Original Research Article

Immunoinformatics approach identified two highly conserved B and T cell epitopes, LEASKRWAF and DSPLEASKRWAFRTG, for effective vaccine design against Ebola and Marburg Viruses

.

ABSTRACT

Aims: Ebola and Marburg viruses cause fatal hemorrhagic fever in both human and non-human primates. Absence of any licensed vaccine has further deteriorated the problem. In the present study, we aimed to design potential epitope based vaccines against these viruses using computational approaches.
Methodology: By using various bioinformatics tools and databases, we analyzed the conserved glycoprotein sequences of Ebola and Marburg viruses and predicted two potential epitopes which may be used as peptide vaccines.
Results: Using various B-cell and T-cell epitope prediction servers, four highly conserved epitopes were identified. Epitope conservancy analysis showed that "LEASKRWAF" and "DSPLEASKRWAFRTG" epitopes were 100% and 93.62% conserved and the worldwide population coverage of "LEASKRWAF" interacting with MHC class I molecules and "DSPLEASKRWAFRTG" interacting with MHC class II molecules were 78.74% and 75.75% respectively. Immunoinformatics analysis showed that they are highly immunogenic, flexible and accessible to antibody. Molecular docking simulation analysis demonstrated a very significant interaction between epitopes and MHC molecules with lower binding energy. Cytotoxic analysis and ADMET test also supported their potential as vaccine candidates.
Conclusion: In sum, our in silico approach demonstrated that both "LEASKRWAF" and "DSPLEASKRWAFRTG" holds the promise for the development of common vaccine against life threatening Ebola and Marburg viruses.

Keywords: B cell; T cell; Vaccine; Epitope; Ebola and Marburg viruses

1. INTRODUCTION

Ebola virus (EBOV) and Marburg virus (MARV), belong to the family Filoviridae (filoviruses), are among the deadliest human pathogenic viruses which cause the outbreak of viral hemorrhagic fever in Africa with high fatality rate [1, 2]. These viruses can be transmitted between humans and from non-human hosts through contact with infectious bodily fluids [3, 4]. Their natural reservoirs are fruit bats, predominantly the Egyptian fruit bat (Rousettus aegyptiacus), which makes its transmission particularly dangerous [5]. Both viruses are classified as category A pathogens with no licensed vaccine or treatment available for human use and are handled in maximum containment laboratories [2]. The genus Ebolavirus is composed of five species such as, Bundibugyo virus (BDBV; species Bundibugyo ebolavirus); Ebola virus (EBOV; species Zaire ebolavirus); Sudan virus (SUDV; species Sudan ebolavirus); Tai Forest virus (TAFV; species Tai Forest ebolavirus) and Reston virus (RESTV; species Reston ebolavirus), with the newly discovered currently unclassified Bombali virus (BOMV; species Bombali ebolavirus) [6]. In contrast, the genus Marburgvirus has only one species, the Marburg marburgvirus, with two known strains Marburg virus (MARV) and Ravn virus (RAVV), which has approximately 20% divergent at the amino acid level [2].

Filoviruses are filamentous in appearance and have non-segmented single strand negative sense RNA genome which is approximately 19 kb in length [17]. The viral RNA genome encode seven proteins which are translated from a single monocistronic mRNA, such as nucleoprotein [18], major (VP40) and minor (VP24) matrix proteins, RNA-dependent RNA polymerase (L), polymerase cofactor (VP35), transcription activator (VP30), and a glycoprotein (GP) [19, 20]. The genome is tightly associated with the nucleoprotein [18] and viral protein 30 (VP30), which along with viral protein 35 (VP35) and the L-polymerase (L) protein form the central nucleocapsid core [20]. The nucleocapsid core is surrounded by a matrix, comprising viral protein 40 (VP40) and viral protein 24 (VP24) and a host-derived lipid envelope composed of anchored glycoprotein (GP) [17]. The MARV VP40 has been known to inhibit protein tyrosine phosphorylation of STAT thereby blocking the Jak-STAT pathway. On the other hand, EBOV VP24 obstructs the interferon induced pathway by preventing nuclear accumulation of phosphorylated STAT1 [21, 22]. VP35 is another protein that impedes interferon production by inhibiting retinoic-acid inducible gene-I (RIG-I)-like receptor (RLR) activity [23, 24]. However, among these proteins, GP is the most promising as it protrudes outward as 7 to 10 nM spikes. Filovirus GP is involved in cell selection and entry by promoting receptor binding and membrane fusion [25, 26] and has the most immunogenic potential, therefore, serves as a possible vaccine candidate [27, 28].

The lethal consequences of Filoviruses become more terrifying due to the absence of any approved vaccine or drug either to induce protective immunity or to control viral infection. Small inhibitor molecules have been developed to inhibit viral entry, but further testing proved the method ineffective in deterring the diseases [29]. The rVSV-ZEBOV vaccine against EBOV was developed in 2003, and was first used in 2016 to immunize patients [30, 31]. The vaccine was successful in some cases, but it exhibited adverse effects in half of the patients, and reports of its 100% efficacy were unsupportable [32]. The passive administration of monoclonal antibodies (mAbs) appeared as a promising treatment option during 2013 to 2016 Western African epidemic [33-38]. Although several monoclonal antibodies based vaccination strategy has been developed recently and undergone clinical study, they are limited to single member of the Ebola virus genus [39, 40]. Recently, several human neutralizing mAb based cocktail immunotherapy has been developed which provide broad protection [41-43]. Another study found complete protection against Ebola and Marburg viruses in two strains of mice using T-cell epigraph vaccine [44]. So far, no universal vaccine has been licensed which can provide protection against all Filoviruses irrespective of their genetic variations.

Nowadays, epitope based vaccine design against lethal viruses through bioinformatics has become popular because of its short study time, increased strength to predict effective epitopes and the availability of ample sequence data. This approach has been validated in various studies to fight diseases such as malaria, human immunodeficiency virus, tuberculosis etc. Conserved epitope prediction by computational biology approaches not only save time, but also reduces the cost associated with the vaccine development process. In the current study, we used various bioinformatics tools to select peptides with high level of conservation and mapped the evolutionary conserved epitopes for entire Filovirus family. We have predicted a potential conserved epitope candidate which may be used to immunize patients against both Ebola and Marburg virus.

2. MATERIAL AND METHODS

The flow chart showing graphical outline of the approaches used for peptide based vaccine design againt Ebola and Marburg virus has been depicted in Figure 1.

