influenza a virus induced by fusion peptide of the hemagglutinin in comparison to ectodomain of M2 protein. Acta Virol. 2011, 55, 61-67.

[46] Oscherwitz, J.; Yu, F.; Cease, K.B. A synthetic peptide vaccine directed against the 2ss2–2ss3 loop of domain 2 of protective

antigen protects rabbits from inhalation anthrax. J. Immunol. 2010, 185, 3661-3668.

[47] Solares, A.M.; Baladron, I.; Ramos, T.; Valenzuela, C.; Borbon, Z.; Fanjull, S.; Gonzalez, L.; Castillo, D.; Esmir, J.; Granadillo, M.; et al. Safety and immunogenicity of a human papillomavirus peptide vaccine (CIGB-228) in women with high-grade cervical intraepithelial neoplasia: first-in-human, proof-of-concept trial. ISRN Obstet. Gynecol. 2011.

[48] Bernhardt, S.L.; Gjertsen, M.K.; Trachsel, S.; Møller, M.; Eriksen, J.A.; Meo, M.; Buanes, T.; Gaudernack, G. Telomerase peptide vaccination of patients with non-resectable pancreatic cancer: A dose escalating phase I/II study. Br. J. Cancer 2006, 95, 1474-1482.

[49] Brunsvig, P.F.; Aamdal, S.; Gjertsen, M.K.; Kvalheim, G.; Markowski-Grimsrud, C.J.; Sve, I.; Dyrhaug, M.; Trachsel, S.; Møller, M.; Eriksen, J.A.; et al. Telomerase peptide vaccination: A phase I/II study in patients with non-small cell lung cancer. Cancer Immunol. Immunother. 2006, 55, 1553-1564.

[50] Brunsvig, P.F.; Kyte, J.A.; Kersten, C.; Sundstrøm, S.; Møller, M.; Nyakas, M.; Hansen, G.L.; Gaudernack, G.; Aamdal, S. Telomerase peptide vaccination in NSCLC: A phase II trial in stage III patients vaccinated after chemoradiotherapy and an 8-year update on a phase I/II trial. Clin. Cancer Res. 2011, 17, 6847-6857.

[51] Kyte, J.A.; Gaudernack, G.; Dueland, S.; Trachsel, S.; Julsrud, L.; Aamdal, S. Telomerase peptide vaccination combined with temozolomide: A clinical trial in stage IV melanoma patients. Clin. Cancer Res. 2011, 17, 4568-4580.

[52] Greten, T.F.; Forner, A.; Korangy, F.; N'Kontchou, G.; Barget, N.; Ayuso, C.; Ormandy, L.A.; Manns, M.P.; Beaugrand, M.; Bruix, J. A phase II open label trial evaluating safety and efficacy of a telomerase peptide vaccination in patients with advanced hepatocellular carcinoma. BMC Cancer 2010, 10, e209.

[53] Kyte, J.A.; Trachsel, S.; Risberg, B.; Thor, S.P.; Lislerud, K.; Gaudernack, G. Unconventional cytokine profiles and development of T cell memory in long-term survivors after cancer vaccination. Cancer Immunol. Immunother. 2009, 58, 1609-1626. 25.

[54] 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 PeptideVaccine Against Zika Virus Envelope Glycoprotein. Indian J. Pharm. Biol. Res. 2014; 2(4):44-57.

[55] Andrea J. Sant and Andrew McMichael. Revealing the role of CD4+ T cells in viral immunity. Journal of Experimental Medicine. 2012; 209(8): 1391-1395.

[56] Paul S1, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide binding repertoires of different size, affinity and immunogenicity. Journal of Immunology. 2013; 191(12):5831-9.

[57] A. Arnaiz-Villena, J. Moscoso, J.I. Serrano-Vela, J. Martinez-Laso. The uniqueness of amerindians according to HLA genes and the peopling of the Americas. Journal of Inmunología. 2006; 25: 13-24.

[58] Lizcano Fernández, Francisco "Composición Étnica de las Tres Áreas Culturales del Continente Americano al Comienzo del Siglo XXI".Convergencia (in Spanish) (Mexico: Universidad Autónoma del Estado de México, Centro de Investigación en Ciencias Sociales y Humanidades). 2005; 38: 185-232.

[59] Zika virus infection outbreak, Brazil and the Pacific region". Stockholm: European Centre for Disease Prevention and Control. 25 May 2015. p. 4. Retrieved 12 February 2016.

[60] Testa, J.S.; Philip, R. Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. Future Virol. 2012, 7, 1077-1088. 28. Lanier, J.G.; Newman, M.J.; Lee, E.M.;

[61] Sette, A.; Ahmed, R. Peptide vaccination using nonionic block copolymers induces protective anti-viral CTL responses. Vaccine 1999, 18, 549-557.