

Epitope-based peptide vaccine against Bombali Ebolavirus viral protein 40: An immunoinformatics combined with molecular docking studies



Mujahed I. Mustafa ^{a,*}, Shaza W. Shantier ^b, Miyssa I. Abdelmageed ^b, Abdelrafie M. Makhawi ^a

^a Department of Biotechnology, University of Bahri, Khartoum, Sudan

^b Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan

ARTICLE INFO

Keywords:

Bombali ebolavirus
Matrix protein
Viral protein (VP40)
Vaccine design
Immunoinformatics

ABSTRACT

Bombali Ebolavirus belongs to the RNA group of viruses which causes lethal hemorrhagic fever with a high mortality rate. Despite these facts, yet there is no approved vaccine have been developed for the eradication of Bombali Ebolavirus infections. Therefore, this study described a multi epitope-based peptide vaccine against Bombali Ebolavirus VP40, using several immunoinformatics tools combined with molecular docking studies. The VP40 sequences of the sixty-six Ebolavirus strains were retrieved from National Center for Biotechnology Information (NCBI) and Uniprot, and then submitted to VaxiJen for identification of the most antigenic protein. B-cell and T-cell prediction, population coverage and molecular docking analysis were achieved to determine the most promising Bombali VP40 epitopes. The VP40 (YP_009513276.1) of Bombali Ebola virus was found to be the most antigenic protein among all the analyzed ones, and thus selected for the further predictions. For T cell prediction, two epitopes showed high affinity to MHC class I with high population coverage against Africa and the world. Furthermore, in MHC class II, three promising epitopes were found associated with most common MHC class II alleles. The above results indicate that these peptides capable of provoking T-cell response and being interacted with a wide range of HLA molecules that suggested a strong potential for a vaccine candidate against Bombali Ebolavirus.

1. Introduction

Ebola virus (EBOV) is an enveloped, non-segmented, negative-stranded RNA virus which belongs to the Filoviridae family [1,2]. Although, they are causing lethal hemorrhagic fever with high mortality rate [3–5], and have been transferred from several mammal groups to the human species [6], their main source is still unknown. In addition to Africa, it has been reported in the United Kingdom, and Italy, beside secondary cases (Health-care workers) which has also occurred in the United States and Spain [7,8].

Symptoms are characterized by an immediate onset of flu-like illness, followed a preliminary incubation period of 2–21 days. After this preliminary period of infection, the symptoms became noticeable which include chest pain, cough, nausea, vomiting and mucosal hemorrhage [9].

The Ebola virus infection is thought to be the most lethal, with a death rate of up to 90% in human cases [10]. It is a biothreat pathogen due to its high fatality rates, lack of therapeutic options, and lack of immunization [11].

Various strains of EBOV have been identified including Sudan (SEBOV), Tai Forest, Zaire (ZEBOV), Reston, Bundibugyo and the most recently identified strain Bombali virus. Among those strains, ZEBOV causing infection is considered to be the most frequently occurring with the highest number of deaths.

The Ebola virus matrix protein VP40 (viral protein 40 kDa) is the most abundantly expressed filoviral protein which serves as the primary matrix protein and coordinates virion assembly at the plasma membrane through interactions with both viral and cellular components [12]. VP40 can assemble either as a hexamer, which appears to be involved in budding, or as an octamer that functions in genome replication and RNA binding [13,14].

The matrix proteins VP40 with their novel adjuvant have already been demonstrated to be safe when administered intramuscularly or subcutaneously, and therefore, they are closer to clinical trials than adjuvants whose safety profiles are unknown [15]. With all the available knowledge, several attempts have been made to produce an effective vaccine against Ebola viruses due to its deadly nature [16].

Computer-based tools for prediction of an ideal vaccine candidate

* Corresponding author.

E-mail address: mujahedimustafa@gmail.com (M.I. Mustafa).

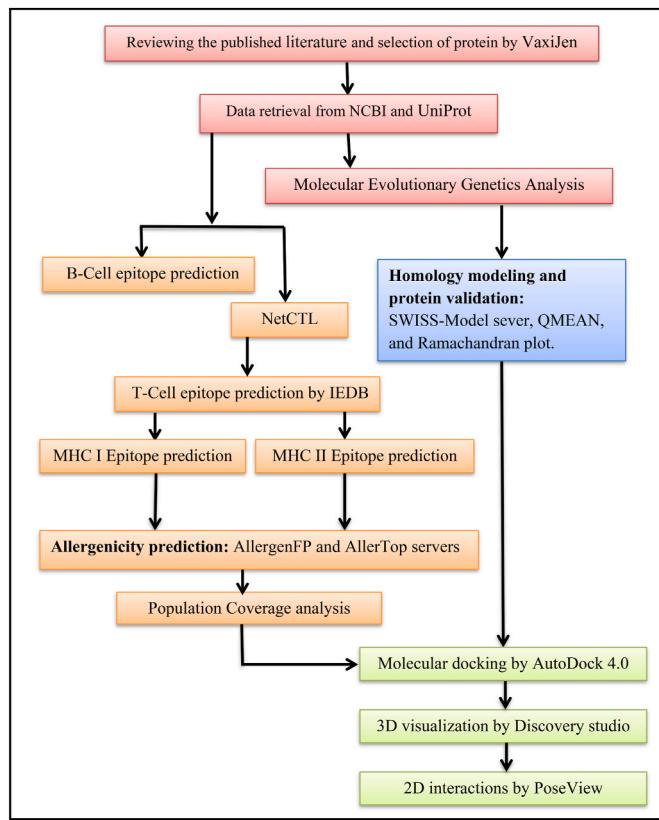


Fig. 1. Illustrated the immunoinformatics approaches used for vaccine design against Bombali Ebolavirus VP40.

Table 1

Predicted conserved regions of Ebola matrix protein with their antigenicity scores by VaxiJen.

Conserved Peptide	Start	End	VaxiJen Score
LPTAPP	6	11	-0.1907
LRPIADD	51	57	-0.2058
SAFILEA	71	77	0.3281
VNVISGPKVLMKQIP	79	93	-0.0984
WLPLGV	95	100	2.8912
YSFDSTTAIAML	106	117	0.6757
SYTITHPGK	119	127	0.9166
LVRVNRLG	132	139	1.0252
GIPDHPLRLLRGNQAFLQEFLVLPVQLPQYFTFDL	141	176	0.4180
ALKLITQPLPA	178	188	-0.0193
LRPGISFHPKLRPILLP	203	219	1.5593
LQDLKIVPIDP	244	254	2.1846
IMGIEVPE	258	265	0.1958
LVHKLGTGKK	267	275	-0.1479
GQPIIPVLLPK	281	291	0.3458
IGLDPV	293	298	2.0539
LTMVIT	303	308	0.9332
CHSPAS	314	319	-0.1678

have helped in reducing the number of validation experiments and time for epitope prediction. Thus, the approach of epitope-based vaccine design has been performed against several life-threatening diseases [17–20]. Although, most of epitope-based vaccines are developed based on B cell epitopes, yet the potentiality of T cell epitope-based vaccine is also promising because CD8⁺ T cell can induce a more effective immune response of the host cell towards the infected T cells [21].

