BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 13(12)

Hypothesis

Computer aided epitope design as a peptide vaccine component against *Lassa* virus

Ar-Rafi Md. Faisal[1, $], Syed Hassan Imtiaz[2, $], Tasnim Zerin[1], Tania Rahman[1], Hossain Uddin Shekhar[1*]

[1]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh; [2]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh; Hossain Uddin Shekhar – E-mail: hossainshekhar@du.ac.bd; *Corresponding Author; [$]Equal contribution

Received November 26, 2017; Revised December 5, 2017; Accepted December 5, 2017; Published December 31, 2017;

Abstract:
Lassa virus (LASV) is an arena virus causing hemorrhagic fever and it is endemic in several regions of West Africa. The disease-causing virus records high mortality rate in endemic regions due to lack of appropriate treatment and prevention strategies. Therefore, it is of interest to design and develop viable vaccine components against the virus. We used the *Lassa* virus envelope glyco-proteins as a vaccine target to identify linear peptides as potential epitopes with immunogenic properties by computer aided epitope prediction tools. We report a T-cell epitope 'LLGTFTWTL' and a B-cell epitope 'AELKCFGNTAVAKCNE' with predicted potential immunogenicity for further *in vivo* and *in vitro* consideration.

Keywords: Lassa virus, envelope glycoprotein, epitope-based vaccine

Background:
Lassa virus is a single-stranded, enveloped, bi-segmented, ambisense RNA virus belonging to *Arenaviridae* virus family **[1]**. It causes severe forms of acute viral hemorrhagic fever known as Lassa hemorrhagic fever, endemic in regions of West Africa **[2]**. In West Africa, *Lassa* virus (LASV) causes as many as 300,000 infections and approximately 5,000 deaths per year **[3]**. This high rate of mortality in endemic areas, along with the absence of effective treatment and vaccination options, makes LASV an important pathogen to study. LASV was first identified in Lassa village, Borno State at the northeastern region of Nigeria in 1969 **[4]**. This zoonotic virus exhibits persistent, asymptomatic infection with profuse urinary virus excretion in *Mastomys natalensis*, the ubiquitous and highly commensal rodent host **[2]**. Human transmission occurs through food or household items contaminated with infected *Mastomys* rats' urine or feces. Person-to-person transmission can occur through direct contact with the blood, urine, feces or other bodily secretions of infected person and indirect contact with environments contaminated with such fluids, which makes it highly susceptible for endemics or epidemics **[5]**. Sexual transmission has been reported and pregnant patients with Lassa fever results in spontaneous abortions **[6]**. Both sexes and all age groups of people appear to be affected by this virus and there is no epidemiological evidence supporting airborne spread between humans (WHO, 2017).

The incubation period of Lassa fever ranges from 6–21 days **[7]**. About 80% of people who become infected with LASV have no symptoms. 1 in 5 infections result in severe disease where the virus affects several organs such as liver, spleen and kidneys. The onset of the disease, when it is symptomatic, is usually gradual, starting with fever, general weakness, and malaise followed by sore throat, muscle pain, chest pain and in severe cases facial edema, fluid in the lung cavity, bleeding from the mouth, nose, vagina or gastrointestinal tract may develop **[2]**. Sensory-neural hearing loss (SNHL) is one of the common complications affecting as many as 25% of patients and rendering an estimated 1 to 2% of the population hearing impairment in areas with high rates of LASV infection **[4]**. Death usually occurs within 14 days of onset in fatal cases. Because of diverse and non-specific symptoms, it is often difficult to clinically diagnose Lassa fever and distinguish it from other viral hemorrhagic fevers such as *Ebola* virus disease, typhoid fever and yellow fever especially in the early course of the disease **[8]**. LASV outbreak first appeared

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

in 1972 in Zorzor, Liberia **[9]**. According to WHO, there have been reports of re-emerged LASV infections followed by high mortality endemic outbreaks in Nigeria (2012); Nigeria, Benin, Togo, Sweden, Liberia (2016); Nigeria, Benin, Togo and Burkina Faso (2017) (WHO, 2017).

LASV genome contains two RNA segments coding for two proteins each. The larger segment is approximately 7.2kb and encodes a small zinc-binding protein regulating transcription, replication and RNA polymerase **[10]**. The smaller segment is approximately 3.4kb encoding the nucleoprotein and the envelope glycoprotein **[11]**. Even though the mortality caused by LASV was first reported in almost 45 years ago, little effort has been made to cure and/or prevent its detrimental effects till date. Treatment with the antiviral drug ribavirin seems to be effective for Lassa fever but it must be administered in the first week of illness for optimal efficacy **[12]**. However, even early commencement of ribavirin therapy seems not to offer protection against development of SNHL **[4]**. The best immediate prospect to control this disease in endemic areas lies with the use of a vaccine.

Innovative stride towards the development of LASV vaccine was made earlier in this century and it was reported that trials on primates were successful with the vaccine, but there were no reports on the human trials or further advancement of that approach **[13]**. Currently, there is no effective vaccine against Lassa fever (WHO, 2017). Developing inactivated vaccines would be an option but they are less effective and need regular booster injections **[14]**. Therefore, developing peptide vaccines would be more effective to restrain the spread or new emerging epidemics of *Lassa* virus **[15]**. Hence, we performed analysis to predict T and B cell epitope based peptides with LASV envelope glycoprotein having known immunity **[16]**. Various immuno-informatics approaches were used to find effective T and B cell epitopes for the design and development of a viable vaccine against LASV.

Methodology:
An outline of the methodology used in this study shown in **Figure 1**.

LASV protein sequences:
The 12 sequences for glyco-proteins of different isolates of LASV were retrieved from NCBI (National center for biotechnology information) database. These sequences were obtained from different strains of LASV at different time frame (like at 1987, 1989, 2002, 2010, 2011 and 2017) in different endemic regions of Africa.

Multiple Sequence Alignment (MSA):
Retrieved sequences were subjected to multiple sequence alignment using MEGA 7.0.26 software package (http://www.megasoftware.net). The CLUSTALW algorithm along with 1000 bootstrap value and other default parameters were adopted to fabricate the alignment. The sequences were analyzed to recognize the immunologically relevant regions that were achieved by predicting epitopic peptides. An amino acid stretch must be of a minimum length for being considered as an epitope that we are aiming to design. Due to representative length of peptide that binds to HLA molecules, nonamers were selected as the minimum length of the conserved sequences for the prediction of epitope-based peptide in the current study.

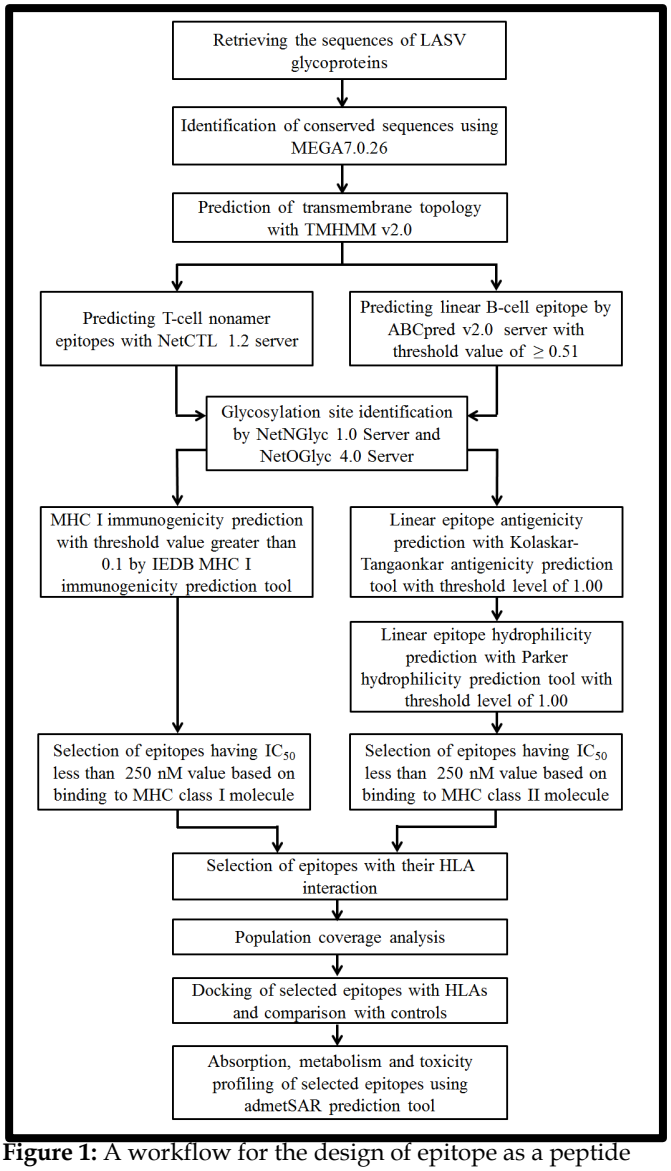

Figure 1: A workflow for the design of epitope as a peptide vaccine component against LASV.