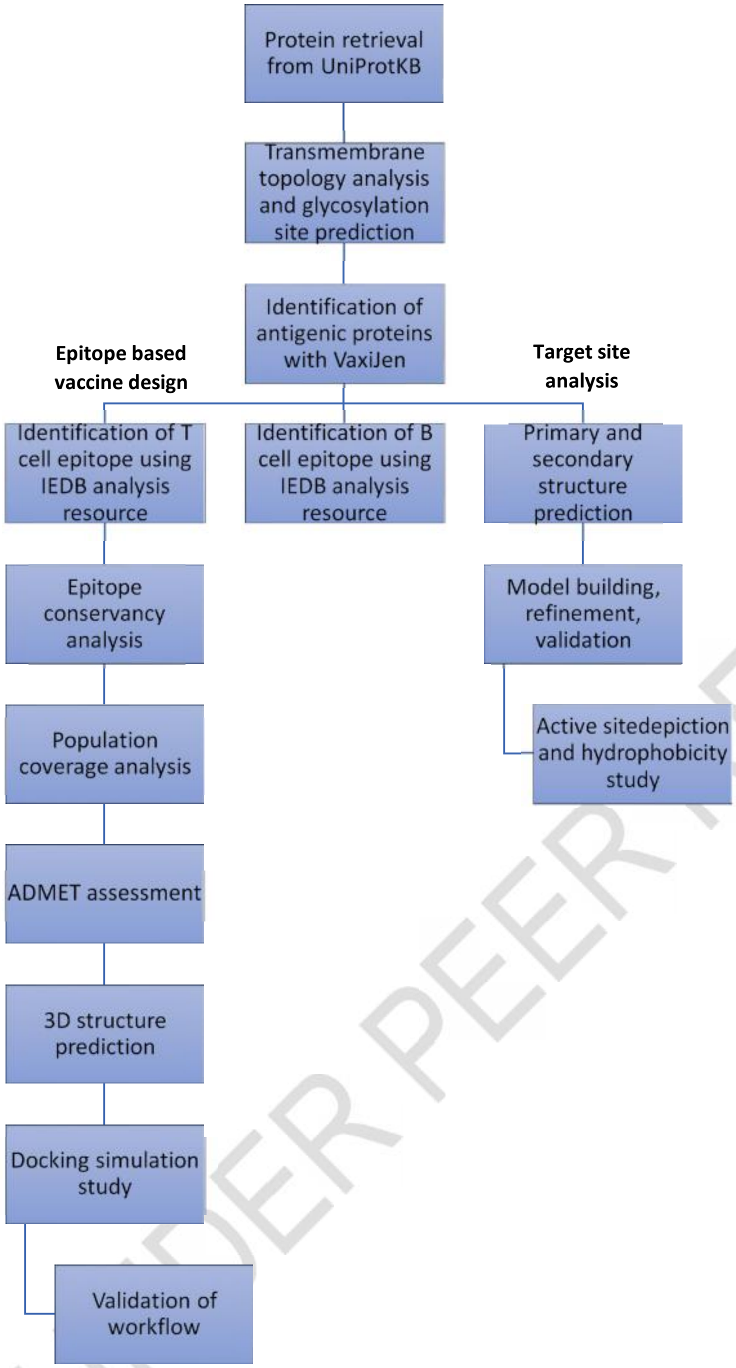


Figure 1. Graphical outline of the peptide based vaccine design againt Ebola and Marburg virus.

2.1. Sequence retrieval and conserved region identification

A total of 47 glycoprotein (GP) sequences of both Marburg virus (30) and Ebola virus (17) were retrieved from UniProtKB database and downloaded in FASTA format. The length of the glycoprotein sequence was 681 amino acids. Mega 7.0 tool was used to determine the conserved sequences through multiple sequence alignment with MUSCLE algorithm, and the results were verified with Jalview [45-47].

2.2. Variability analysis of the glycoprotein

The conserved sequences were fed into the Protein Variability Server (PVS) to determine the absolute site variability using Shannon entropy analysis [48]. Several other variability measures were also computed to calculate the absolute variation in the alignment.

2.3. Transmembrane topology analysis and glycosylation site prediction

As the epitopes need to be in the exposed regions of the protein to yield the best response, they were analyzed using TMHMM v2.0 server to identify the inner, outer and transmembrane helix regions [49]. The protein was then analyzed to identify the glycosylation sites using NetOGlyc 4.0 Server, and the results were verified using NetNGlyc 1.0 Server [50, 51]. The epitopes without glycosylation sites were used in further analyses.

2.4. Prediction of antigenicity

Antigenicity determines the success of a subunit vaccine by inducing an immune response and providing protection from future infections. The conserved sequence was tested using VaxiJen v2.0 server [52], which calculates antigenicity based on physiochemical properties of the protein and is not dependent on sequence alignment.

2.5. Identification of the B cell epitope

B lymphocytes recognize B cell epitopes on viral surface proteins and mount immune response through the differentiation of plasma and memory cells. Plasma cell releases antibody for opsonization, while memory cells retain immunity. IEDB provides different methods to predict linear epitopes from protein sequences using amino acid scales and Hidden Markov Models (HMM) [53]. Bepipred Linear Epitope Prediction, Chou & Fasman Beta-Turn Prediction, Emini Surface Accesibility Prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Angenicity, Parker Hydrophilicity Prediction tools were used to predict the B cell epitopes, and the results were cross-referenced with each other to obtain epitopes that fulfilled all the criteria of a highly immunogenic peptide vaccine and finally verified with ABCpred server [54-58].

2.6. Prediction of epitope conservancy

Prediction of epitope conservancy is important to determine the effectiveness of the vaccine among population. IEDB based epitope conservancy analysis tool was used to calculate the ratio of protein sequences having the epitope at a given identity level [53]. Sequence identity threshold was set at least 80% for calculating the conservancy score.

2.7. Prediction of population coverage

Population coverage is a tool used to calculate the ratio of individual, which can mount immune response to a set of epitopes with fixed MHC molecules. Allelic frequency of the interacting HLA alleles was exploited to predict the population coverage for each epitope [59].

2.8. Identification of T cell epitope and their interaction to MHC class I and MHC class II molecules

T cell epitope is expressed on antigen presenting cell bound with Major Histocompatibility Complex (MHC) to initiate T cell immune response. IEDB analysis resource provides several tools to predict T cell epitope [60-62]. T cell epitopes were identified by NetCTL prediction method which predicts epitopes based on proteosomal processing, TAP transport and MHC binding affinity. Artificial Neural Network (ANN) method was used to determine the half-maximal inhibitory concentration (IC50) values [63, 64]. All the alleles from this site with some extra alleles relevant to this study from external source were used for binding analysis. The length of the peptide was set at 9.0 to predict the epitope with MHC I molecule. T cell epitopes binding to MHC class II molecules were also identified using combinatorial library, SMMalign (Stabilized matrix method) and Sturniolo methods to obtain IC50 values [65].

2.9. Prediction of 3-D structure and Molecular Docking Analysis of HLA and epitopes

The docking analysis was performed using pdb files for HLA obtained from RCSB PDB and pdb files for the epitopes created using PEP-FOLD3 server [66]. The HLA pdb files extracted from RCSB PDB were prepared by removing all unnecessary molecules, adding polar hydrogens and Kollman charges. AutoDock Vina was then used to carry out the docking analysis with 1.00 A° spacing and exhaustiveness = 8 [67]. The output files were then viewed with AutoDock Tools and the conformation with the highest binding affinity at the correct binding site was selected. The non-bond interactions (H-bonds) were then observed between the ligand and the H-bond surface of the receptor in BIOVIA Discovery Studio Visualizer v17 [68].

2.10. ADMET assessment of target peptides

Peptide based subunit vaccine development is promising, but toxicity of the peptide epitopes interferes the success of peptide based therapy. The ADMET profile of the target peptides was determined using the SwissADME tool and the results were verified using admetSAR server [69, 70].

2.11. Validation of the workflow

The entire study was dependent on computational analyses that needed to be verified before a stable conclusion was drawn. The entire workflow was put to the test by using a negative and a positive control. For the negative control, a random 681 amino acid sequence was analyzed using the workflow. In contrast, for the positive control, six linear B-cell epitopes of VP1 protein of coxsackievirus A16 were tested using the protein sequence extracted from NCBI [71].

3. RESULTS AND DISCUSSION

3.1. The envelope glycoprotein is highly conserved in both Ebola and Marburg viruses

The degree of conservancy of specific proteins among various strains or species provides important information about its evolutionary history, structure, function, and immunological properties**.** To determine the degree of conservation, the retrieved sequences were aligned properly and an MSA was carried out with MUSCLE. MSA analysis by MUSCLE revealed that envelope glycoprotein is well conserved in all sequences and the absolute variability computed by PVS suggested 8 highly conserved regions (Figure 2a, 2b and Table 1). These regions were therefore selected for further analysis.