The scientific community is currently faced with the issues of ensuring an effective and long lasting response to the vaccine candidate, as well as protection against various viral types [22,23]. The creation of peptide-based vaccinations provides a safer alternative to

live-attenuated or inactivated vaccines, utilizing the immunoinformatics techniques. This approach gives immunogenic peptides that were investigated in an in vitro and in vivo system [24,25].

In the present study, consensus based predictions and molecular docking tools were employed to obtain the peptides containing multiple epitopes of Ebola matrix protein which have the potential to interact with a wide range of HLA molecules. Simultaneously, a genome wide search was achieved to recognize the most appropriate vaccine target site by using in silico tools. This study will therefore, enhance further upcoming laboratory-based approach to develop effective vaccine against Bombali Ebolavirus infection.

2. Material and methods

The immunoinformatics approaches combined with molecular docking studies used for vaccine design are illustrated in Fig. 1.

2.1. Sequence retrieval

The FASTA format of the reference sequences of Bundibugyo Ebola (NC_014373), Tai Forest Ebola (NC_014372), Zaire Ebola (NC_002549), Bombali Ebola (NC_039345), Sudan Ebola (NC_006432), Reston Ebola (NC_004161), and the matrix protein (VP40) of Bombali Ebola (YP_009513276.1) were retrieved from the National Center of Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>).

The FASTA format of 66 sequences of Ebola matrix protein with 326 aa long, were then retrieved from UniProt (<https://www.uniprot.org/>) [26]. The accession numbers are: AOA075WBZ0, AOA096YGW8, AOA0E3TL71, AOA0E3TLS6, AOA0E3TMB6, AOA0E3TML4, AOA0E3XK76, AOA0F6PDU8, AOA0F6PEC8, AOA0F6PGG2, AOA0F6PGI8, AOA0F6PHJ6, AOA0F7DDX8, AOA0F7IMB4, AOA0F7IMQ7, AOA0F7IPK1, AOA0G2YAY4, AOA0G2YD18, AOA0G2YK96, AOA0H3VWT4, AOA0K1NY98, AOA0K1NYV4, AOA0S0GCL5, AOA0S0H2H3, AOA0S0H2R4, AOA0S0HBW4, AOA0S2MMJ0, AOA0U2JQR9, AOA0U3V1M8, AOA0U3V4D3, AOA0U3VI29, AOA0U3VJF3, AOA0U4CXB3, AOA0U4CZ38, AOA0U4DB64, AOA0U4DBF7, AOA0U4E032, AOA0U4EET9, AOA0U4EF82, AOA0U4ER08, AOA0U4ERC6, AOA0U4ET02, AOA1C4HD11, AOA1L2BJ47, AOA1S6GW91, AOA288QR57, AOA2I4PE05, AOA2U9QPQ5, AOA2Z2FH29, AOA3G2LM27, AOA3G2XD69, AOA3Q8HH57, AOA3Q8HHB0, AOA3S7SNR5, AOA3S7SQ02, AOA3S7SQ58, A9QPL8, B8XCN8, G8DB48, I7F2J6, L7QHU6, X5H596, Q2PDK5, Q05128, Q5XX06, Q77DJ6.

2.2. Antigenicity prediction

Antigenicity prediction of the matrix protein (VP40) of the six Ebola strains was performed by VaxiJen using threshold of 0.4 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [27]. VaxiJen Server allows the classification of antigens solely based on the physicochemical properties of proteins without referring to sequence alignment.

2.3. Analyses of the conserved region

Pairwise and multiple sequence alignment of the 66 retrieved sequences were performed using ClustalW (<https://www.genome.jp/tools-bin/clustalw>), with gap opening penalty of 10.00 for both alignment, gap extension penalty of 0.10 for pairwise sequence alignment and 0.20 for multiple sequence alignment [28].

2.4. Phylogenetic analysis

Phylogenetic analysis reveals the evolutionary relationship between the different strains of the Ebola virus and trace back patterns of common ancestry between their lineages. The analysis was performed using

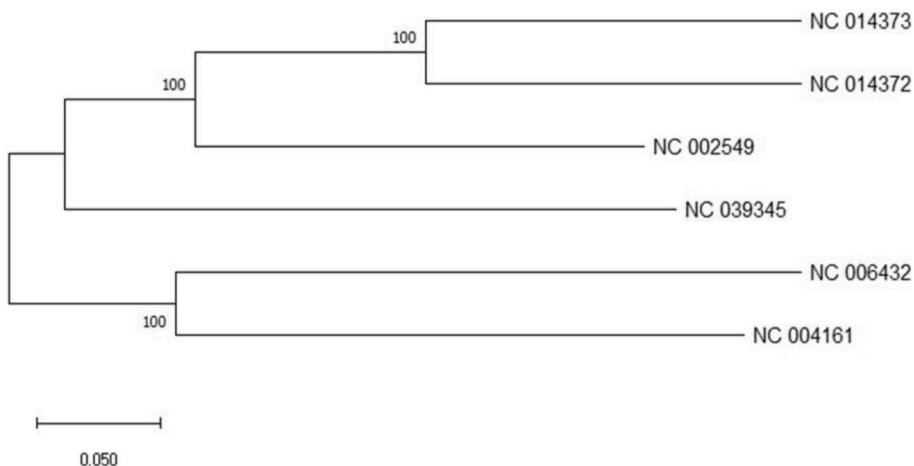


Fig. 2. Illustrates the maximum likelihood phylogenetic tree, which constructed based on whole-genome sequence alignment of the six Ebola virus strains.

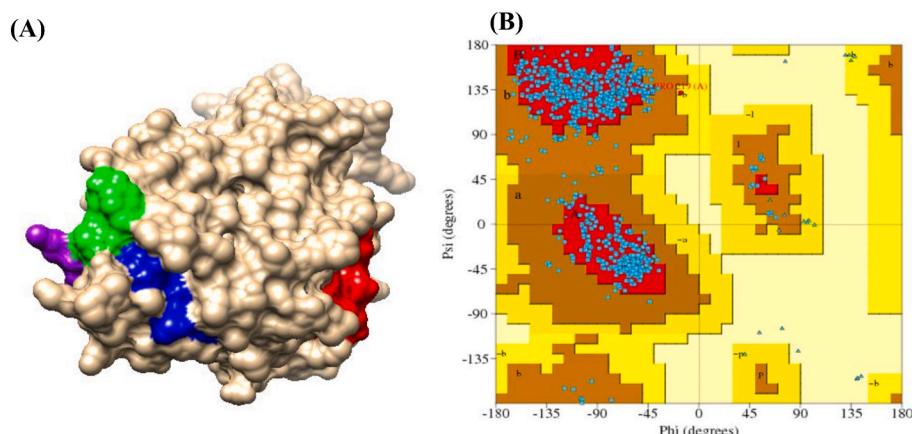


Fig. 3. (A) Surface view showing the location of the predicted peptides, YSFDSSTTAA (red); VQLPQYFTF (green); LPQYFTFDL (blue); QEFVLPPVQ (purple); within the modeled 3D structure of VP40. (B) Ramachandran plot analysis to validate the 3D predicted structure revealing 92% of the residues of VP40 protein are located in the most favored region. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Molecular Evolutionary Genetics Analysis (MEGA-X) (version 10.1.18) (<https://www.megasoftware.net>) [29]. It is a software applied for sequence alignment, inferring phylogenetic trees, and testing evolutionary relationship. It can analyze DNA, RNA and Protein sequences [29]. The maximum likelihood phylogenetic tree was constructed with bootstrapping value of 300.