Transmembrane Topology Analysis:
The conserved regions from the glyco-proteins were analyzed to distinguish their soluble and membrane regions. Each selected conserved sequence was subjected to transmembrane topology prophecy using TMHMM v0.2 server to identify the inner, outer and transmembrane portion with high degree of accuracy **[17]**. It should be noted that vaccine components can be effectively designed with epitopes at the exposed regions of the membrane.

T-Cell Epitope prediction in conserved regions:
The cytotoxic T lymphocyte (CTL) epitopes from the conserved peptides were predicted using the NetCTL 1.2 server, which is based on neural network architecture and predicts candidate epitopes based on the processing of the peptides *in vivo* **[18]**. During analysis, threshold, sensitivity and specificity were set at 0.75, 0.8 and 0.97 respectively. A combined algorithm integrating MHC class I binding, transporter of antigenic peptides (TAP) transport efficiency and proteasomal cleavage prediction was involved to predict a total/overall score. Weight on C-terminal cleavage 0.15 and weight on TAP (transporter of antigenic peptides) transport efficiency was determined 0.05. The best epitope candidates were chosen based on a combined score. Primarily selected epitopes were undertaken for further analysis. For note, these epitopes were nonamer in length.

Linear B-Cell Epitope Prediction from the Conserved Sequences:
Linear B-cell epitope prediction was done with ABCpred v2.0 server, which uses artificial neural network **[19]**. This webserver is based on recurrent neural network (machine based technique) using fixed length patterns. The epitope length was selected as sixteen-mer. Epitopes, which had crossed the threshold level of 0.51, were selected for further analysis. The epitopes were also checked with Bepipred linear epitope prediction tool by IEDB **[20]**.

Prediction of glycosylation sites:
NetNGlyc 1.0 Server and NetOGlyc 4.0 Server were used to check the glycoprotein sequences respectively. The epitopes, which had no glycosylation site, were considered for further analysis.

MHC Class I and T-Cell Epitope Interaction:
The immune epitope database and analysis resource (IEDB) MHC class I binding prediction tool was used to calculate IC_{50} (half maximal inhibitory concentration) values for peptides binding to specific MHC I molecules. This tool is based on stabilized matrix method-based prediction (SMM) **[21]**. Epitopes taken from NetCTL analysis and having no glycosylation site were used here. T-cell epitopes were then analyzed with IEDB MHC I immunogenicity scores. The parameters for immunogenicity detection (TAP score, proteasomal score and IC_{50} values) were normalized on a scale of 0 to 1 and were given a weighted score to prioritize the parameters in order to determine immunogenicity **[22]**.

MHC Class II and B-Cell Epitope Interaction:
Prior to predicting epitopes interacting with MHC class II, they were first checked for their antigenic properties like antigenicity and hydrophilicity by Kolaskar-Tangaonkar antigenicity prediction tool and Parker hydrophilicity prediction tool respectively **[23, 24]**. The epitopes that didn't possess any glycosylation sites were analyzed in these tools and those, which crossed the threshold level of 1.00 for each of the properties, were further checked to have association with MHC II alleles. IEDB MHC II binding tool was used to check the association with MHC II alleles **[25]**. Likewise in MHC I binding interaction analysis, SMM was also used to analyze HLA relationship with epitopes. The epitopes, which had highest interactions, were selected for further analysis.

Population Coverage:
Selected T-cell and B-cell epitopes were used for population coverage analysis using population coverage tool from the IEDB analysis resource **[26]**. The allele frequency of the interacting HLA alleles was used to measure the population coverage for the corresponding epitope.

Molecular Docking Study of HLA-Epitope Interaction:
Retrieving 3D Structure of HLAs:
The 3D structure of HLA-A*02:01 (PDB ID: 3UTQ) and HLA-DRB1*01:01 (PDB ID: 1AQD) were downloaded from the Protein Data Bank (PDB) database. Prior to docking, all the water molecules were removed from the 3D structure of epitope free HLA-A*02:01 and HLA-DRB1*01:01.

Designing of the 3D Epitope Structure:
The 3D structures of epitopes were designed using the PEP-FOLD peptide structure prediction server at the Ressource Parisienne en Bioinformatique Structurale Mobyle Portal **[27]**.

HLA-Epitope Binding Prediction:
The AutoDock tool from the Molecular Graphics Laboratory software package (version 1.5.6) was employed for docking purpose. Both the proteins and ligands (epitopes) files were converted to PDBQT format for using in this docking study. The grid/ space box center for B-cell epitope and HLA-DRB1*01:01 was set at 32.4206, -0.6947 and 6.2001 A˚ in the x, y and z axes respectively, so that the epitope could bind at the binding groove of HLA-DRB1*01:01. The grid/ space box center for T-cell epitope and HLA-A*02:01 was set at 11.9650, 0.3823 and 16.3061 A˚ in the x, y and z axes respectively, so that the epitope could bind at the binding groove of HLA-A*02:01. The size was set at 104.8110, 83.3764 and 133.0247 A˚ in the x-, y- and z-dimensions, respectively.

All the analysis was done at 1.00 A˚ spacing. The number of outputs was set at 10, while the exhaustiveness was kept at the default 8.00. AutoDock Vina program was used to perform the actual docking based on these parameters **[28]**.

Outputs were again visualized in the PYMOL molecular graphics system and the best output was selected based on higher binding energy. The binding affinity or the docking results were collected in .csv format which opens in Microsoft excel. The non-bond interaction between proteins and the ligands were visualized using Accelrys Discovery Studio Visualizer (version 2017 R2). The types of bonding, their distance and the categories of bonding were also visualized. The hydrophobic and ionizable receptor surfaces were determined using this software.

Absorption, Metabolism and Toxicity Profiling of Candidate Peptide Vaccines:

Absorption, metabolism and toxicity profiles of the peptide epitopes were predicted using "admetSAR" prediction server (http://lmmd.ecust.edu.cn/admetsar1/).

Results:

Identification of conserved regions in antigen sequences:

A total of 12 sequences of glycoproteins from different isolates of LASV have been retrieved from GenBank at NCBI (**Table S1**). The CLUSTALW program in MEGA software generated several conserved sequences with varying lengths. A total of 10 conserved sequences were found.

Transmembrane Topology Determination:

TMHMM v2.0 prediction analysis revealed that 6 (out of 10) conserved regions of envelope glycoproteins fulfilled the criteria of outer membrane characteristics (**Table S2**).

Immunogenicity Prediction of T-Cell Epitope:

NetCTL tool predicted 66 possible interactions with MHC super-types, which were, determined with a combinatorial score of TAP score, MHC I binding score, proteasomal cleavage score and antigen processing score.

Linear B-Cell Epitope Prediction:

ABCpred server was used to predict sixteen-mer epitopes from the selected conserved sequences. This tool found 29 epitopes. The Bepipred linear epitope prediction tool was used to analyze conserved regions for epitope identification. This is followed by accuracy crosscheck for epitope prediction using ABCpred (**Figure 2**).

Glycosylation site prediction:

11 N-glycosylation sites were found but there were no sites for O-glycosylation in the target protein sequence (**Figure 3**). Among the 66 nonamer T cell epitopes and 29 sixteen-mer B-cell epitopes, 51 T cell epitopes and 22 B-cell epitopes did not possess any glycosylation sites (**Table S3, S5**).

MHC Class I and T-Cell Epitope Interaction:

All the 51 T-cell epitopes were then analyzed with an stabilization matrix method in IEDB MHC-I binding prediction tool having IC_{50} score of less than 250 nM (IC_{50}< 250 nM) matching 81 possible MHC-I allele interactions with 17 different T-cell epitopes (**Table S4**). Top 6 epitopes were selected having IEDB MHC I antigenicity score over 0.2 (**Table 1**). They were further checked with population coverage and molecular docking.