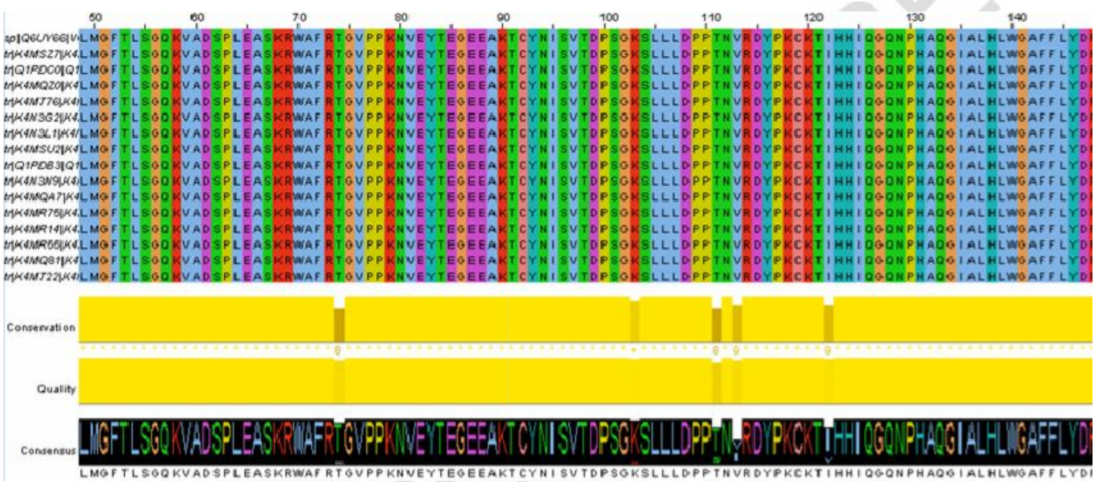


Figure 2.a. Multiple sequence alignment of the retrieved sequences in Jalview. These regions are highly conserved.

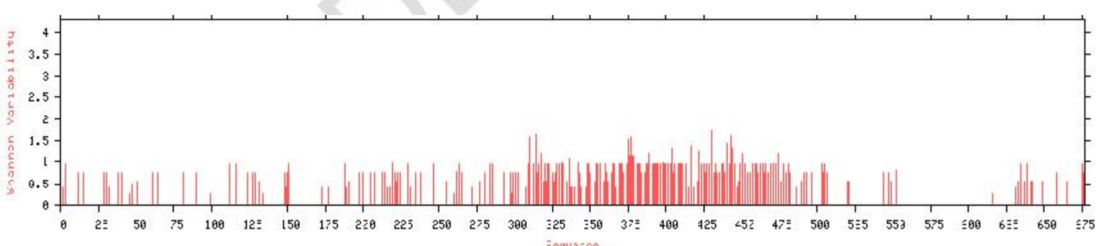


Figure 2.b. Protein variability index of G protein determined by PVS server. The conservancy threshold was 1.0 in this analysis. X axis indicates the amino acid position in sequences and Y axis indicates the Shannon entropy.

Table 1. Transmembrane topology of GP protein analyzed using THMM 2.0 server.

Conserved Regions	Topology
34-73	Outer membrane
75-102	Outer membrane
104-121	Outer membrane

123-157	Outer membrane
159-200	Outer membrane
511-546	Outer membrane
548-595	Outer membrane
597-649	Outer membrane

3.2. The envelope glycoprotein is highly antigenic and has large extracellular stretches

A protein must be antigenic enough to provoke sufficient immune response to be a vaccine candidate. Evaluation of the envelope glycoprotein by the VaxiJen v2.0 server suggested it as a probable antigen with the prediction value of 0.5453. A very large region of the protein (1-649) was purely on the outer membrane, while only two small segments were on the inner membrane (650-672) and transmembrane helix (673-681). The conserved regions were cross-referenced to obtain short stretches that were on the outer membrane (Table 1). The glycosylated regions were excluded from further analysis (Figure 3.).

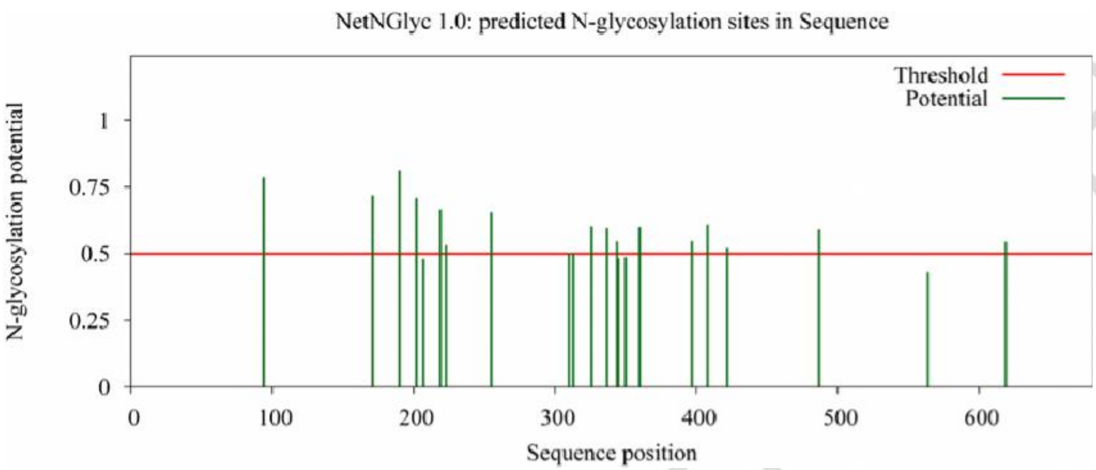


Figure 3. The N-glycosylation sites of GP protein identified using NetNGlyc 1.0 server.

3.3. The highly antigenic B cell epitopes are flexible, hydrophilic and surface accessible

Several B cell epitope prediction software packages are currently used for B cell epitope prediction. Each software provides its own dataset and exploits a specific method for epitope prediction. Hence the predicted epitopes for a given protein differ from one software to another [72, 73], accurate identification of immunogenic regions in a given antigen is complicated, and prediction of false positive epitopes is a common problem [74]. Therefore, we utilized six different software packages for the B cell epitope prediction. ABCpred identified 66 16-mer epitopes with score higher than 0.5. These epitopes were cross-referenced with the results of IEDB linear B cell epitope prediction. The epitopes with higher surface accessibility scores, flexibility scores, hydrophilicity scores, and antigenicity scores were then selected (Figure 4 and Table 2).