2.5. Homology modeling and protein validation

The tertiary structure of Bombali Ebolavirus matrix protein was generated using SWISS-Model sever (<https://swissmodel.expasy.org/>). Discovery studio 2020 (<https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/visualization/>) was used to visualize the most promising peptides for vaccine design. The confirmation of the predicted 3D structure was done using QMEAN (<https://swissmodel.expasy.org/qmean/>), while the protein quality was tested using PROCHECK server (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) by Ramachandran plot.

2.6. B-cell epitope prediction

B-cells have been known to get actively involved in antibody generation, humoral immunity, and also result in memory cell formation towards encountering future pathogens. Therefore, the hydrophilicity, flexibility, surface accessibility and antigenicity of the proposed

epitopes were predicted using different B cell predication methods at the Immune Epitope Database (IEDB) [30].

2.7. T-cell epitope prediction tools

2.7.1. CTL epitope prediction

Consistent CTL epitope predictions are critical for developing a cohesive vaccine. NetCTL1.2 36, an internet-based server designed to identify human CTL epitopes in a target protein, was utilized for this purpose. The total score was computed by adding the TAP transport efficiency, proteasomal cleavage, and MHC I molecules binding affinity values. The parameter was set at 0.5 which have sensitivity and specificity of 0.89 and 0.94, respectively [31].

2.7.2. Peptide binding to MHC class I molecules

To predict the interaction with different Histocompatibility Complex class I (MHC Class I) alleles, the Major MHC Class I binding prediction tool on the IEDB (<http://tools.iedb.org/mhci/>) was used. It services distinct approaches to measure the binding affinity of selected sequence to a definite MHC class I molecule. The half maximal inhibitory concentration (IC₅₀) values of peptide binding to MHC class I molecule was calculated by artificial neural network (ANN) approach [32]. All alleles having a binding affinity of IC₅₀ that are equal or less than 100 nM were selected for further analysis.

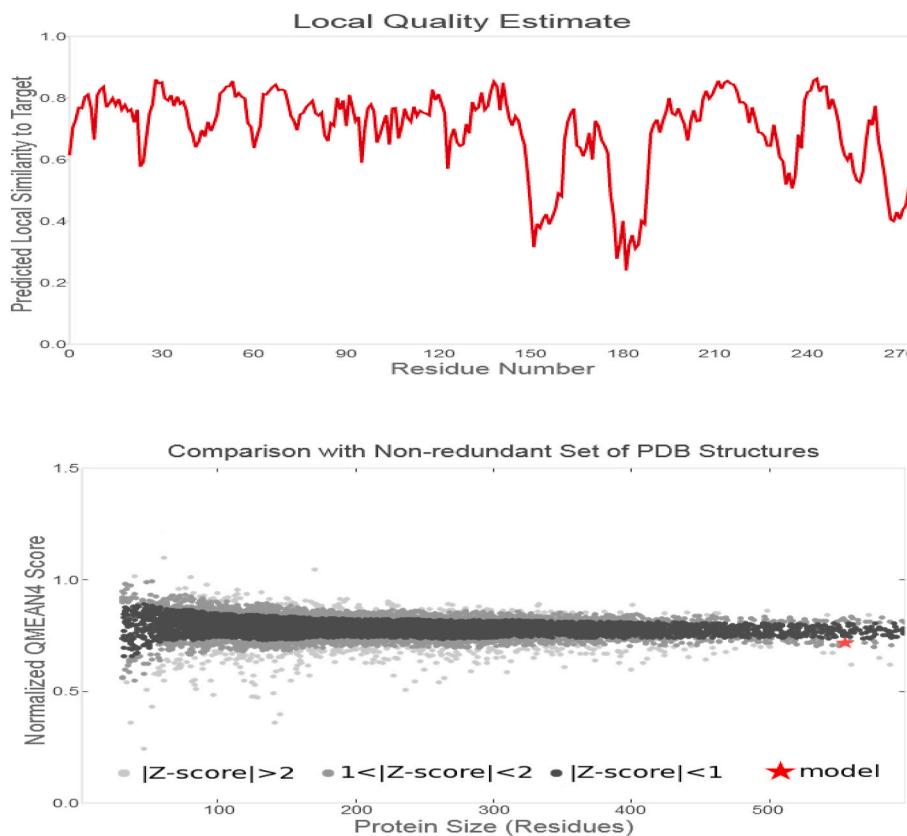


Fig. 4. Assessment of structural superiority by QMEAN valuation.

Table 2
The obtained physicochemical properties.

Characteristics	Finding	Remark
Number of amino acids	326	Suitable
Molecular weight	35408.92 Da	Average
Theoretical pI	8.40	Slightly basic
Chemical formula	C ₁₅₉₇ H ₂₅₅₂ N ₄₂₂ O ₄₆₆ S ₉	–
Extinction coefficient (at 280 nm in H ₂ O)	23045	–
Estimated half-life (mammalian reticulocytes, in vitro)	30 h	–
Estimated half-life (yeast-cells, in vivo)	>20 h	–
Estimated half-life (<i>E. coli</i> , in vivo)	>10 h	–
Instability index of vaccine	37.84	Stable
Aliphatic index of vaccine	96.01	Thermostable
Grand average of hydropathicity (GRAVY)	-0.134	Hydrophilic

2.7.3. Peptide binding to MHC class II molecules

The MHC Class II prediction tool (<http://tools.iedb.org/mhcii/>) provided by the IEDB was used to predict the peptide binding to MHC class II molecules, consistently of human allele references sets was used. The ANN prediction method was selected to recognize the binding affinity of MHC Class II grooves and MHC Class II binding core epitopes. All epitopes that bind to many alleles at a score equal to or less than 250, IC50 were chosen for additional investigation. The Human allele reference sets (HLA DR, DP, and DQ) were included in the prediction.

Table 3
Bepriled, Surface accessibility and Antigenicity values for the Predicted Epitope.

Test	Peptide	Threshold	Average score	Minimum score	Maximum score	Predicted score
Bepriled Linear Epitope Prediction	IPTAPPDYTEALYP	0.35	0.27	-0.001	2.068	–
Emini Surface Accessibility Prediction	IPTAPPDYTEALYP	1	1	0.001	36.497	2.621
Kolaskar & Tongaonkar Antigenicity	IPTAPPDYTEALYP	1.041	1.043	1.001	1.071	1.044

2.8. Physicochemical and allergenicity prediction

The major problem with vaccines is the occurrence of allergy when administered to the human body and therefore, it is crucial to check for allergenicity and physicochemical properties. ProtParam server (<http://web.expasy.org/protparam/>) was used to anticipate the physicochemical features of the protein under study and to understand the fundamental nature of the vaccine [33]. The molecular weight, theoretical pI, instability index, half-life, aliphatic index, and Grand Average of Hydropathy (GRAVY) were all anticipated. These parameters are important in defining a protein's pH properties, the extinction coefficient (which shows how much light a protein absorbs at a specific wavelength), and the predicted half-life and stability of the protein. AllergenFP v.1.0 [34], and AllerTop v 2.0 servers [35] were also used to evaluate the allergenicity of the predicted peptides.