MHC Class II and B-Cell Epitope Interaction:

The 22 B cell epitopes that were selected by ABCpred server and did not contain any glycosylation site within them were checked with Kolaskar-Tongaonkar antigenicity prediction tool and Parker hydrophilicity prediction tool for predicting the antigenicity and the hydrophilicity of the epitopes (**Table S5**). 7 epitopes crossed the threshold level of 1.00 for each of the properties (antigenicity and hydrophilicity) and were selected for further analysis. IEDB MHC II binding prediction tool then provided interactions between these epitopes and MHC II alleles. Interactions, which had IC50 value less than 250 nM, were then selected (**Table S6**). Finally three 16-mer epitopes were selected based on prominent interactions with MHC II alleles (**Table 2**). These epitopes were further checked with population coverage and molecular docking.

Population Coverage:

For the epitopes of class I MHC molecules, Europe showed a coverage of 59%, North America showed almost 50%, West Africa showed 22.14% and the worldwide coverage was 50.04% (**Table S7**). 'LLGTFTWTL' showed highest coverage among the T-cell epitopes, which is 38% (**Figure 4a**). For the epitope of class II MHC molecules, epitopes provided the coverage of 36.78% in Europe, which was the highest, and worldwide coverage was just at 28.63%. West Africa showed population coverage of 12.69% (**Table S8**). 'AELKCFGNTAVAKCNE' showed 18% coverage, which is the highest in case of B-cell epitopes (**Figure 4b**).

Molecular Docking Study of HLA-Epitope Interaction:

Binding models of the good epitopes ('LLGTFTWTL' and 'AELKCFGNTAVAKCNE' epitopes; **Figure 5**) to its specific HLA molecules were completed using AutoDock Vina.

T cell epitope 'LLGTFTWTL' bound to the binding groove of HLA-A*02:01 with a binding energy of -10.5 kcal/mol is shown (**Figure 6**). The binding energy of control peptide 'FLNKDLEVDGHFVTM' to the binding grooves of the HLA-A*02:01 were found to be -11.4 kcal/mol (PDB ID: 4U6Y). This comparison corrects that T-cell epitope 'LLGTFTWTL' binds with HLA-A*02:01 maintaining almost the same binding energy. The peptide ligand 'LLGTFTWTL' interacts with leu 9, arg 6, thr 8, trp 7 and thr 6 of the HLA-A*02:01 protein molecule. There are electrostatic, hydrophobic and conventional hydrogen bond formed between the ligand and the HLA-A*02:01 protein (**Figure S1**).

The binding energy for B cell epitope 'AELKCFGNTAVAKCNE' to the binding groove of HLA-DRB1*01:01, was estimated to be -14.1 kcal/mol (**Figure 7**). The peptide ligand 'AELKCFGNTAVAKCNE' forms electrostatic bond, conventional hydrogen bond, carbon hydrogen bond, alkyl and pi-alkyl bond with HLA-DRB1*01:01 (**Figure S2**). The key amino acids of the HLA-DRB1*01:01 protein that are situated at the active site and interacts with the ligand are Glu 2, Glu 16, Asn 8, Asn 15, Lys 13, Lys 4, Cys 5, Ala 10, Val 11 and phe 6. Altogether, it can be concluded that both the epitopes showed satisfactory binding affinities with respective HLA allele.

Absorption, Metabolism and Toxicity Profiling of Candidate Peptide Vaccines:

Through "admetSAR" we performed the screening of absorption, metabolism and toxicity profiles of these two epitope-based vaccine candidates and found that both the projected epitopes were predicted to be impermeable to blood brain barrier, non-inhibitor of P-glycoprotein and non-inhibitor of renal organic cation transporter (**Table 3**). Moreover, they were predicted to be

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

non-mutagenic (according to the result of AMES toxicity prediction), non-carcinogenic and non-inhibitor of a variety of CYP450 enzymes. Both the epitopes had acute oral toxicity level III (Category III includes compounds with LD50 values greater than 500mg/kg but less than 5000mg/kg) indicating large amount will be required to cause toxicity, which in turn predicted the non-toxic characteristics of the epitopes. Both the predicted epitopes showed good characteristics with prediction analyses and are predicted to be the potential candidates for *in vitro* and *in vivo* analysis for further consideration.

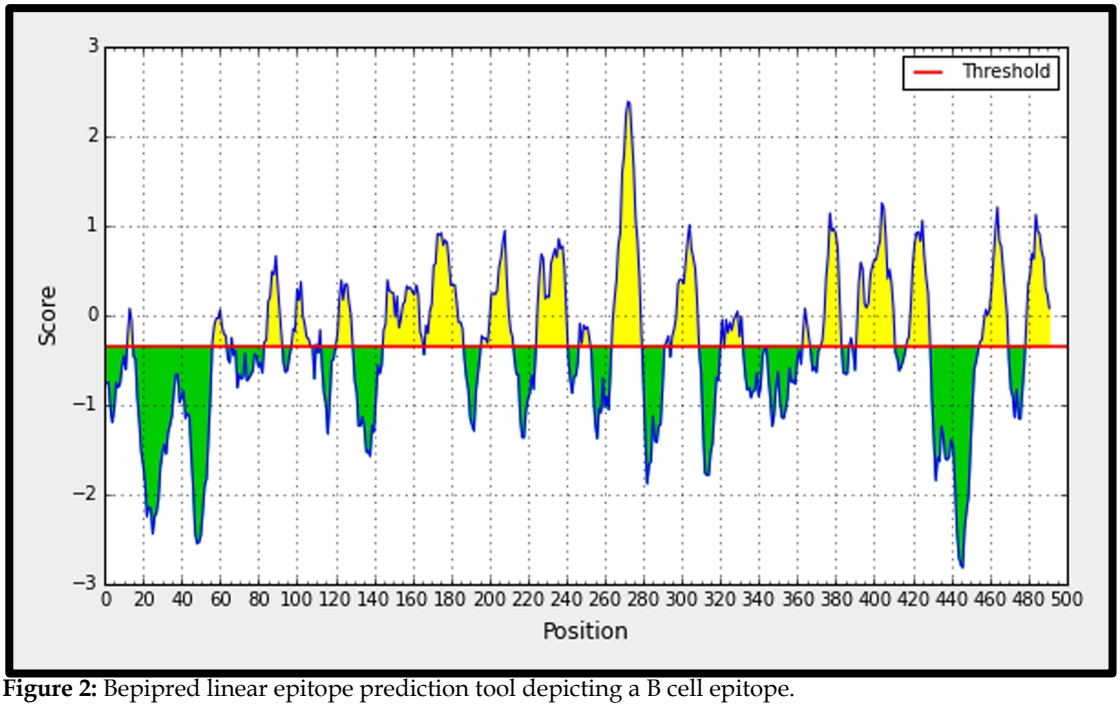

Figure 2: Bepipred linear epitope prediction tool depicting a B cell epitope.

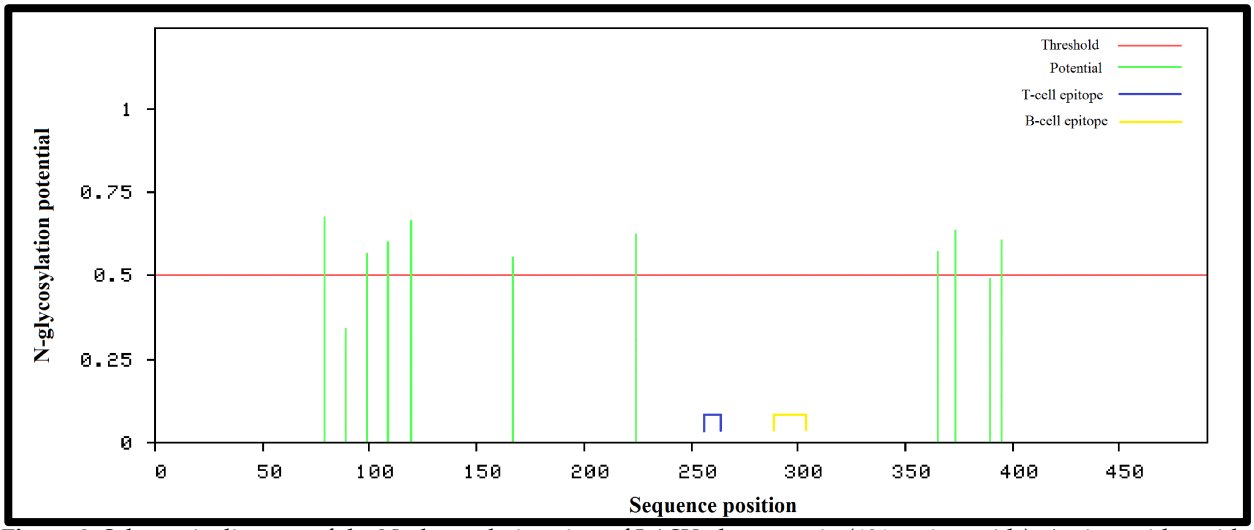

Figure 3: Schematic diagram of the N-glycosylation sites of LASV glycoprotein (491 amino acids). Amino acids residue numbers 79, 89, 99, 109, 119, 167, 224, 365, 373, 390 and 395 are predicted to be N-glycosylation sites. The blue and the yellow area indicate the predicted T-cell candidate epitope 'LLGTFTWTL' region (258-266) and predicted B-cell candidate epitope 'AELKCFGNTAVAKCNE' region (288-303). Both the epitopes do not possess any glycosylation site.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