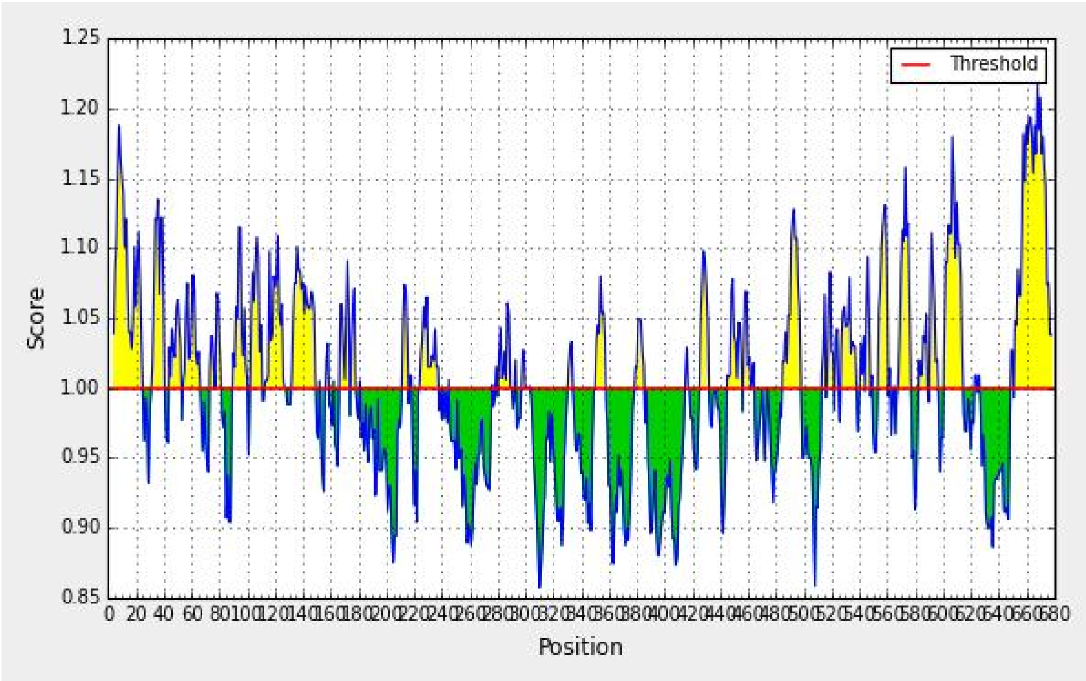


Figure 4. Kolaskar and Tongaonkar antigenicity prediction of the proposed epitope with a threshold value of 1.00. Residues in yellow regions are antigenic in nature.

Table 2. Predicted B-cell linear epitopes with ABCpred score, antigenicity score and hydrophilicity score.

Epitope	**Position**	**ABCpred score**	**Antigenicity (IEDB)**	**Hydrophilicity (IEDB)**
PLEASKRWAFRTGVPP	63-78	0.89	0.98	1.61
GKSLLLDPPTNVRDYP	102-117	0.69	1.05	1.27
LHLWGAFFLYDRIAST	137-152	0.86	1.06	1.44
ASTTMYRGKVFTEGNI	150-165	0.85	0.98	1.73

3.3. The T cell epitopes are bound and processed by MHC molecules

The 9-mer T cell epitopes were cross-referenced with MHC I processing and binding results. Only the epitopes with a total score (proteosomal processing, TAP transport, MHC binding) above 0.5 and an IC50 < 250 nM were selected for further analysis (Table 2). Finally, only 5 epitopes were selected based on the criteria which interacted with several HLA alleles. Following this, T cell epitopes interacting with MHC II molecules were also identified based on MHC II binding results where lower total percentile ranks and IC50 < 500 nM. A total of 5 epitopes, which interacted with several HLA alleles, with similarities to the ones identified before were selected in this case (Table 3 and 4).

Table 3. Predicted epitopes for CD8+ T-cell along with their interacting MHC class I alleles with affinity < 250 nM.

Epitope	Position	MHC class I allele with total score having IC50 values < 250 nM
LEASKRWAF	64-72	HLA-B*18:01(1.05), HLA-B*15:03(.91), HLA-B*41:03(.57), HLA-B*41:04(.37), HLA-B*41:02(.32), HLA-B*44:02(.23), HLA-B*44:27(.23), HLA-B*44:08(.06)
LLLDPPTNV	105-113	HLA-A*02:11(1.09), HLA-A*02:03(.68), HLA-A*02:16(.65), HLA-A*02:50(.58), HLA-A*02:12(.58), HLA-A*02:01(.46), HLA-A*02:02(.38), HLA-A*02:19(.3), HLA-A*02:06(.2)
IALHLWGAF	135-143	HLA-B*15:03(1.23), HLA-B*15:17(.77), HLA-B*15:02(.47), HLA-B*35:01(.41), HLA-A*32:07(.21), HLA-B*15:01(.15)
HLWGAFFLY	138-146	HLA-A*29:02(1.88), HLA-A*80:01(1.35), HLA-B*15:03(.97), HLA-A*32:07(.59), HLA-A*68:23(.56), HLA-A*30:02(.52), HLA-A*32:01(.48), HLA-A*32:15(.28), HLA-B*35:01(.2), HLA-A*03:01(.19), HLA-A*03:02(.14)
TTMYRGKVF	152-160	HLA-B*15:17(1.32), HLA-B*15:03(.8), HLA-C*12:03(.73), HLA-A*26:02(.43), HLA-

		C*14:02(.08)

Table 4. Predicted CD4+ T-cell epitopes along with their interacting MHC class II alleles with affinity (IC50) < 500 nM and respective total scores.

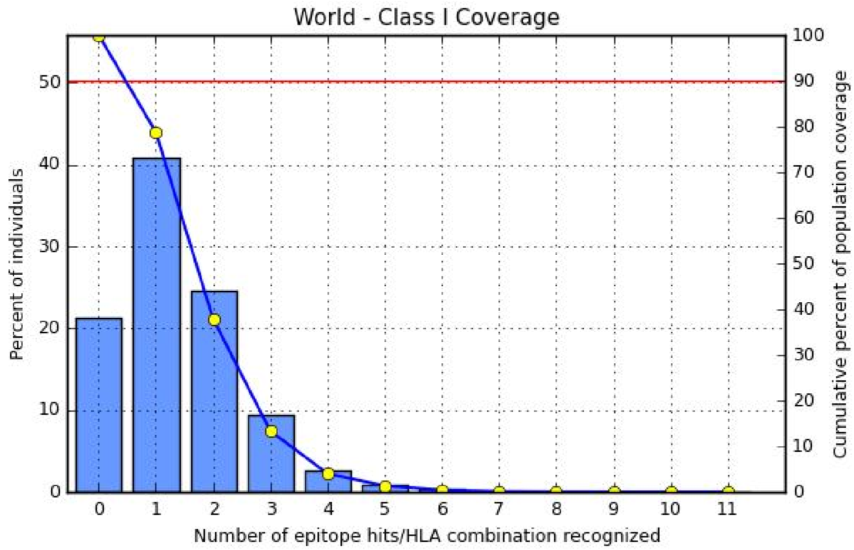
Epitope	Position	MHC class II allele with percentile rank having IC50 values < 500 nM
DSPLEASKRWAFRTG	61-75	HLA-DRB1*03:01 (5.77), HLA-DRB1*09:01 (10.07), HLA-DRB3*01:01 (11.91), HLA-DRB1*07:01 (14.01), HLA-DRB1*15:01 (19.58)
GKSLLLDPPTNVRDY	102-116	HLA-DRB1*03:01 (0.25), HLA-DRB3*01:01 (1.5), HLA-DRB1*13:02 (2.3), HLA-DRB1*04:01 (3.26), HLA-DRB3*02:02 (6.5), HLA-DRB1*12:01 (12.6), HLA-DRB1*04:05 (14.63), HLA-DRB1*01:01 (18.99)
AQGIALHLWGAFFLY	132-146	HLA-DPA1*01:03/DPB1*02:01 (0.12), HLA-DQA1*01:01/DQB1*05:01 (1.96), HLA-DRB1*15:01 (2.42), HLA-DPA1*01/DPB1*04:01 (2.43), HLA-DPA1*02:01/DPB1*01:01 (5.21)
IALHLWGAFFLYDRI	135-149	HLA-DPA1*01/DPB1*04:01 (0.01), HLA-DPA1*01:03/DPB1*02:01 (0.02), HLA-DPA1*02:01/DPB1*01:01 (1.05), HLA-DQA1*01:01/DQB1*05:01 (1.24), HLA-DPA1*03:01/DPB1*04:02 (2.51), HLA-DRB1*15:01 (2.77), HLA-DPA1*02:01/DPB1*05:01 (4.67)
IASTTMYRGKVFTEG	149-163	HLA-DQA1*01:02/DQB1*06:02 (14.69), HLA-DRB1*15:01 (15.04), HLA-DPA1*01/DPB1*04:01 (17.46)