2.9. Population coverage

The population coverage for each epitope was calculated by the IEDB population coverage tool at (http://tools.iedb.org/tools/population/iedb_input). All epitopes and their binding to MHC Class I and MHC Class II molecules were assessed against population covering the World and Africa.

Table 4

Predicted epitopes binding with MHC class I alleles along with their allergenicity prediction.

Peptide	Alleles	Position	Allergenicity	IC50	COMB score ^a
YSFDSTTAA	HLA-A*02:01, HLA-A*02:06, HLA-A*68:02, HLA-B*15:01, HLA-B*35:01, HLA-B*46:01, HLA-C*03:03, HLA-C*12:03	106–114	Non-allergen	20.63	0.509
VQLPQYFTF	HLA-A*02:06, HLA-A*23:01, HLA-A*24:02, HLA-A*32:01, HLA-B*15:01, HLA-C*07:02	166–174	Non-allergen	17.61	0.513

^a Combined score of peptide MHC class I binding, proteasomal C terminal cleavage score and TAP transport efficiency.

Table 5

Predicted epitopes binding with MHC Class II alleles along with their allergenicity prediction.

Peptide	Alleles	Allergenicity	IC50
LPQYFTFDL	HLA-DPA1*01:03/DPB1*04:01, HLA-DPA1*01:03/DPB1*04:02, HLA-DPA1*01:03/DPB1*06:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DQA1*01:01/DQB1*05:01, HLA-DQA1*03:03/DQB1*04:02, HLA-DQA1*01:02/DQB1*05:01, HLA-DQA1*04:01/DQB1*04:02, HLA-DRB1*04:02, HLA-DRB1*04:03, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:01, HLA-DRB1*08:02, HLA-DRB1*09:01	Non-allergen	23.2
QEFVLPPVQ	HLA-DPA1*01:03/DPB1*04:02, HLA-DPA1*01:03/DPB1*06:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*02:01/DPB1*05:01, HLA-DQA1*01:02/DQB1*05:01, HLA-DQA1*01:03/DQB1*06:03, HLA-DQA1*04:01/DQB1*04:02, HLA-DQA1*05:01/DQB1*03:02, HLA-DQA1*06:01/DQB1*04:02	Non-allergen	28.4
YSFDSTTAA	HLA-DQA1*01:04/DQB1*05:03, HLA-DQA1*02:01/DQB1*04:02, HLA-DQA1*03:01/DQB1*03:02, HLA-DQA1*05:01/DQB1*03:03, HLA-DQA1*05:01/DQB1*04:02, HLA-DRB1*01:01, HLA-DRB1*01:03, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:02, HLA-DRB1*04:03, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*08:02	Non-allergen	106.0

Table 6

Population coverage for all epitopes binding to MHC class I and MHC class II in the World and Africa.

MHC Class	Population coverage		
	World	East Africa	Central Africa
I	94.78%	87.28%	79.66%
II	68.40%	85.51	75.1

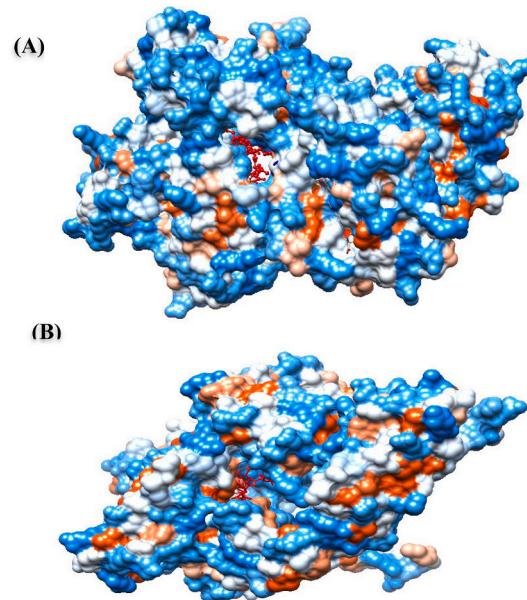


Fig. 5. Representative surface view for Docking analysis of the predicted epitope YSFDSTTAA with HLA-A*02:01 (A) and HLA-DRB1*07:01 (B).

2.10. Molecular docking analysis

In order to estimate the binding affinities between the proposed epitopes and molecular structure of T cells, in silico molecular docking was used. Sequences of proposed epitopes were selected from Ebola virus reference sequence using Chimera 1.10 and saved as PDB file. The obtained files were then optimized and energy minimized. The 3D structures of MHC Class I allele HLA-A*02:01 (PDB ID: 4UQ3) and MHC Class II alleles HLA-DRB1*07:01 (PDB ID: 6HBY), HLA-DPA1*02:01/DPB1*14:01 (PDB ID: 6VB3) were retrieved from RCSB PDB (protein data bank) database. Swiss PDB viewer V.4.1.0 software [36] was used for structure optimization and energy minimization.

Molecular docking was then performed using AutoDock 4.0 software [37]. The active residue of the protein was selected, the results less than 1.0 Å in positional root-mean-square deviation (RMSD) were then considered ideal and clustered together for finding the favorable binding [38]. The highest binding energy (most negative) was considered as the ligand with maximum binding affinity. The 3D and 2D interactions of the resultant docking files with poses showing the lowest binding energies were visualized using DS Visualizer Client 2020 and the PoseView [39] at the ProteinPlus web portal [40], respectively.

3. Results

3.1. Antigenicity and conservation analysis

The alignment of 66 VP40 sequences predicted 18 conserved peptides with length more than 5 aa (Table 1). Analysis of those conserved regions by VaxiJen revealed eight antigenic regions as they met the criteria of default threshold ≥ 0.5 in VaxiJen (Table 1).

3.2. Phylogenetic analysis

The obtained multiple sequence alignment of the six Ebola virus strains was used to construct the phylogenetic tree shown in Fig. 2.

3.3. 3D structure prediction and evaluation

The tertiary structure of Bombali Ebolavirus VP40 was generated using SWISS-Model sever (Fig. 3, A.), while the Ramachandran plot

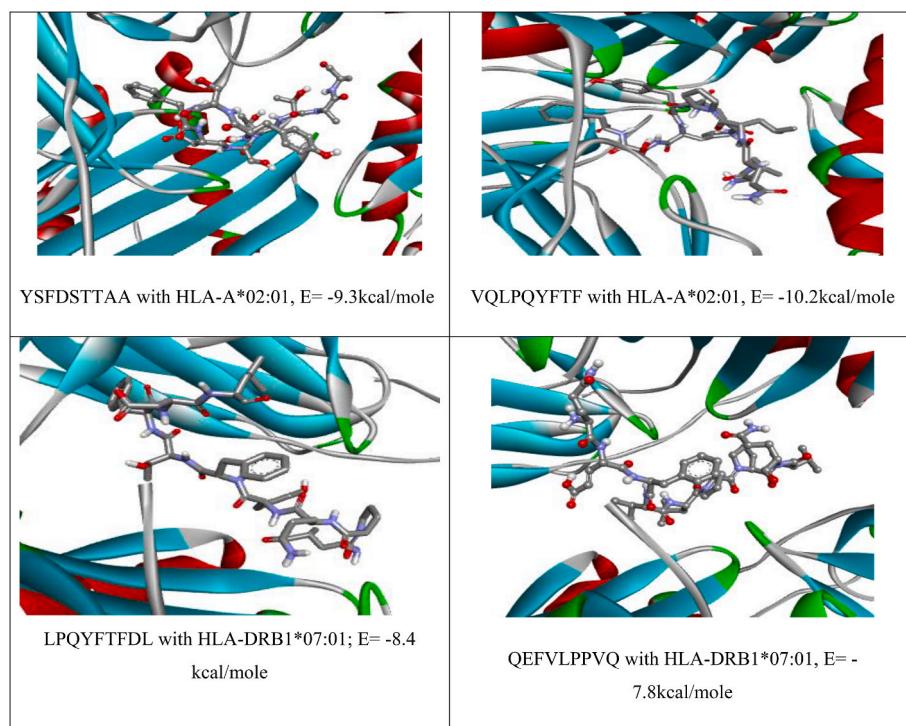


Fig. 6. 3D structures for the most promising peptides bound to MHC I and MHC II alleles.