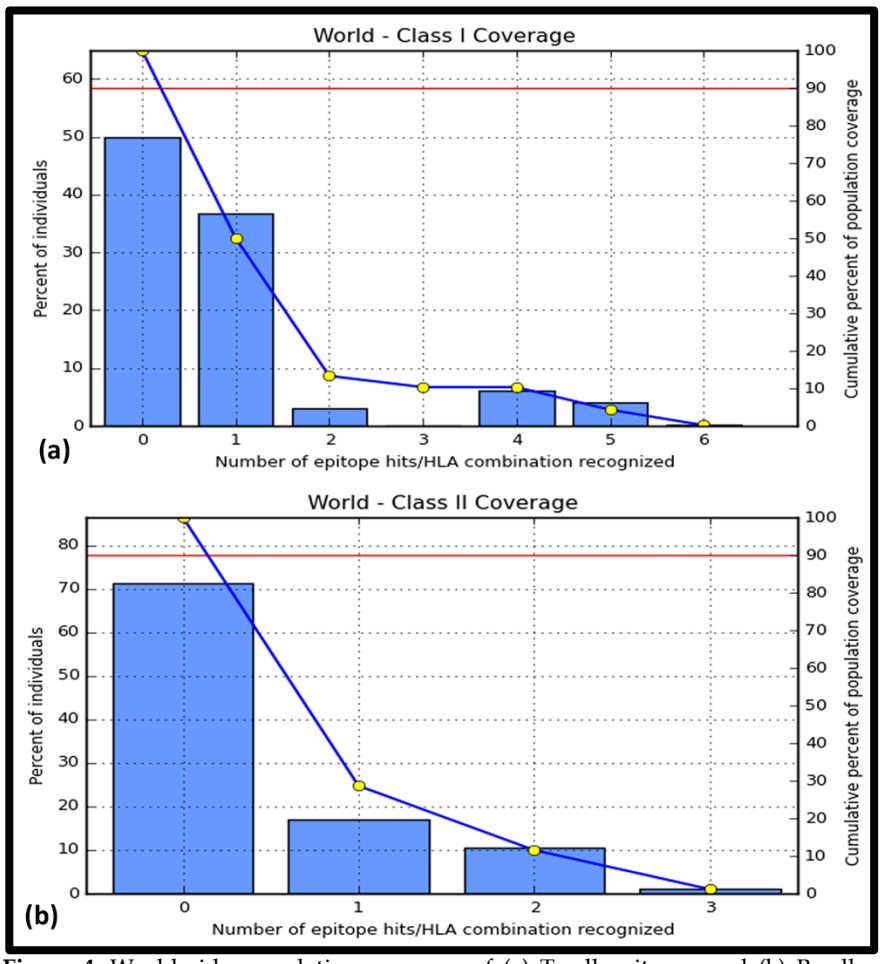

Figure 4: Worldwide population coverage of (a) T-cell epitopes and (b) B-cell epitopes with MHC class I alleles and MHC class II alleles respectively. Bar '0' indicates the percentage of people not elucidating immune response by the epitopes. In the figures (a) and (b), bar '1' indicates the highest interaction showed by the epitopes 'LLGTFTWTL' and 'AELKCFGNTAVAKCNE' respectively.

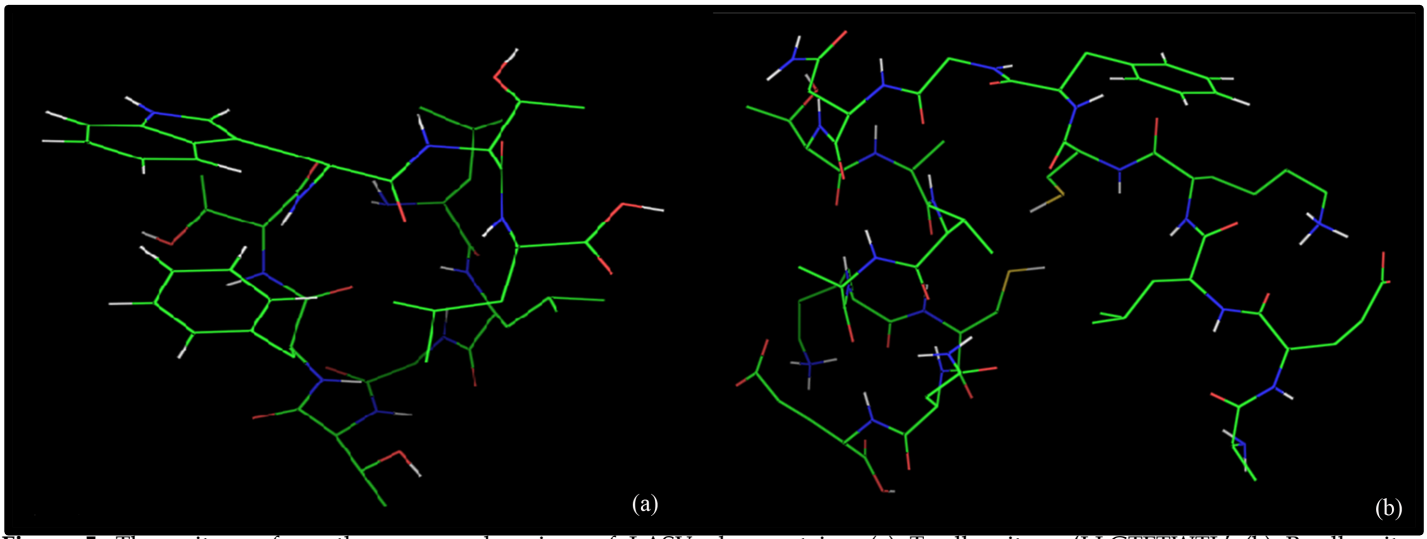

Figure 5: The epitopes from the conserved regions of LASV glycoproteins. (a) T-cell epitope 'LLGTFTWTL' (b) B-cell epitope 'AELKCFGNTAVAKCNE'

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 13(12): 417-429 (2017)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

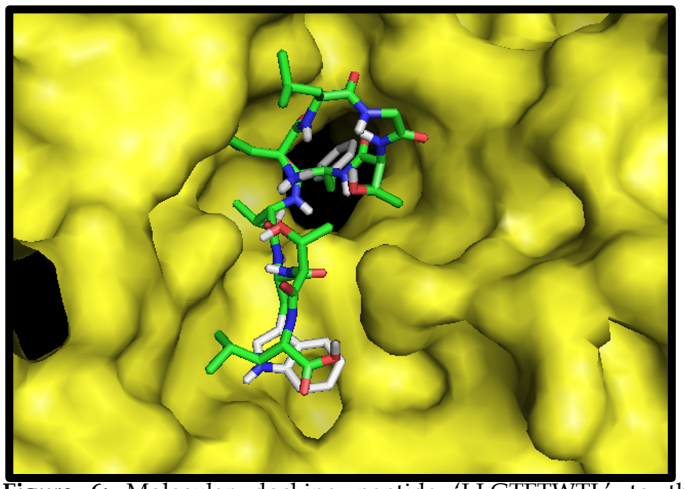

Figure 6: Molecular docking peptide 'LLGTFTWTL' to the binding grooves of MHC class I molecule, HLA-A*02:01. Binding energy was -10.5 kcal/mol.