3.5. The candidate epitopes are highly conserved and cover large portions of the population

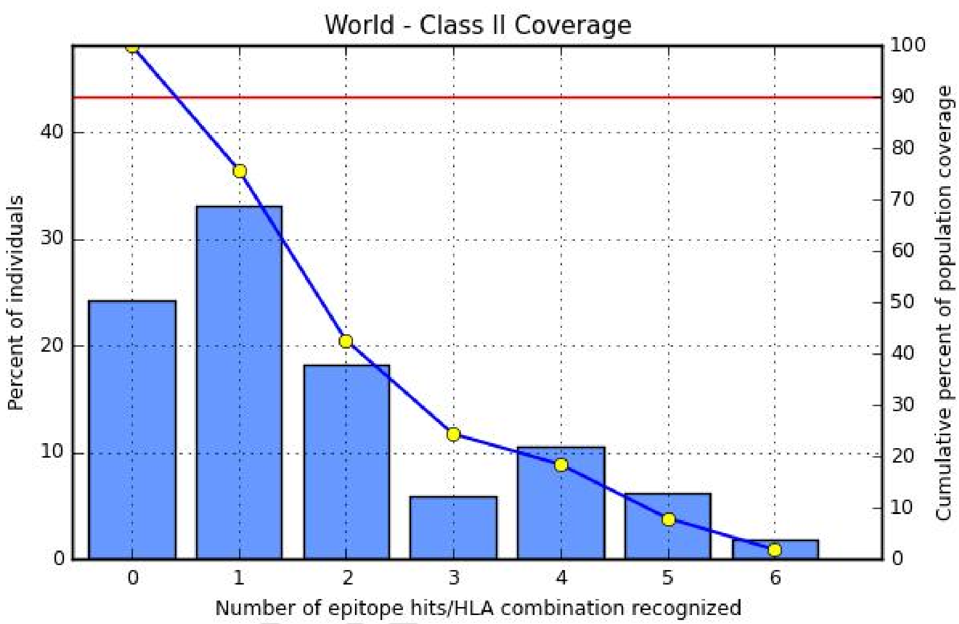
Selection of conserved epitopes confers broader protection against multiple strains, or even species, than epitopes selected from highly variable regions. Therefore, in an epitope based vaccine approach, an ideal epitope should be highly conserved. The epitopes identified in the previous assays were tested for conservancy using the IEDB resources. The epitopes “LEASKRWAF” and “DSPLEASKRWAFRTG” had 100% and 93.62% conservancy in the 47 glycoprotein (GP) sequences (Table 5). Population coverage analyses were also carried out for the epitopes, and it revealed that epitopes interacting with MHC class I molecules had a worldwide coverage of 78.74% (Figure 5.a). On the other hand, the epitopes interacting with MHC class II molecules had a worldwide coverage of 75.75% (Figure 5.b).

Table 5. Conservancy analysis of all the epitopes identified in the study.

Epitope sequence	Epitope length	Conservancy	Minimum identity	Maximum identity
HLWGAFFLY	9	100.00% (47/47)	100.00%	100.00%
TTMYRGKVF	9	80.85% (38/47)	88.89%	100.00%
IALHLWGAF	9	100.00% (47/47)	100.00%	100.00%
LLLDPPTNV	9	55.32% (26/47)	77.78%	100.00%
LEASKRWAF	9	100.00% (47/47)	100.00%	100.00%
DSPLEASKRWAFRTG	15	93.62% (44/47)	93.33%	100.00%
GKSLLLDPPTNVRDY	15	55.32% (26/47)	86.67%	100.00%
AQGIALHLWGAFFLY	15	100.00% (47/47)	100.00%	100.00%
IALHLWGAFFLYDRI	15	82.98% (39/47)	93.33%	100.00%
IASTTMYRGKVFTEG	15	63.83% (30/47)	93.33%	100.00%



a.

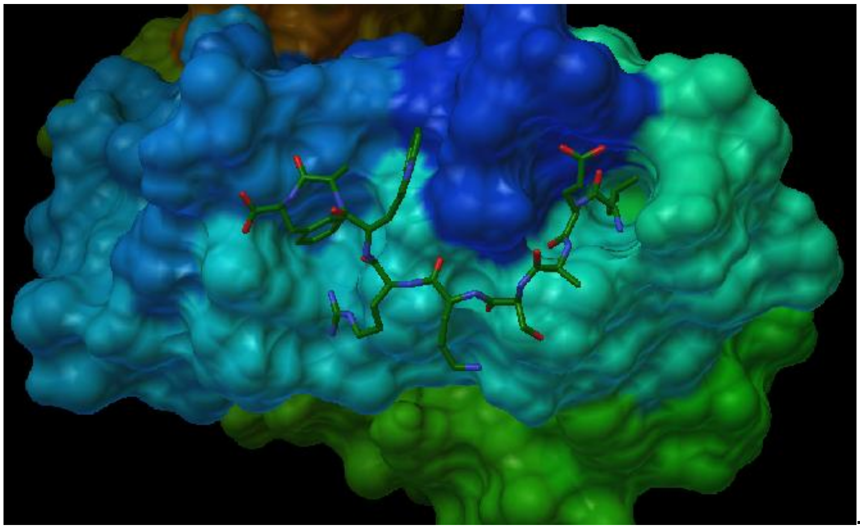


b.

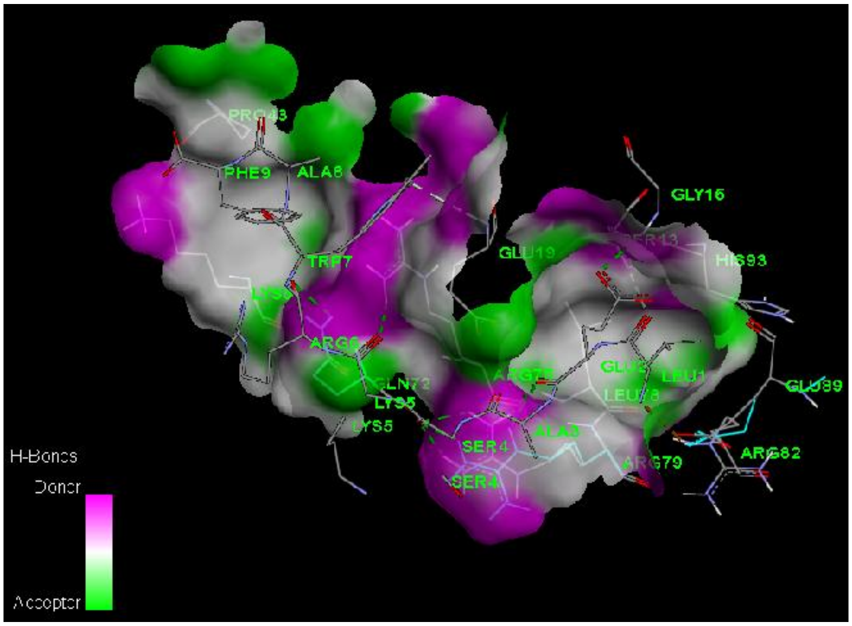
Figure 5. Worldwide population coverage of epitopes with (a) MHC class I alleles and (b) MHC class II alleles respectively.