created by the PROCHECK server showed that about 92% of VP40 protein residues are located in the most favored region (Fig. 3, B). Moreover, the quality of the 3D structure was verified using QMEAN (Fig. 4) QMEAN analysis resulted in a Z-score of -1.20 , and the total score was 0.640 . These values denote a higher quality of the model, where the acceptable score ranges between 0 and 1 [41]. It is worth to mentioned that the protein model having $>90\%$ of the residues in the core and allowed regions can be designated as a high-quality model [42]. Based on the fore mentioned results, the 3D model was found reliable and thus considered for the further study (see Figs. 5, 6).

3.4. Physicochemical properties

The physicochemical properties of the protein were assessed as shown in Table 2.

3.5. B-cell epitope prediction tools

3.6. T-cell epitope prediction tools

The MHC Class I and MHC Class II binding prediction tools predicted 102 peptides and 313 peptides from the VP40 protein that could interact with different MHC Class I and MHC Class II alleles, respectively. Tables 4 and 5 summarizes the most promising peptides bound to MHC Class I and MHC Class II alleles along their allergenicity predication.

3.7. Population coverage

Population coverage analysis was also predicted for the total peptides (Table 6). Obtained results showed that the proposed peptides binding to MHC class I have a 94.78% projected population coverage in the world, and 98.40% in Africa, while the population coverage results for the total number of peptides binding to MHC class II alleles showed a 68.40% estimated population coverage in the world, and 85.51% in Africa.

3.8. Molecular docking

4. Discussion

For decades, EBOV has been a burden to humanity, harming the lives of millions of people all over the world. The virus succeeds continuously to escape the eradication enzymes found in the lysosome by attaching to its membrane [43–45]. In order to overcome this behavior, there are different ways could be followed: either using drug that prevent the virus from binding to NPCI in order to be destroyed inside the lysosome; design an engineered antibody to be attached to the virus glycoprotein which will prevent the virus from binding to the host cell and eventually kill it [46–49]; or design an effective vaccine [50–52].

The current available vaccines have serious adverse effects, and the dread of a viral relapse is always present among the people who have had them. In order to develop a new vaccination that is both safe and free of side effects, the present study was first designed and then formulated a multi-epitope subunit vaccine that can elicit an immunological response against Bombali Ebolavirus using matrix protein VP40 as a target.

Classification of epitopes, which only based on experimental method, is consuming a lot of time and resources. Recent advances of immunoinformatics methods has made it easier to identify potential epitopes, accordingly, decreasing the number of *in vivo* and *in vitro* experiments for epitopes validation [21]. Peptide vaccines are more superior to traditional or single-epitope-based vaccines due to many factors including the presence of multiple HLA epitopes than can be detected by many T-cell receptors, providing a long-term memory; additionally, designed vaccine can be associated with adjuvant thereby boosting its immunogenicity [21]. This idea is based on replacing a whole genome with small antigenic proteins [53–55], which have the ability to strongly evoke the immune system. Thus, the present study was devoted to pinpoint the 100% conserved regions which are then selected to predict the highly immunogenic epitopes for T-cells, the key molecules of cell mediated and humoral immunity as vaccine candidates

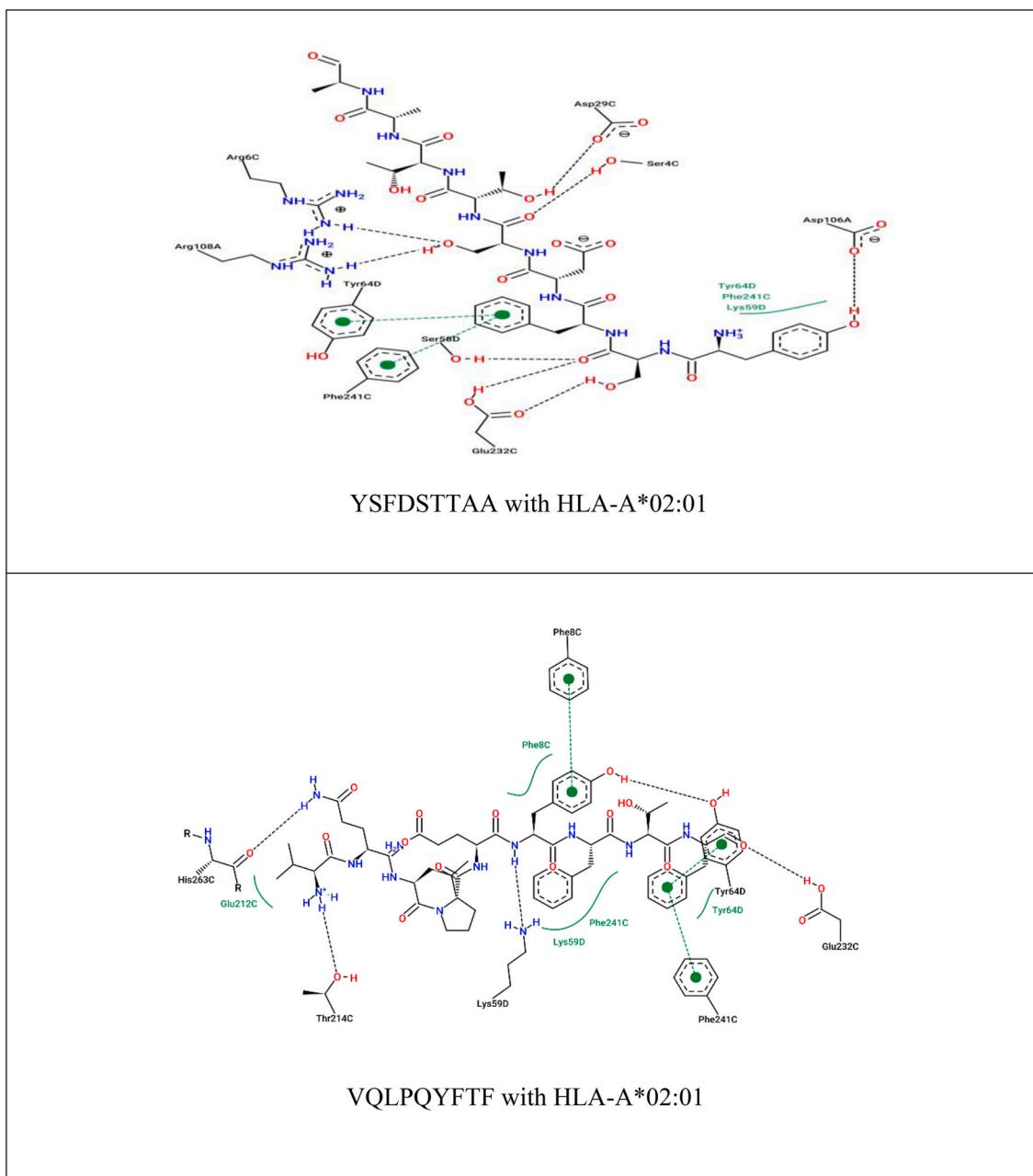


Figure 7. 2D structure for the most promising peptides bound to MHC class I alleles.

for the deadly Bombali virus infection using matrix protein VP40 as a target. It is well known that peptides can be recombinant or construct epitopes, which are targeted against surface or intracellular proteins (such as our target) [56]. Although some studies had determined the ability of VP40 of Ebolavirus alone to evoke strong immune response against Ebolavirus [57,58], still there is no epitope-based vaccine has been predicted for Bombali Ebolavirus.