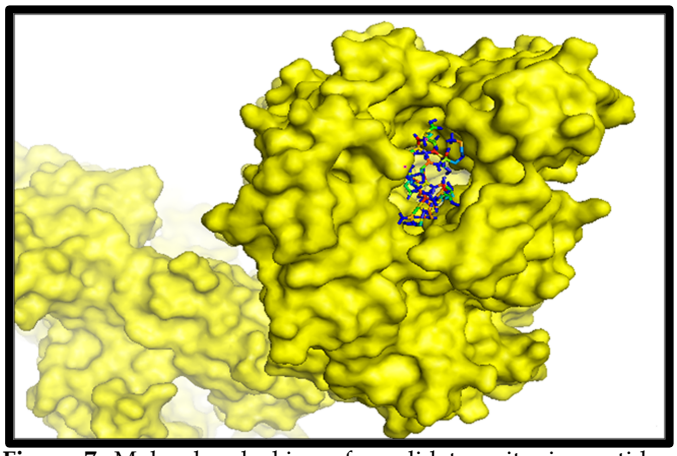

Figure 7: Molecular docking of candidate epitopic peptides to MHC class II molecule, HLA-DRB1*01:01. The bindings of predicted peptide, 'AELKCFGNTAVAKCNE' to the binding grooves of HLA-DRB1*01:01 and binding energy was found to be -14.1 kcal/mol.

Table 1: MHC class I specific nonamer epitopes with antigenicity score (over 0.2) having IC_{50} values less than 250 nM

Sequence	Antigenicity score	HLA class I alleles
LLGTFTWTL	0.41022	HLA-A*02:01, HLA-C*12:03, HLA-C*03:03, HLA-B*15:02
RLLGTFTWT	0.34068	HLA-C*03:03, HLA-C*12:03, HLA-C*14:02, HLA-A*02:06
KHDEEFCDM	0.28311	HLA-C*12:03, HLA-C*05:01, HLA-C*14:02, HLA-C*07:02
RWMLIEAEL	0.24721	HLA-C*12:03, HLA-C*03:03, HLA-C*14:02
SQRTRDIYI	0.2373	HLA-C*12:03, HLA-A*30:01
MLRLFDFNK	0.23554	LA-C*12:03, HLA-A*30:01, HLA-C*12:03, HLA-A*03:01, HLA-A*31:01, HLA-A*11:01, HLA*68:01

Table 2: MHC class II specific 16 mer B-cell linear epitopes with alleles, ABCpred score, antigenicity score and hydrophilicity score

Sequence	MHC II HLA	ABCpred Score	Antigenicity (IEDB)	Hydrophilicity (IEDB)
AELKCFGNTAVAKCNE	HLA-DRB1*07:01, HLA DRB1*01:01, HLA-DRB1*04:04, HLA-DQA1*05:01/DQB1*03:01	0.95	1.04	2.431
AEAQMSIQLINKAVNA	HLA-DRB4*01:01, HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DRB1*11:01	0.82	1.025	1.331
YKGVYELQTLELNMET	HLA-DRB1*01:01, HLA-DRB1*04:05, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*03:01/DPB1*04:02	0.55	1.015	1.181

Table 3: ADMET predicted profile classification for predicted T cell epitope 'LLGTFTWTL' and B cell epitope 'AELKCFGNTAVAKCNE'

Absorption	**Predicted T cell epitope 'LLGTFTWTL'**		**Predicted B cell epitope 'AELKCFGNTAVAKCNE'**	
Model	**Result**	**Probability**	**Result**	**Probability**
Blood-Brain Barrier	BBB-	0.8154	BBB-	0.9273
P-glycoprotein Inhibitor	Non-inhibitor	0.8743	Non-inhibitor	0.9490
Renal Organic Cation Transporter	Non-inhibitor	0.9127	Non-inhibitor	0.9444
Metabolism	**Predicted T cell epitope 'LLGTFTWTL'**		**Predicted B cell epitope 'AELKCFGNTAVAKCNE'**	
Model	**Result**	**Probability**	**Result**	**Probability**
CYP450 1A2 Inhibitor	Non-inhibitor	0.8878	Non-inhibitor	0.8708
CYP450 2C9 Inhibitor	Non-inhibitor	0.8367	Non-inhibitor	0.8816
CYP450 2D6 Inhibitor	Non-inhibitor	0.9006	Non-inhibitor	0.8891
CYP450 2C19 Inhibitor	Non-inhibitor	0.8146	Non-inhibitor	0.8031
CYP450 3A4 Inhibitor	Non-inhibitor	0.9172	Non-inhibitor	0.8443
Toxicity	**Predicted T cell epitope 'LLGTFTWTL'**		**Predicted B cell epitope 'AELKCFGNTAVAKCNE'**	
Model	**Result**	**Probability**	**Result**	**Probability**
AMES Toxicity	Non AMES toxic	0.8561	Non AMES toxic	0.7732
Carcinogens	Non-carcinogens	0.8709	Non-carcinogens	0.8836
Acute Oral Toxicity	III	0.5729	III	0.6490

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

Discussion:

LASV is an endemic virus. It holds a catastrophic potential to destroy the medical scenario in several regions of the world. So, developing a vaccine as a prevention method against LASV is important. Therefore, it is of interest to investigate potential vaccine candidates from LASV glycoprotein. Glycoprotein remains in the outer membrane portion of the virus and it triggers the initial action for viral infection or entry into the host. Vaccine development using glycoprotein targets is highly viable in the context of virus infection **[29]**. A pipeline for developing envelope glycoprotein epitope based vaccine is found to be highly efficient against LASV infection.

A previous study (Verma *et al.* 2015) was completed using glycoprotein (Accession Number NP_694870.1) without considering other isolates specific to LASV endemic regions in different time frame **[30]**. The available 12 LASV glycoprotein sequences from different isolates at different time frame in endemic regions were retrieved for this study and conserved regions among them were identified and illustrated. Viruses tend to mutate faster and therefore, it is important that the candidate epitopes should be in the highly conserved region so that the vaccine designed remain active and provide protection for longer period and against a longer range of virus strains. This, we believe, would generate more acceptable epitope(s) that should be effective universally. Moreover, the epitopes are also needed to be outside the membrane as immune cells activity is modulated by the outer membrane portion as they encounter with the virus during the initial period of infection. Thus, the conserved sequences that fulfill the outer membrane characteristics were used for further analyses that were neglected in the earlier study **[30]**.

Our interest in designing B cell epitopes is because of its role to induce the production of antibodies synthesized by B cells and mediates its effector functions **[31]**. However, over time, humoral response from memory B cells can easily be overcome by surge of antigens whereas cell mediated immunity (T cell immunity) often elicits lasting immunity **[32]**. Cytotoxic T lymphocytes (CTL) restrict the spread of pathogens (like virus) by recognizing and killing infected cells and secreting specific antiviral cytokines, which in turn elicit strong immune response **[33]**. Hence, T-cell epitopes were predicted for the design.

NetCTL tool was used to predict cytotoxic T cell epitopes as it integrates prediction of peptide MHC class I binding, proteasomal C terminal cleavage, TAP transport efficiency. The best 66-epitope candidates were chosen based on a combined score. While projecting the B-cell epitopes, ABCpred server was used and it was found that 29 epitopes cross the threshold level of 0.51. Moreover, BepiPred was also used to predict B cell epitopes with other tools so that everybody can compare the results showed by different prediction tools. This study using different tools will make the study more concrete and increase the probability of epitope as a favorable candidate for further wet-lab verifications.

Another important concern with glycoprotein is to elicit immunity against protein antigen attached with sugar moiety. Protein glycosylation have crucial influence on uptake and proteolytic processing of protein antigens by sterically blocking the action of proteases **[34]**. This in turn can affect MHC presentation and subsequent immune response **[35]**. For example, gp120 subunit of HIV envelope glycoprotein is heavily N-glycosylated which facilitates viral escape from the host immune system by constraining proteolytic processing of the protein antigen required for antigen presentation and cytotoxic T-cell priming **[36]**. Moreover, the N-glycans can also block access of neutralizing antibodies to critical epitopes **[35]**. Thus, protein sequences were checked for glycosylation site. Only the non-glycosylated epitopes (51 out of 66 T cell epitopes and 22 out of 29 B-cell epitopes) were taken for further analyses. This approach was not considered in previous reports.

Interaction analysis between different MHC supertypes/HLA alleles and epitopes in IEDB, SMM method was used among others (like NetMHCpan and CombLib) in this study. This is because both input and output are quantitative in these tools. The output is easy to interpret and specific computational strategies are built to handle experimental noise **[21]**.

Predicted T cell epitope ('LLGTFTWTL') showed adequate MHC I antigenicity score (0.41022). It is also found to interact with MHC-I alleles with IC_{50} score of less than 250 nM having high population coverage of 38%. Moreover, this T cell epitope bound to the binding groove of HLA-A*02:01 with a binding energy of -10.5 kcal/mol similar to the control peptide's binding affinity (-11.4 kcal/mol).