3.6. The T cell epitope and B cell epitope has high affinity for HLAs

The T cell epitope “LEASKRWAF” interacted with MHC class I allele HLA-B*18:01 (PDB ID: 4XXC) at its binding pocket (Figure 6). This yielded binding affinity of -7.2 kcal/mol indicates a good interaction, while epitope “LLLDPPTNV” interacted with HLA-A*02:03 (PDB ID: 3OX8) with a binding affinity of -8.4 kcal/mol. On the other hand, epitope “DSPLEASKRWAFRTG” interacted with MHC class II allele HLA-DRB1*15:01 (PDB ID: 5V4M) yielded binding affinity of -6.9 kcal/mol (Figure 6). The epitope “GKSLLLDPPTNVRDY”, however, interacted with HLA-DRB1*04:01 (PDB ID: 5JLZ) with binding affinity of -6.6 kcal/mol.

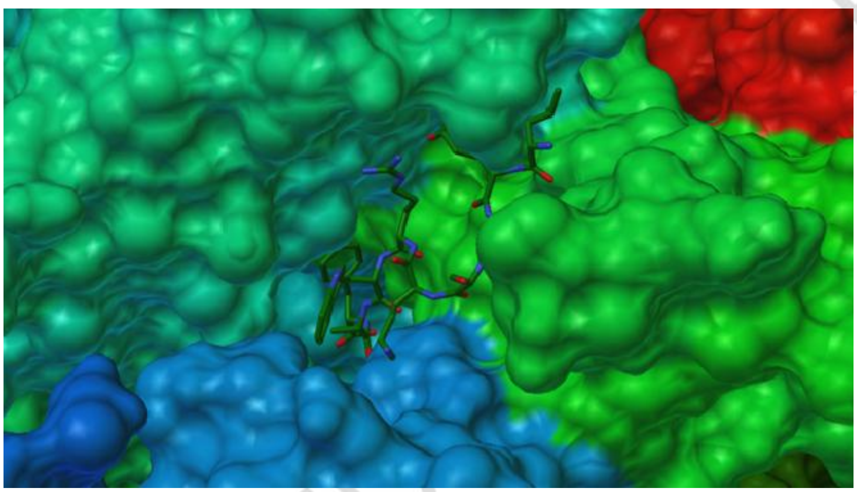
.

a.

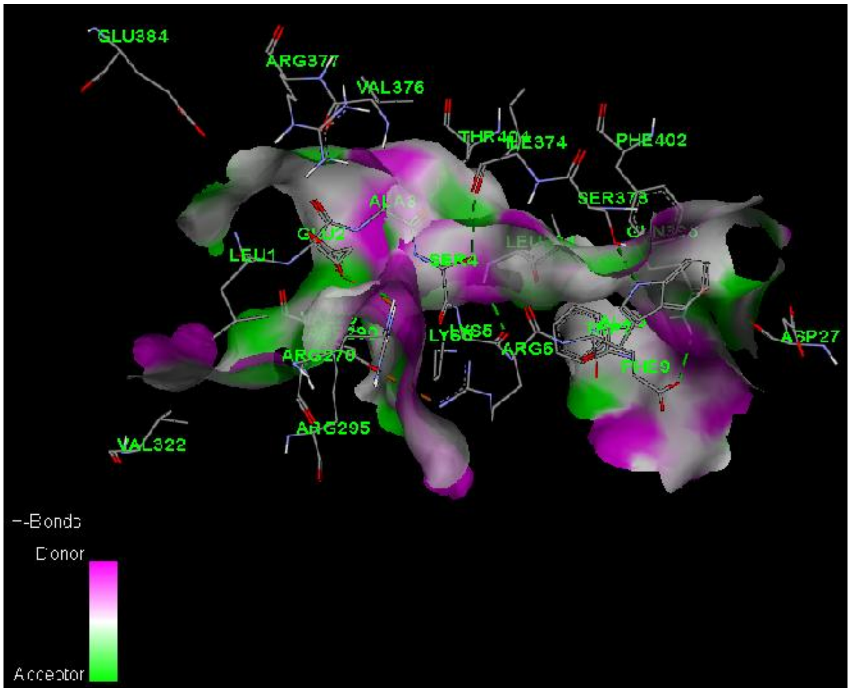


b.

Figure 6. (a) Molecular docking of epitope “LEASKRWAF” with HLA-B*18:01 (PDB ID: 4XXC) yielded binding affinity = -7.2 kcal/mol; (b) H-bond receptor surface of HLA-B*18:01 depicting non-bond interactions.



a.



b.

Figure 7. (a) Molecular docking of epitope “DSPLEASKRWAFRTG” with HLA-DRB1*15:01 (PDB ID: 5V4M) yielded binding affinity = -6.9 kcal/mol (b) H-bond receptor surface of HLA-DRB1*15:01 depicting non-bond interactions.

3.7. The peptide vaccine candidates are non-toxic and do not cross the blood-brain barrier

The ADMET analysis results carried out with SwissADME tool and were cross-referenced with those of admetSAR server. It was found that both of the peptide vaccine candidates could not cross the blood brain barrier, but they were readily absorbed in the human intestine. These epitopes are non-inhibitors of P-glycoproteins, renal organic cation transporter, and many of the CYP450 enzymes. They also have a low CYP inhibitory promiscuity and Non-AMES toxic and non-carcinogens in nature (Table 6).

Table 6. ADMET assessment of epitope “LEASKRWAF” and “DSPLEASKRWAFRTG”.

Model	**Result**	**Probability**	**Result**	**Probability**
Absorption	“LEASKRWAF”		“DSPLEASKRWAFRTG”	
Blood-Brain Barrier	BBB-	0.8969	BBB-	0.9856
Human Intestinal Absorption	HIA+	0.8349	HIA+	0.8617
P-glycoprotein Inhibitor	Non-inhibitor	0.8835	Non-inhibitor	0.6331
Renal Organic Cation Transporter	Non-inhibitor	0.7958	Non-inhibitor	0.7665
Metabolism				
CYP450 1A2 Inhibitor	Non-inhibitor	0.821	Non-inhibitor	0.8043
CYP450 2C9 Inhibitor	Non-inhibitor	0.8141	Non-inhibitor	0.8002
CYP450 2D6 Inhibitor	Non-inhibitor	0.8809	Non-inhibitor	0.898
CYP450 3A4 Inhibitor	Non-inhibitor	0.7562	Inhibitor	0.5
CYP Inhibitory Promiscuity	Low CYP Inhibitory Promiscuity	0.9103	Low CYP Inhibitory Promiscuity	0.868
Toxicity				
AMES Toxicity	Non-AMES toxic	0.7156	Non-AMES toxic	0.7249
Carcinogens	Non-carcinogens	0.9137	Non-carcinogens	0.8413
Acute Oral Toxicity	III	0.5991	III	0.5795

3.8. The in vivo results verify the in-silico workflow

The results of the study remained questionable until it was tested and found to be concordant with in vivo results. The negative control or random sequence failed to pass through the steps of the workflow. On the contrary, four of the six peptides tested by Shi et al. [71] were identified as antigenic epitopes in our workflow as well. However, PEP37 and PEP71 were filtered out in our workflow. Random sequence used as negative control failed to pass the first step of the workflow.

4. DISCUSSION

Ebola and Marburg viruses are classified as category A and biosafety level 4 list pathogens which cause severe hemorrhagic fever with high mortality rate [75]. Although four decades have elapsed after the first discovery of these, still there is no licensed vaccine available in the market [2]. Several attempts have been taken by scientists to develop vaccine but none has shown promising efficacy in preclinical or clinical trial to be approved for market availability [76]. Though the incidents of the breakout of these viruses have been found mainly in African countries, it has the potential to spread all over the world within a very short time [77]. Therefore, development of viable universal vaccine has become an urgent issue.