Indeed, Zaire Ebola is responsible for most of the outbreaks with the highest case-fatality rates of all the ebolaviruses, Bombali Ebola virus was found to be the most antigenic one and hereby selected for the present study.

The conservation analysis utilizing multiple alignments of all the sequences revealed only eight conserved regions which formed the basis for the further analysis as they had been predicted antigenic by the VaxiJen server (Table 1).

Phylogenetic analysis reveals the evolutionary relationship between the different strains of Ebola virus and trace patterns of common

ancestry between their lineages [59] (Fig. 2). represents maximum likelihood phylogenetic tree which constructed based on whole genome sequence alignment of the six Ebola virus strains.

The taxa represent Bundibugyo Ebola (NC_014373), Tai Forest Ebola (NC_014372), Zaire Ebola (NC_002549), Bombali Ebola (NC_039345), Sudan Ebola (NC_006432), and Reston Ebola (NC_004161). Bundibugyo Ebola, Tai Forest Ebola founds on the same clade [60], also Sudan Ebola, Reston Ebola on the same clade [61] and both Zaire and Bombali Ebola are in individual clades. Zaire strain was found to be an ortholog with Bombali strain, which supported by the high bootstrapping values on each node.

The assessed physicochemical properties revealed that the nature of the vaccine, which determined by theoretical PI value, to be acidic. Instability index suggested by server tools indicated that the protein would remain stable after synthesis (Table 2). In contrast, the GRAVY value and aliphatic index portrayed the vaccine to be hydrophilic and thermostable, respectively. Those favorable physicochemical properties

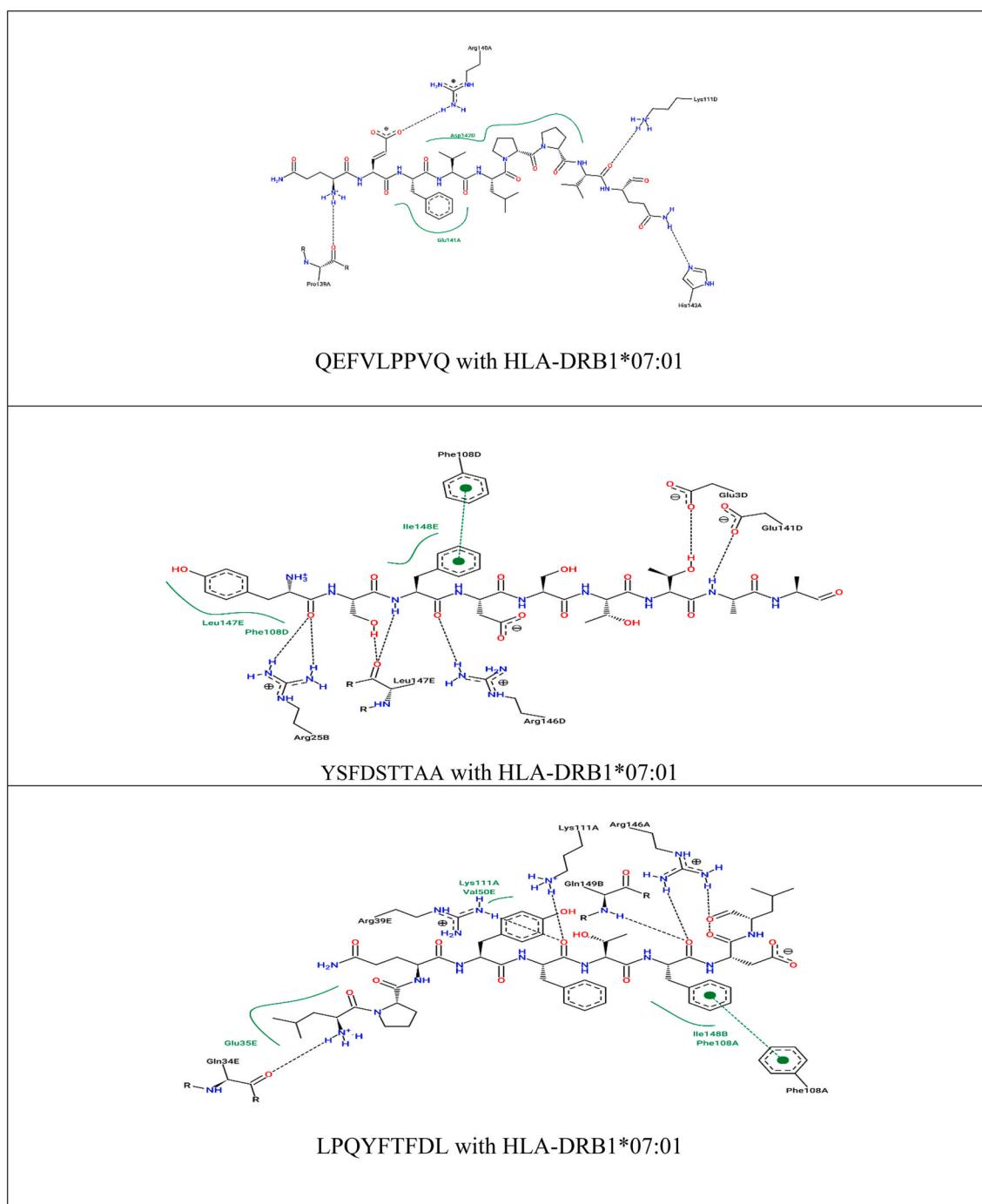


Fig. 8. 2D structures for the most promising peptides bound to MHC II alleles.

predicted for the vaccine and all the scores on different parameters relies on a high possibility to confer this vaccine as a valid candidate against Bombali Ebolavirus VP40.

Additionally, it is well known that a strong and effective immune response depends on the recognition of epitopes by HLA molecules with high affinity. Hence, a peptide is classified by its highest number of bound HLA alleles (termed as promiscuous) and safety, and thus will have the best potential to evoke a strong immune response [62,63].

Among the analyzed antigenic sequences, two epitopes (YSFDSTTAA and VQLPYFTF) were predicted as most promising taking into consideration their conservancy and binding affinity to the highest numbers of MHC class I alleles (Table 4). On the other hand, 3 conserved epitopes (LPQYFTFDL, QEFVLPPVQ, and YSF DSTTAA) were predicted

to interact with several MHC class II alleles named HLA-D, and Q (Table 5), specifying that extra attention required to be devoted to this region.