However, the B cell epitope (288–303), 'AELKCFGNTAVAKCNE' showed adequate score like ABCpred Score (0.95), antigenicity score (1.04) and hydrophilicity score (2.431). It also showed dominant interactions with MHC II alleles with IC_{50} score of less than 250 nM having high population coverage of 18%. The binding energy for the epitope at the binding groove of HLA-DRB1*01:01, was estimated to be -14.1 kcal/mol. Meulen *et al.* (2004) reported a highly conserved region having an overlap of two CD4+T-cell epitopes defined by amino acids (289–301) and (282–294) with this epitope (288-303). The epitope between residues 289–301 is 100% conserved in known arena viruses implying its evolutionary and functional importance **[37]**. Thus, the significance of the epitope for further *in vivo* and *in vitro* consideration is established.

Conclusion:

Identification of epitopes in LASV is non-trivial towards the development of a peptide based vaccine formulation. We report a T-cell epitope 'LLGTFTWTL' and a B-cell epitope 'AELKCFGNTAVAKCNE' with predicted immunogenicity for further *in vivo* and *in vitro* consideration.

Conflict of Interest:

The authors declare no conflict of interests.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

Acknowledgement:
The authors sincerely thank Mr. M. Sadman sakib, IMPRS Neurosciences doctoral candidate, AG Fischer, Deutsches Zentrum für Neurodegenerative Erkrankungen e. V. (DZNE) in der Helmholtz-Gemeinschaft, Germany for his valuable comments during preparation of the paper.

References:

[1] Meyer B *et al.* Curr Top Microbiol Immunol. 2002 **262**:139 [PMID: 11987804]
[2] Brosh-Nissimov T. Disaster Mil Med. 2016 **2**:8 [PMID: 28265442]
[3] Ogbu O *et al.* J Vector Borne Dis. 2007 **44**:1 [PMID: 17378212]
[4] Okokhere PO *et al.* J Med Case Rep. 2009 **3**:36 [PMID: 19178735]
[5] Fisher-Hoch S *et al.* Proc Natl Acad Sci U S A 1989 **86**:317 [PMID: 2911575]
[6] McCormick J B *et al.* J Infect Dis. 1987 **155**:445 [PMID: 3805772]
[7] Yun N E *et al.* Viruses 2012 **4**:2031 [PMID: 23202452]
[8] Asogun DA *et al.* PLoS Negl Trop Dis. 2012 **6**:e1839 [PMID: 23029594]
[9] Bloch A Bull World Health Organ. 1978 **56**:811 [PMID: 310723]
[10] Djavani M *et al.* Virology 1997 **235**:414 [PMID: 9281522]
[11] Cao W *et al.* Science 1998 **282**:2079 [PMID: 9851928]
[12] Fisher-Hoch S P *et al.* Expert Rev Vaccines. 2004 **3**:189 [PMID: 15056044]
[13] Fisher-Hoch S *et al.* J Virol. 2000 **74**:6777 [PMID: 10888616]
[14] Sinha J *et al.* A Text Book of Immunology 2006.
[15] Yasmin T *et al.* Scand J Immunol. 2016 **83**:321 [PMID: 26939891]
[16] Banerjee N *et al.* Virus Disease. 2016 **27**:1 [PMID: 26925438]
[17] Krogh A *et al.* J Mol Biol. 2001 **305**:567 [PMID: 11152613]
[18] Larsen M V *et al.* BMC bioinformatics. 2007 **8**:424 [PMID: 17973982]
[19] Saha S *et al.* Proteins. 2006 **65**:40 [PMID: 16894596]
[20] Larsen JE *et al.* Immunome Res. 2006 **2**:2 [PMID: 16635264]
[21] Peters B *et al.* BMC bioinformatics. 2005 **6**:132 [PMID: 15927070]
[22] Calis JJ *et al.* PLoS Comput Biol. 2013 **9**:e1003266 [PMID: 24204222]
[23] Kolaskar A *et al.* FEBS Lett. 1990 **276**:172 [PMID: 1702393]
[24] Parker J *et al.* Biochemistry. 1986 **25**:5425 [PMID: 2430611]
[25] Nielsen M *et al.* BMC bioinformatics. 2007 **8**:238 [PMID: 17608956]
[26] Bui H-H *et al.* BMC bioinformatics. 2006 **7**:153 [PMID: 16545123]
[27] Maupetit J *et al.* Nucleic Acids Res. 2009 **37**:W498 [PMID: 19433514]
[28] Trott O *et al.* J Comput Chem. 2010 **31**:455 [PMID: 19499576]
[29] Graham BS Immunol Rev. 2013 **255**: 230 [PMCID: 3821995]
[30] Verma SK *et al.* Adv Biomed Res. 2015 **4**: 201[PMCID: 4620608]
[31] Cooper NR *et al.* Invest Dermatol. 1984 **83**:121 [PMID: 6376646]
[32] Bacchetta R *et al.* Autoimmun Rev. 2005 **4**:491 [PMID: 16214084]
[33] Garcia, KC *et al.* Annu Rev Immunol. 1999 **17**:369 [PMID: 10358763]
[34] Hanisch FG *et al.* Curr Protein Pept Sci. 2006 **7**:307[PMID: 16918445]
[35] Wolfert MA *et al.* Nat Chem Biol. 2013 **9**: 776 [PMCID: 3966069]
[36] Li H *et al.* J Immunol. 2009 **182**:6369 [PMID: 19414790]
[37] ter Meulen, J *et al.* Virology. 2004 **321**:134 [PMID: 15033572]

Edited by P Kangueane
Citation: **Faisal *et al.*** Bioinformation 13(12): 417-429 (2017)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

Supplementary Material:

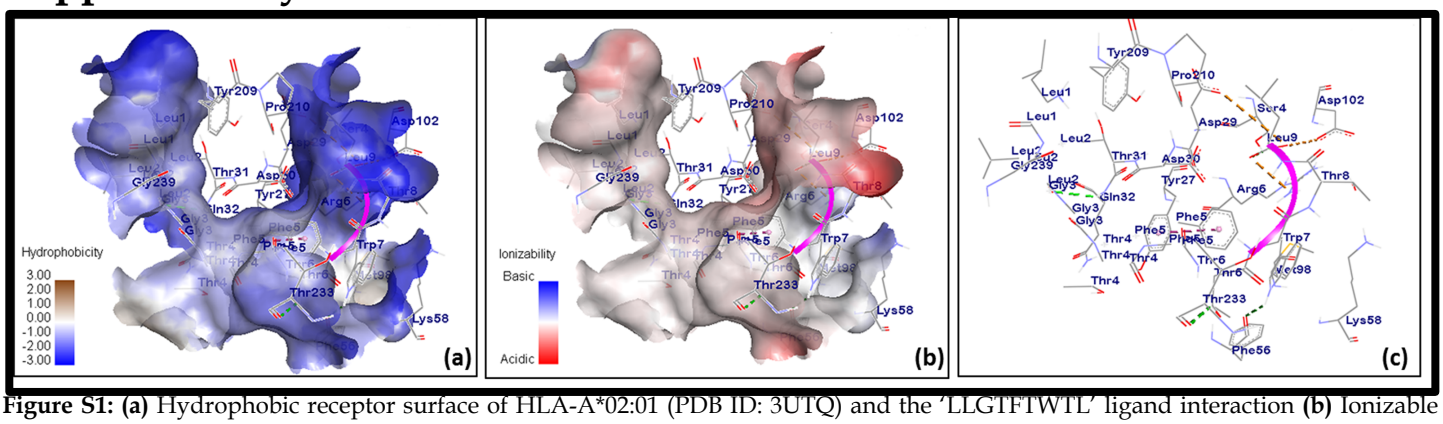

Figure S1: (a) Hydrophobic receptor surface of HLA-A*02:01 (PDB ID: 3UTQ) and the 'LLGTFTWTL' ligand interaction **(b)** Ionizable receptor surface of HLA-A*02:01 (PDB ID: 3UTQ) and the interaction of 'LLGTFTWTL' ligand and **(c)** Non-bond interaction between HLA-A*02:01 (PDB ID: 3UTQ) and 'LLGTFTWTL' ligand.