Most vaccine currently available is based on either inactivated or live-attenuated pathogen, but the major drawback of these vaccines is the safety issue as they may reactivate in the human body and cause deleterious effect. In this case, epitope based vaccine can mitigate or avoid the possible harmful effects as it contains only a short peptide. Currently vaccine development using bioinformatics has gained popularity as it reduces time consuming trial and error process and can be exploited to develop vaccine against emerging viruses within a very short time. In a previous study, Raju Das et al. [78] designed an epitope based vaccine against Ebola virus and in another study, Anum Munir et al. [79] proposed another epitope based peptide vaccine against Marburg virus. But to our best knowledge till now, there is no combined single vaccine design against these two deadly viruses.

In our study, we focused on designing epitope based universal vaccine with global efficacy against these two deadly viruses. For that, we selected the glycoprotein (GP) out of seven different proteins produced by both viruses as it contains large conserved region positioned on the outer membrane that may easily facilitate to mount immune response. From the epitope conservancy analysis, the two epitopes “LEASKRWAF” (64 a.a-72 a.a.) and “DSPLEASKRWAFRTG” (61 a.a.-75 a.a) had been found 100% and 93.62% conserved in the 47 GP sequences respectively and population coverage analysis revealed that epitopes “LEASKRWAF” interacting with MHC class I molecules and “DSPLEASKRWAFRTG” interacting with MHC class II molecules had worldwide coverage of 78.74% and 75.75% respectively. As the high epitope conservancy and large population coverage are the prerequisites of vaccine candidate, the both peptides fulfill these criteria. ABCpred and IEDB software identified the B cell epitope “PLEASKRWAFRTGVPP” (63 a.a-78 a.a) which has higher surface accessibility scores, hydrophilicity scores and antigenicity scores that are the crucial requirements of an epitope to be considered as vaccine. Most importantly, B cell and T cell epitope has sequence similarity that indicates same epitope can induce both B cell and T cell mediated immunity. From the molecular docking analysis, it was found that the binding affinity of “LEASKRWAF” epitope interacted with MHC class I allele HLA-B*18:01 was -7.2 kcal/mol and “DSPLEASKRWAFRTG” interacted with MHC class II allele HLA-DRB1*15:01 was -6.9 kcal/mol, which indicates good interaction between epitope and allele. The ADMET analysis revealed that both peptide vaccine candidates were not susceptible to cross the blood brain barrier, non-AMES toxic and non-carcinogens in nature. Finally, the epitopes were category III oral toxic compounds, but the dosage needed to cause toxicity is very high (500-5000 mg/kg), and therefore poses minimal risk.

5. CONCLUSION

In sum, this study suggests an epitope based vaccine against both Ebola and Marburg viruses with low side effects. Our results are based on sequence data analysis, binding interaction between MHC molecule and epitopes, toxicity test and the predicted epitopes can be used as a target for the development of pan-filovirus vaccine. Both in vitro and in vivo experiments are needed to test the effectiveness of these vaccine candidates.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

REFERENCES

References must be listed at the end of the manuscript and numbered in the order that they appear in the text. Every reference referred in the text must also present in the reference list and vice versa. In the text, citations should be indicated by the reference number in brackets [3].

1. Yang, X.L., et al., Characterization of a filovirus (Mengla virus) from Rousettus bats in China. Nat Microbiol, 2019.

2. Reynolds, P. and A. Marzi, Ebola and Marburg virus vaccines. Virus Genes, 2017. 53(4): p. 501-515.

3. Brainard, J., et al., Risk factors for transmission of Ebola or Marburg virus disease: a systematic review and meta-analysis. Int J Epidemiol, 2016. 45(1): p. 102-16.

4. Gordon, T.B., et al., Host and Viral Proteins Modulating Ebola and Marburg Virus Egress. Viruses, 2019. 11(1).

5. Towner, J.S., et al., Isolation of Genetically Diverse Marburg Viruses from Egyptian Fruit Bats. PLoS Pathogens, 2009. 5(7): p. e1000536.

6. Goldstein, T., et al., The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. Nat Microbiol, 2018. 3(10): p. 1084-1089.

7. Beeching, N.J., M. Fenech, and C.F. Houlihan, Ebola virus disease. Bmj, 2014. 349: p. g7348.

8. Brainard, J., et al., Presence and Persistence of Ebola or Marburg Virus in Patients and Survivors: A Rapid Systematic Review. PLoS Negl Trop Dis, 2016. 10(2): p. e0004475.

9. Adjemian, J., et al., Outbreak of Marburg hemorrhagic fever among miners in Kamwenge and Ibanda Districts, Uganda, 2007. J Infect Dis, 2011. 204 Suppl 3: p. S796-9.

10. Nyakarahuka, L., et al., Knowledge and attitude towards Ebola and Marburg virus diseases in Uganda using quantitative and participatory epidemiology techniques. PLoS Negl Trop Dis, 2017. 11(9): p. e0005907.

11. Coltart, C.E., et al., The Ebola outbreak, 2013-2016: old lessons for new epidemics. Philos Trans R Soc Lond B Biol Sci, 2017. 372(1721).

12. Feldmann, H. and T.W. Geisbert, Ebola haemorrhagic fever. Lancet, 2011. 377(9768): p. 849-862.

13. Martini, G.A., Marburg virus disease. Clinical syndrome, in Marburg virus disease. 1971, Springer. p. 1-9.

14. Mehedi, M., et al., Clinical aspects of Marburg hemorrhagic fever. Future Virol, 2011. 6(9): p. 1091-1106.

15. Qin, E., et al., Clinical Features of Patients With Ebola Virus Disease in Sierra Leone. Clin Infect Dis, 2015. 61(4): p. 491-5.

16. Fauquet, C.M., et al., Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses. 2005: Academic Press.

17. Feldmann, H., H.D. Klenk, and A. Sanchez, Molecular biology and evolution of filoviruses. Arch Virol Suppl, 1993. 7: p. 81-100.

18. Li, Q., et al., STK11 rs2075604 Polymorphism Is Associated with Metformin Efficacy in Chinese Type 2 Diabetes Mellitus. International Journal of Endocrinology, 2017. 2017: p. 6.

19. Taylor, D.J., R.W. Leach, and J. Bruenn, Filoviruses are ancient and integrated into mammalian genomes. BMC Evolutionary Biology, 2010. 10: p. 193-193.

20. Martin, B., et al., Filovirus proteins for antiviral drug discovery: A structure/function analysis of surface glycoproteins and virus entry. Antiviral Res, 2016. 135: p. 1-14.

21. Xu, W., et al., Ebola virus VP24 targets a unique NLS binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated STAT1. Cell Host Microbe, 2014. 16(2): p. 187-200.

22. Valmas, C., et al., Marburg virus evades interferon responses by a mechanism distinct from ebola virus. PLoS Pathog, 2010. 6(1): p. e1000721.

23. Ramanan, P., et al., Structural basis for Marburg virus VP35-mediated immune evasion mechanisms. Proc Natl Acad Sci U S A, 2012. 109(50): p. 20661-6.

24. Albarino, C.G., et al., Genomic analysis of filoviruses associated with four viral hemorrhagic fever outbreaks in Uganda and the Democratic Republic of the Congo in 2012. Virology, 2013. 442(2): p. 97-100.

25. Babirye, P., et al., Identity and validity of conserved B cell epitopes of filovirus glycoprotein: towards rapid diagnostic testing for Ebola and possibly Marburg virus disease. BMC Infect Dis, 2018. 18(1): p. 498.