Besides the binding with the MHC molecules, the predicted peptides must be non-toxic and non-allergen, hence, their safety was predicted using AllergenFP v.1.0 [34] and AllerTop v 2.0 servers [35]. The result showed that all the selected peptides are non-allergen and non-toxin (Tables 4 and 5).

Moreover and in order to stimulate immunological responses, the predicted peptides should interact effectively with the MHC Class I and MHC Class II molecules [60]. Therefore, molecular docking was performed to study their binding affinities selecting the alleles HLA-A*02:01, and HLA-DRB1 as targets due to their diversity and

immunogenic association [65,66].

H-bond formation is an important parameter to deduce the stability of the conformation through the simulation period in the protein ligand complex [67]. Thus, the PoseView at the ProteinPlus web portal was used to illustrate their 2D interactions and bonding with MHC molecules. In regards to MHC Class I, the obtained results showed that YSF DSTTAA and VQLPQYFTF bound to the groove of HLA-A*02:01 with binding energies of -9.3 kcal/mol and -10.2 kcal/mol, respectively. Additionally, the 2D interactions viewed that YSF DSTTAA formed eight hydrogen bonds with Asp29, Asp106, Ser4, Ser58, Arg6, Arg108, Tyr64 and Glu232, while VQLPQYFTF formed five hydrogen bonds with Lys59, Tyr64, Thr214, Glu232 and His263 (Fig. 7). In contrast to MHC II, the most promising peptides (LPQYFTFDL, YSF DSTTAA, QEFVLPPVQ) were bound to MHC Class II molecules with binding energies of -8.4 kcal/mol, -7.3 kcal/mol and -7.8 kcal/mol, respectively. They formed a number of hydrogen bonds ranged from 3 to 6 hydrogen bonds with different amino acids (Fig. 8).

VP40 protein was also searched for B cell epitopes which can induce both primary and secondary immunity. After cross-referencing several tools of IEDB generated results, it was found that the most favorable region from 6 to 20 (IPTAPPDYTEALYP) as potent 15-mer B cell epitope (Table 3).

As a result of these interesting outcomes, the proposed epitopes can be used as a basis for formulating a vaccine against most of the known strains of Ebola virus, and even possibly against newly emerging strains because the basis of this study was the conservation of protein sequences in various strains. To the best of our knowledge, this is the first study to identify specific peptides in the VP40 as candidates for Bombali Ebolavirus. Accordingly, these epitopes were highly recommended as promising epitope vaccine to provoke a strong immune response against Bombali Ebolavirus using VP40 as a target.

5. Conclusion

Bombali Ebolavirus is one of the most re-emerging viruses in certain regions scattering across the world. With no approved vaccine or drug against this deadly virus, the situation during an outbreak gets exacerbated. Therefore, this study devoted to serve as a platform to hasten vaccine development through the design of an epitope-based peptide vaccine against Bombali Ebolavirus viral Protein 40 using an immunoinformatics approach combined with molecular docking studies. The results are originated from a systematic analysis, which suggest that, effective vaccine candidates linking the best-fit epitopes (YSF DSTTAA, VQLPQYFTF, LPQYFTFDL and QEFVLPPVQ) and the complete analysis of this proposed vaccine discloses the high tendency of the vaccine to provoke a strong immune response against Bombali Ebolavirus using matrix protein VP40 as a target. Finally, both in vivo and in vitro experiments are suggested to support these findings.

Data availability

All data underlying the results are available as part of the article, and no additional source data are required.

Authors' contributions

MIM: conceptualization, formal analysis, methodology, validation, and writing (original draft); SWS: conceptualization, formal analysis, visualization, validation, and writing (original draft) and editing the final manuscript version; MIA: conceptualization, formal analysis, methodology, writing (original draft); and AMM: data curation, conceptualization, project administration, supervision, and writing (review and editing). All authors have read and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the Deanship of Scientific Research at University of Bahri for the supportive cooperation.

References

- [1] Beer B, Kurth R, Bukreyev A. Characteristics of Filoviridae: marburg and ebola viruses. *Naturwissenschaften* Jan 1999;86:8–17.
- [2] Kiley MP, Bowen ET, Eddy GA, Isaäcson M, Johnson KM, McCormick JB, et al. Filoviridae: a taxonomic home for Marburg and Ebola viruses? *Intervirology* 1982; 18:24–32.
- [3] Feldmann H, Geisbert TW. Ebola haemorrhagic fever. *Lancet* 2011;377:849–62.
- [4] Pocock L, Rezaeian M. Review : ebola haemorrhagic fever. *Middle East J Fam Med* 2014;12:10. 01.
- [5] Otutume S. Ebola haemorrhagic fever in Zaire. *J Afr Stud* 1996;1996:95–7.
- [6] Badshah S, Ahmed N, Naeem A, Farman A, Khan K. The structure and pathophysiology of the Ebola virus. 2020. p. 1–14.
- [7] Evans DK, Goldstein M, Popova A. Health-care worker mortality and the legacy of the Ebola epidemic. *Lancet Glob Health Aug* 2015;3:e439–40.
- [8] de la Calle-Prieto F, Arsuaga-Vicente M, Mora-Rillo M, Arnalich-Fernandez F, Arribas JR. [Ebola virus disease: update]. *Enferm Infect Microbiol Clin* 2016;34: 452–60.
- [9] Genton C, Cristescu R, Gatti S, Levrero F, Bigot E, Caillaud D, et al. Recovery potential of a western lowland gorilla population following a major Ebola outbreak: results from a ten year study. *PLoS One* 2012;7:e37106.
- [10] Sanchez A, Geisbert TW, Feldmann H. Filoviridae: marburg and ebola viruses. *Fields Virol* 05/21 2013;1409–48.
- [11] Jacob ST, Crozier I, Fischer 2nd WA, Hewlett A, Kraft CS, Vega MA, et al. Ebola virus disease. *Nat Rev Dis Prim Feb* 20 2020;6(13).
- [12] Madara JJ, Han Z, Ruthel G, Freedman BD, Harty RN. The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target. *Future Virol* May 2015;10:537–46.
- [13] Timmins J, Schoehn G, Kohlhaas C, Klenk HD, Ruigrok RW, Weissenhorn W. Oligomerization and polymerization of the filovirus matrix protein VP40. *Virology* Aug 1 2003;312:359–68.
- [14] Bornholdt ZA, Noda T, Abelson DM, Halfmann P, Wood MR, Kawaoka Y, et al. Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle. *Cell Aug* 15 2013;154:763–74.
- [15] Feng H, Nakatsu S, Lopes T, Imai M, Yamayoshi S, Yamashita M, et al. Identification of novel adjuvants for ebola virus-like particle vaccine. *Vaccines* May 10 2020;8.
- [16] Suschak JJ, Schmaljohn CS. Vaccines against ebola virus and marburg virus: recent advances and promising candidates 2019;15:2359–77.
- [17] Abdelmageed MI, Abdelmoneim AH. Design of a multipeptidopeptide-based peptide vaccine against the E protein of human COVID-19: An immunoinformatics approach 2020;2020:2683286.
- [18] Mohammed AA, Shantier SW. Epitope-based peptide vaccine against glycoprotein G of nipah henipavirus using immunoinformatics approaches, vol. 2020; 2020. p. 2567957.
- [19] Abdelmoneim AH, Mustafa MI. Immunoinformatics design of multipeptidopeptide-based universal cancer vaccine using matrix metalloproteinase-9 protein as a target Mar 2021;44:35–52.
- [20] Anwar S, Mourosi JT, Khan MF, Hosen MJ. Prediction of epitope-based peptide vaccine against the chikungunya virus by immuno-informatics approach. *Curr Pharmaceut Biotechnol* 2020;21:325–40.
- [21] Patronov A, Doychinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* 2013;3:120139.
- [22] Sridhar S. Clinical development of Ebola vaccines. *Therapeut Adv Vaccines* 2015;3 (5–6):125–38.
- [23] Hoenen T, Groseth A, Feldmann H. Current ebola vaccines. *Expert Opin Biol Ther* 2012;12(7):859–72.
- [24] Oliveira FM, Coelho IE, Lopes MD, Taranto AG, Junior MC, Santos LL, Vil lar JA, Fonseca CT, Lopes DD. The use of reverse vaccinology and molecular modeling associated with cell proliferation stimulation approach to select promiscuous epitopes from *Schistosoma mansoni*. *Appl Biochem Biotechnol* 2016;179(6): 1023–40.
- [25] Lohia N, Baranwal M. Immune response of highly conserved influenza A virus matrix 1 peptides: matrix 1 peptides for influenza vaccine. *Microbiol Immunol* 2017;61(6):225–31.