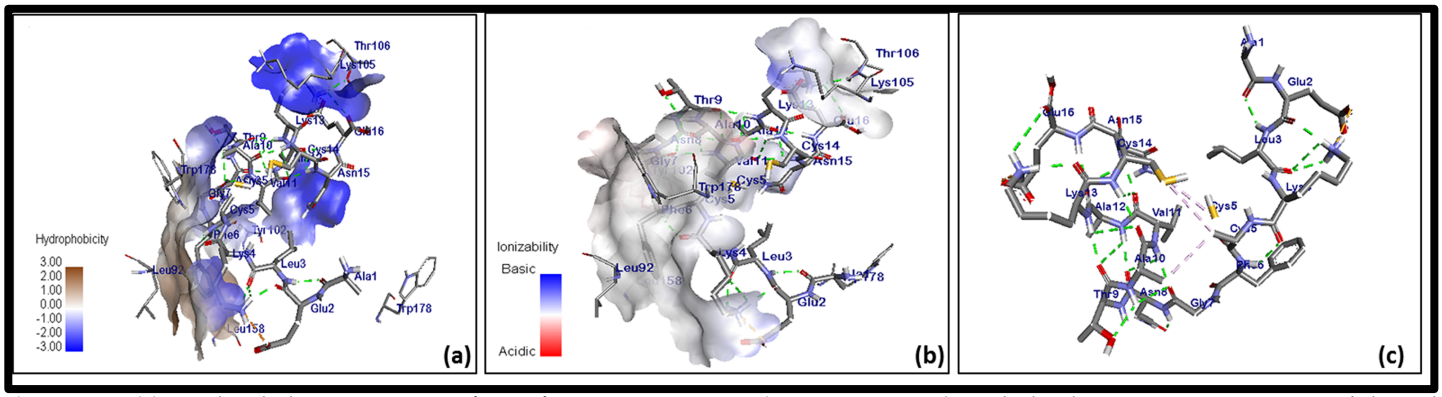

Figure S2: (a) Hydrophobic receptor surface of HLA-DRB1*01:01 (PDB ID: 1AQD) and the 'AELKCFGNTAVAKCNE' ligand interaction **(b)** Ionizable receptor surface of HLA-DRB1*01:01 (PDB ID: 1AQD) and the interaction of 'AELKCFGNTAVAKCNE' ligand **(c)** Non-bonded interaction between HLA-DRB1*01:01 (PDB ID: 1AQD) and the 'AELKCFGNTAVAKCNE' ligand.

Table S1: Glycoprotein sequences from different strains of LASV at different time frame in different endemic regions

LASV glycoprotein sequences	Accession number
LASVGP_1987	AAA46283.1
LASVGP_1989	AAA46286.1
LASVGP_Josiah_Sep2002	NP_694870.1
LASVGP_Josiah_Dec2010	ADY11068.1
LASVGP_recombinant_Josiah_Dec2010	ADY11070.1
LASVGP_Josiah_Aug2011_Sierra Leone	AEY85213.1
LASVGP_Josiah_Jan2017_Sierra Leone	APT69664.1
LASVGP_Josiah_Jan2017_Sierra Leone	APT69609.1
LASVGP_Jan2017_Guinea	APT69607.1
LASVGP_Josiah_Jan2017_Sierra Leone	APT69605.1
LASVGP_Josiah_Jan2017_Sierra Leone	APT69592.1
LASVGP_Jan2017_Guinea	APT69572.1

Table S2: Transmembrane topology analysis of LASV glycoprotein conserved regions using TMHMM v2.0 server

Conserved regions	Nature
1-27	Inner membrane + Transmembrane
29-37	Transmembrane
59-131	Outer membrane
133-173	Outer membrane
175-220	Outer membrane
222-375	Outer membrane
380-409	Outer membrane
411-427	Outer membrane
428-461	Inner + Transmembrane
463-478	Inner membrane

Table S3: Results from non-glycosylated T-cell epitope prediction with MHC class I binding score, processing score, proteasomal cleavage score, TAP score interaction along with super-type information is shown.

Sequence	MHC super-type	MHC score	processing score	proteasomal cleavage score	TAP score	Total score
ITEMLQKEY	A1	0.5415	2.2993	0.9549	2.818	2.5834
	B62	0.2637	0.5234	0.9549	2.818	0.8076
LSIPNFNQY	A1	0.3242	1.3764	0.9728	3.015	1.6731

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 13(12): 417-429 (2017)

	A26	0.3454	0.927	0.9728	3.015	1.2237
	B58	0.4724	1.0647	0.9728	3.015	1.3614
	B62	0.4597	0.9126	0.9728	3.015	1.2093
LSQRTRDIY	A1	0.2816	1.1957	0.3009	2.985	1.3901
	B62	0.3411	0.6772	0.3009	2.985	0.8716
NWDCIMTSY	A1	0.2562	1.0876	0.8691	3.002	1.368
FSRPSPIGY	A1	0.2354	0.9996	0.7407	2.87	1.2542
	B58	0.3128	0.705	0.7407	2.87	0.9596
	B62	0.5028	0.9981	0.7407	2.87	1.2527
LRDIMGIPY	A1	0.214	0.9087	0.9175	2.936	1.1931
	B27	0.2951	0.7845	0.9175	2.936	1.0689
MRMAWGGSY	A1	0.1321	0.5611	0.8509	3.138	0.8456
	B27	0.5452	1.4493	0.8509	3.138	1.7339
	B62	0.4623	0.9178	0.8509	3.138	1.2023
SLYKGVYEL	A2	0.7619	1.1357	0.9778	1.224	1.3435
	B39	0.3597	1.1517	0.9778	1.224	1.3596
	B62	0.2898	0.5752	0.9778	1.224	0.7831
LLGTFTWTL	A2	0.7622	1.1362	0.9783	0.947	1.3303
RMAWGGSYI	A2	0.6266	0.934	0.9571	0.953	1.1252
	B62	0.4864	0.9656	0.9571	0.953	1.1568
RLFDFNKQA	A2	0.5849	0.8719	0.8557	-0.112	0.9946
CIMTSYQYL	A2	0.5237	0.7806	0.1666	0.968	0.854
IISTFHLSI	A2	0.4681	0.6978	0.6683	0.736	0.348
ALINDQLIM	A2	0.4559	0.6796	0.7948	0.678	0.8327
	A26	0.2836	0.7611	0.7948	0.678	0.9143
	B62	0.3129	0.6213	0.7948	0.678	0.7744
MAWGGSYIA	A2	0.4616	0.688	0.8474	-0.132	0.8086
RLLGTFTWT	A2	0.4924	0.734	0.5216	-0.292	0.7977
MTMPLSCTK	A3	0.6468	1.2174	0.9195	0.696	1.3901
MLRLFDFNK	A3	0.5932	1.1164	0.8549	0.441	1.2667
LINDQLIMK	A3	0.5415	1.0191	0.7773	0.657	1.1686
QMSIQLINK	A3	0.5685	1.0699	0.3017	0.749	1.1526
RTRDIYISR	A3	0.4475	0.8422	0.6413	1.707	1.0238
MLQKEYMER	A3	0.3859	0.7262	0.9385	1.352	0.9346
LMSIISTFH	A3	0.424	0.798	0.0903	-0.26	0.7986
QYEAMSCDF	A24	0.577	1.2286	0.6609	2.809	1.4682
VYELQTLEL	A24	0.5585	1.1893	0.9583	1.107	1.3877
	B39	0.2581	0.8262	0.9583	1.107	1.0246
RWMLIEAEL	A24	0.9396	1.1489	0.9371	1.52	1.3655
TFHLSIPNF	A24	0.5158	1.0982	0.4659	2.765	1.3064
GYCLTRWML	A24	0.4916	1.0467	0.5185	1.019	1.1755
EFCDMLRLF	A24	0.4277	0.9107	0.1669	2.34	1.0527
	A26	0.4232	1.1357	0.1669	2.34	1.2777
AWGGSYIAL	A24	0.3554	0.7567	0.8072	1.218	0.9387
EGKDTPGGY	A26	0.5752	1.5435	0.9526	2.5060	1.8117
DCIMTSYQY	A26	0.5009	1.3442	0.2989	0.185	1.3983
TTWEDHCQF	A26	0.1784	0.4788	0.9761	2.805	0.7654
SPIGYLGLL	B58	0.5342	1.2039	0.9761	2.805	1.4906
ISRRLLGTF	B8	0.3758	1.2834	0.8407	-0.075	1.4058
LINKAVNAL	B7	0.3062	0.5907	0.782	0.944	0.7552
	B8	0.2316	0.7908	0.374	2.595	0.9766
YISRRLLGT	B8	0.2604	0.8894	0.0430	-0.592	0.8662
RRLLGTFTW	B58	0.3064	0.6906	0.9146	1.0960	0.8826
SIQLINKAV	B8	0.1972	0.6733	0.4906	0.5	0.7719
	B27	0.5307	1.4109	0.9784	1.409	1.628
QRLKAEAQM	B27	0.5114	1.3594	0.8823	0.551	1.5192
KHDEEFCDM	B44	0.2765	0.6853	0.1647	0.944	0.7572
TLELNMETL	B44	0.6825	1.6912	0.3874	0.745	1.7866
HDEEFCDML	B39	0.256	0.8197	0.6187	0.513	0.9381
	B39	0.21	0.6725	0.9134	0.76	0.8475
AEAQMSIQL	B39	0.2054	0.6577	0.3779	1.081	0.7684
EEFCDMLRL	B44	0.747	1.8512	0.3779	1.081	1.9619
RDIYISRRL	B44	0.3093	0.7664	0.4343	0.97	0.88
TEMLQKEYM	B44	0.2782	0.6895	0.609	0.222	0.792
VANGVLQTF	B62	0.3842	0.7628	0.9593	0.904	0.9519