26. Lee, J.E., et al., Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. Nature, 2008. 454(7201): p. 177-82.

27. Klenk, H.-D. and H. Feldmann, Ebola and Marburg viruses: molecular and cellular biology. 2004: Garland Science.

28. Shabman, R.S., et al., Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. MBio, 2014. 5(6): p. e02011.

29. Côté, M., et al., Small molecule inhibitors reveal Niemann-Pick C1 is essential for ebolavirus infection. Nature, 2011. 477(7364): p. 344-348.

30. Marzi, A., et al., Antibodies are necessary for rVSV/ZEBOV-GP–mediated protection against lethal Ebola virus challenge in nonhuman primates. Proceedings of the National Academy of Sciences of the United States of America, 2013. 110(5): p. 1893-1898.

31. Geisbert, T.W., First Ebola virus vaccine to protect human beings? Lancet, 2017. 389(10068): p. 479-480.

32. Keusch, G., et al., Integrating clinical research into epidemic response: the Ebola experience. Integrating clinical research into epidemic response: the Ebola experience., 2017.

33. Corti, D., et al., Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. Science, 2016. 351(6279): p. 1339-42.

34. Mire, C.E., et al., Therapeutic treatment of Marburg and Ravn virus infection in nonhuman primates with a human monoclonal antibody. Sci Transl Med, 2017. 9(384).

35. Pascal, K.E., et al., Development of Clinical-Stage Human Monoclonal Antibodies That Treat Advanced Ebola Virus Disease in Nonhuman Primates. J Infect Dis, 2018. 218(suppl_5): p. S612-S626.

36. Qiu, X., et al., Ebola GP-specific monoclonal antibodies protect mice and guinea pigs from lethal Ebola virus infection. PLoS Negl Trop Dis, 2012. 6(3): p. e1575.

37. Qiu, X., et al., Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. Nature, 2014. 514(7520): p. 47-53.

38. Qiu, X., et al., Two-mAb cocktail protects macaques against the Makona variant of Ebola virus. Sci Transl Med, 2016. 8(329): p. 329ra33.

39. Davey, R.T., Jr., et al., A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. N Engl J Med, 2016. 375(15): p. 1448-1456.

40. Sivapalasingam, S., et al., Safety, pharmacokinetics, and immunogenicity of a co-formulated cocktail of three human monoclonal antibodies targeting Ebola virus glycoprotein in healthy adults: a randomised, first-in-human phase 1 study. Lancet Infect Dis, 2018. 18(8): p. 884-893.

41. Wec, A.Z., et al., Development of a Human Antibody Cocktail that Deploys Multiple Functions to Confer Pan-Ebolavirus Protection. Cell Host Microbe, 2019. 25(1): p. 39-48 e5.

42. Bornholdt, Z.A., et al., A Two-Antibody Pan-Ebolavirus Cocktail Confers Broad Therapeutic Protection in Ferrets and Nonhuman Primates. Cell Host Microbe, 2019. 25(1): p. 49-58 e5.

43. Brannan, J.M., et al., Post-exposure immunotherapy for two ebolaviruses and Marburg virus in nonhuman primates. Nat Commun, 2019. 10(1): p. 105.

44. Rahim, M.N., et al., Complete protection of the BALB/c and C57BL/6J mice against Ebola and Marburg virus lethal challenges by pan-filovirus T-cell epigraph vaccine. PLoS Pathog, 2019. 15(2): p. e1007564.

45. Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 2004. 32(5): p. 1792-1797.

46. Kumar, S., G. Stecher, and K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol, 2016. 33(7): p. 1870-4.

47. Waterhouse, A.M., et al., Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics, 2009. 25(9): p. 1189-91.

48. Garcia-Boronat, M., et al., PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res, 2008. 36(Web Server issue): p. W35-41.

49. Krogh, A., et al., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol, 2001. 305(3): p. 567-80.

50. Hansen, J.E., et al., NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J, 1998. 15(2): p. 115-30.

51. Chuang, G.-Y., et al., Computational prediction of N-linked glycosylation incorporating structural properties and patterns. Bioinformatics, 2012. 28(17): p. 2249-2255.

52. Doytchinova, I.A. and D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics, 2007. 8: p. 4.

53. Bui, H.H., et al., Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics, 2007. 8: p. 361.

54. Jespersen, M.C., et al., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Research, 2017. 45(Web Server issue): p. W24-W29.

55. Chen, H., F. Gu, and Z. Huang, Improved Chou-Fasman method for protein secondary structure prediction. BMC Bioinformatics, 2006. 7(Suppl 4): p. S14-S14.

56. Kolaskar, A.S. and P.C. Tongaonkar, A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett, 1990. 276(1-2): p. 172-4.

57. Parker, J.M., D. Guo, and R.S. Hodges, New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry, 1986. 25(19): p. 5425-32.

58. Saha, S. and G.P. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins, 2006. 65(1): p. 40-8.

59. Bui, H.H., et al., Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics, 2006. 7: p. 153.

60. Vita, R., et al., The immune epitope database (IEDB) 3.0. Nucleic Acids Res, 2015. 43(Database issue): p. D405-12.

61. Fleri, W., et al., The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Frontiers in Immunology, 2017. 8: p. 278.

62. Zhang, Q., et al., Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Research, 2008. 36(Web Server issue): p. W513-W518.

63. Agatonovic-Kustrin, S. and R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal, 2000. 22(5): p. 717-27.

64. Bhasin, M. and G.P. Raghava, Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine, 2004. 22(23-24): p. 3195-204.

65. Nielsen, M., C. Lundegaard, and O. Lund, Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics, 2007. 8: p. 238-238.

66. Lamiable, A., et al., PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res, 2016. 44(W1): p. W449-54.

67. Trott, O. and A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. Journal of computational chemistry, 2010. 31(2): p. 455-461.

68. Kemmish, H., M. Fasnacht, and L. Yan, Fully automated antibody structure prediction using BIOVIA tools: Validation study. PLoS ONE, 2017. 12(5): p. e0177923.

69. Daina, A., O. Michielin, and V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports, 2017. 7: p. 42717.

70. Cheng, F., et al., admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. J Chem Inf Model, 2012. 52(11): p. 3099-105.

71. Shi, J., et al., Comparative proteomics profile of lipid-cumulating oleaginous yeast: an iTRAQ-coupled 2-D LC-MS/MS analysis. PLoS One, 2013. 8(12): p. e85532.

72. Blythe, M.J. and D.R. Flower, Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Sci, 2005. 14(1): p. 246-8.

73. Yang, X. and X. Yu, An introduction to epitope prediction methods and software. Rev Med Virol, 2009. 19(2): p. 77-96.

74. Gao, J., et al., BEST: Improved Prediction of B-Cell Epitopes from Antigen Sequences. PLoS ONE, 2012. 7(6): p. e40104.

75. Bente, D., et al., Disease modeling for Ebola and Marburg viruses. Dis Model Mech, 2009. 2(1-2): p. 12-7.

76. Sridhar, S., Clinical development of Ebola vaccines. Ther Adv Vaccines, 2015. 3(5-6): p. 125-38.

77. Kalra, S., et al., The emergence of ebola as a global health security threat: from 'lessons learned' to coordinated multilateral containment efforts. J Glob Infect Dis, 2014. 6(4): p. 164-77.

78. Dash, R., et al., In silico-based vaccine design against Ebola virus glycoprotein. Adv Appl Bioinform Chem, 2017. 10: p. 11-28.

79. Anum Munir1*, S.A., 2, Maria Manzoor1, Azhar Mehmood1, G. Mujtaba Shah3, Sahar Fazal2, In-silico epitopebased vaccine an excellent solution against Marburg virus. Journal of Innovations in Pharmaceutical and Biological Sciences(JIPBS), 2016. 3(4).