MSIISTFHL	B58	0.6342	1.4293	0.9397	1.159	1.6282
	B58	0.5985	1.3489	0.9048	2.565	1.6128
LQTFMRMAW	B62	0.4276	0.8488	0.2256	2.72	1.0186
SQRTRDIYI	B62	0.3315	0.6581	0.5659	0.76	0.781

Table S4: HLA class I specific T-cell epitopes with antigenicity score having IC_{50} < 250 nM.

Sequence	Antigenicity score	HLA class I alleles
LLGTFTWTL	0.41022	HLA-A*02:01, HLA-C*12:03, HLA-C*03:03, HLA-B*15:02
RLLGTFTWT	0.34068	HLA-C*03:03, HLA-C*12:03, HLA-C*14:02, HLA-A*02:06
KHDEEFCDM	0.28311	HLA-C*12:03, HLA-C*05:01, HLA-C*14:02, HLA-C*07:02
RWMLIEAEL	0.24721	HLA-C*12:03, HLA-C*03:03, HLA-C*14:02
SQRTRDIYI	0.2373	HLA-C*12:03, HLA-A*30:01
GYCLTRWML	0.1422	HLA-C*14:02, HLA-C*07:02, HLA-C*12:03, HLA-B*15:02, HLA-C*03:03, HLA-A*23:01
HDEEFCDML	0.11262	HLA-C*12:03, HLA-C*07:02, HLA-B*15:02, HLA-C*05:01
TTWEDHCQF	0.11152	HLA-C*14:02, HLA-A*32:01, HLA-C*03:03, HLA-C*12:03
LRDIMGIPY	0.10786	HLA-C*12:03, HLA-C*14:02, HLA-C*08:02
MLRLFDFNK	0.23554	HLA-C*12:03, HLA-A*30:01, HLA-C*12:03, HLA-A*03:01, HLA-A*31:01, HLA-A*11:01, HLA-A*68:01
MRMAWGGSY	0.16191	HLA-C*12:03, HLA-C*06:02, HLA-C*07:01, HLA-C*14:02, HLA-B*15:02, HLA-B*27:05, HLA-A*30:02, HLA-A*29:02
RMAWGGSYI	0.15871	HLA-C*03:03, HLA-A*02:01, HLA-C*12:03, HLA-C*14:02, HLA-B*15:02, HLA-A*02:06, HLA-A*32:01
LMSIISTFH	0.15525	HLA-C*03:03, HLA-C*12:03, HLA-C*14:02
RRLLGTFTW	0.17626	HLA-C*12:03, HLA-A*32:01, HLA-B*58:01, HLA-C*03:03, HLA-B*27:05, HLA-C*07:02
MSIISTFHL	0.17026	HLA-B*15:02, HLA-B*58:01, HLA-A*68:02, HLA-C*12:03, HLA-C*03:03, HLA-A*02:06
LSQRTRDIY	0.19748	HLA-C*12:03, HLA-C*03:03, HLA-B*15:02, HLA-C*14:02, HLA-A*30:02
RTRDIYISR	0.1809	HLA-A*31:01, HLA-A*30:01, HLA-C*12:03, HLA-C*14:02, HLA-A*68:01

Table S5: Results from non-glycosylated B-cell epitope prediction using the ABCpred server v2.0. Antigenicity and hydrophilicity were calculated using the IEDB antigen sequence properties tool.

Rank	Sequence	Starting number	ABCpred Score	Antigenicity (IEDB)	Hydrophilicity (IEDB)
01	AELKCFGNTAVAKCNE	288	0.95	1.04	2.431
02	DQLIMKNHLRDIMGIP	347	0.90	0.999	0.125
03	LSDSEGKDTPGGYCLT	266	0.90	1.007	3.375
04	TEMLQKEYMERQGKTP	412	0.89	0.946	3.106
05	EEFCDMLRLFDFNKQA	307	0.88	1.006	0.738
06	WGGSYIALDSGRGNWD	196	0.87	0.957	1.875
07	HLSIPNFNQYEAMSCD	141	0.87	1.027	1.625
08	LGTFTWTLSDSEGKDT	259	0.83	0.963	2.569
09	TFMRMAWGGSYIALDS	190	0.83	0.984	0.081
10	MSIISTFHLSIPNFNQ	134	0.83	1.025	-0.431
11	AEAQMSIQLINKAVNA	328	0.82	1.025	1.331
12	QRTRDIYISRRLLGTF	247	0.81	1.008	0.619
13	TAVAKCNEKHDEEFCD	296	0.80	1.023	3.95
14	KDTPGGYCLTRWMLIE	272	0.78	1.006	0.656
15	YISRRLLGTFTWTLSD	253	0.76	1.018	-0.25
16	CLTRWMLIEAELKCFG	279	0.75	1.055	-1.106
17	DHCQFSRPSPIGYLGL	229	0.75	1.067	0.925
18	NQYEAMSCDFNGGKIS	148	0.73	0.981	3.006
19	PLSCTKNNSHHYIMVG	83	0.65	1.047	1.519
20	DFNKQAIQRLKAEAQM	317	0.63	0.986	2.131
21	RPSPIGYLGLLSQRTR	235	0.62	1.032	0.931
22	YKGVYELQTLELNMET	62	0.55	1.015	1.181

Table S6: MHC class II specific B-cell epitopes having predicted IC_{50} less than 250 nM is listed.

Sequence	MHC II HLA
AELKCFGNTAVAKCNE	HLA-DRB1*07:01, HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DQA1*05:01/DQB1*03:01
LSDSEGKDTPGGYCLT	Nil
HLSIPNFNQYEAMSCD	HLA-DRB1*01:01
AEAQMSIQLINKAVNA	HLA-DRB4*01:01, HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DRB1*11:01
PLSCTKNNSHHYIMVG	Nil
YKGVYELQTLELNMET	HLA-DRB1*01:01, HLA-DRB1*04:05, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*03:01/DPB1*04:02
TAVAKCNEKHDEEFCD	Nil

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

Table S7: Population coverage of class I MHC specific T-cell epitopes

Population/area	Class I		
	Coverage	Average hit	pc90
Central Africa	18.78%	0.36	0.12
East Africa	23.39%	0.31	0.13
East Asia	38.61%	0.45	0.16
Europe	59.02%	1.15	0.24
North Africa	34.05%	0.69	0.15
North America	49.7%	0.72	0.2
Northeast Asia	26.32%	0.43	0.14
Oceania	27.89%	0.44	0.14
South America	28.08%	0.44	0.14
South Asia	21.92%	0.54	0.13
Southeast Asia	23.07%	0.29	0.13
Southwest Asia	33.69%	0.61	0.15
West Africa	22.14%	0.33	0.13
West Indies	32.19%	0.32	0.15
World	50.04%	0.88	0.2

Table S8: Population coverage of class II MHC specific B-cell epitopes

Population/area	Class II		
	Coverage	Average hit	pc90
Central Africa	15.2%	0.17	0.12
Central America	12.49%	0.14	0.11
East Africa	12.76%	0.17	0.11
East Asia	17.28%	0.28	0.12
Europe	36.78%	0.54	0.16
North Africa	27.63%	0.3	0.14
North America	31.49%	0.46	0.15
Northeast Asia	10.91%	0.14	0.11
Oceania	2.67%	0.03	0.1
South America	13.17%	0.17	0.12
South Asia	36.27%	0.47	0.16
Southeast Asia	7.13%	0.08	0.11
Southwest Asia	14.43%	0.17	0.12
West Africa	12.69%	0.17	0.11
West Indies	26.71%	0.37	0.14
World	28.63%	0.41	0.14