- [26] UniProt. The universal protein knowledgebase. *Nucleic Acids Res* 2017;45: D158–d169.
- [27] Doychinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* Jan 5 2007;8(4).
- [28] Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinform* 2002;2.
- [29] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary Genetics analysis across computing platforms. *Mol Biol Evol* Jun 1 2018;35: 1547–9.
- [30] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope Database (IEDB): 2018 update. *Nucleic Acids Res* Jan 8 2019;47: D339–d343.
- [31] Larsen MV, Lundsgaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf* Oct 31 2007;8:424.
- [32] Nielsen M, Lundsgaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* May 2003;12:1007–17.
- [33] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999;112:531–52.
- [34] Dimitrov I, Naneva L, Doychinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* Mar 15 2014;30:846–51.
- [35] Dimitrov I, Flower DR, Doychinova I. AllerTOP—a server for in silico prediction of allergens. *BMC Bioinf* 2013;14(Suppl 6):S4.
- [36] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* Dec 1997;18:2714–23.
- [37] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* Dec 2009;30:2785–91.
- [38] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998;19:1639–62.
- [39] Stierand K, Rarey M. PoseView – molecular interaction patterns at a glance. *J Cheminf* 2010;2:P50. P50.
- [40] Fährholfs R, Bietz S, Flachsenberg F, Meyer A, Nittlinger E, Otto T, et al. ProteinsPlus: a web portal for structure analysis of macromolecules. *Nucleic Acids Res* Jul 3 2017;45:W337–w343.
- [41] Benkert P, Tosatto SC, Schomburg D. QMEAN: a comprehensive scoring function for model quality assessment. *Proteins* Apr 2008;71:261–77.
- [42] Dash R, Hosen SM, Sultana T, Junaid M, Majumder M, Ishat IA, et al. Computational analysis and binding site identification of type III secretion system ATPase from *Pseudomonas aeruginosa*. *Interdiscip Sci* Dec 2016;8:403–11.
- [43] Saeed M, Kolokoltsov A, Freiberg A, Holbrook M, Davey R. Phosphoinositide-3 kinase-Akt pathway controls cellular entry of ebola virus. *PLoS Pathog* 2008;4: e1000141.
- [44] Kuroda M, Fujikura D, Nanbo A, Marzi A, Noyori O, Kajihara M, et al. The interaction between TIM-1 and NPC1 is important for the cellular entry of ebola virus. *J Virol* 2015;89:4–8.
- [45] Saeed MF, Kolokoltsov AA, Albrecht T, Davey RA. Cellular entry of ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS Pathog* 2010;6:e1001110.
- [46] Olivier R, Volchkov V. Characterization of a novel neutralizing monoclonal antibody against ebola virus GP. *J Infect Dis* 2015;212(Suppl 2).
- [47] Dye J, Chandran K, Kawaoka Y, Zeitlin L, Ward A, Nykiforuk C, et al. E-105 Antibodies against ebola virus. *JAIDS* 2016;71:58.
- [48] Mendoza E, Racine T, Kobinger G. The ongoing evolution of antibody-based treatments for Ebola virus infection. *Immunotherapy* 2017;9:435–50.
- [49] Davidson E, Bryan C, Fong R, Barnes T, Pfaff J, Mabila M, et al. Mechanism of binding to ebola virus glycoprotein by the ZMapp, ZMAb, and MB-003 cocktail antibodies. *J Virol* 2015;89:8–26.
- [50] Khan S, Khan A, Rehman AU, Ahmad I, Ullah S, Khan AA, et al. Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression. *Infect Genet Evol* Sep 2019;73:390–400.
- [51] Gul H, Ali SS, Saleem S, Khan S, Khan J, Wadood A, et al. Subtractive proteomics and immunoinformatics approaches to explore *Bartonella bacilliformis* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine. *Infect Genet Evol* Nov 2020;85:104551.
- [52] Khan S, Ali SS, Zaheer I, Saleem S, Ziaullah N Zaman, et al. Proteome-wide mapping and reverse vaccinology-based B and T cell multi-epitope subunit vaccine designing for immune response reinforcement against *Porphyromonas gingivalis*. *J Biomol Struct Dyn* Sep 14 2020;1–15.
- [53] Rappuoli R, Bottomley MJ, D’Oro U, Finco O, De Gregorio E. Reverse vaccinology 2.0: human immunology instructs vaccine antigen design. *J Exp Med* Apr 4 2016; 213:469–81.
- [54] Bruno L, Cortese M, Rappuoli R, Merola M. Lessons from Reverse Vaccinology for viral vaccine design. *Curr Opin Virol* Apr 2015;11:89–97.
- [55] Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin Microbiol Infect* Oct 2012;18(Suppl 5):109–16.
- [56] Kumai T, Kobayashi H, Harabuchi Y, Celis E. Peptide vaccines in cancer-old concept revisited. *Curr Opin Immunol* Apr 2017;45:1–7.
- [57] Wilson JA, Bray M, Bakken R, Hart MK. Vaccine potential of Ebola virus VP24, VP30, VP35, and VP40 proteins. *Virology* Aug 1 2001;286:384–90.
- [58] Monreal-Escalante E, Ramos-Vega AA, Salazar-González JA, Bañuelos-Hernández B, Angulo C, Rosales-Mendoza S. Expression of the VP40 antigen from the Zaire ebolavirus in tobacco plants. *Planta Jul* 2017;246:123–32.
- [59] Dudas G, Rambaut A. Phylogenetic analysis of Guinea 2014 EBOV ebolavirus outbreak. *PLoS Curr* May 2 2014;6.
- [60] Oluwagbemi O, Awe O. A comparative computational genomics of Ebola Virus Disease strains: in-silico Insight for Ebola control. *Inform Med Unlocked* 2018;12: 106–19.
- [61] Baize S, Pannetier D, Oestreich L, Rieger T, Koivogui L, Magassouba N, et al. Emergence of Zaire ebola virus disease in Guinea. *N Engl J Med* Oct 9 2014;371: 1418–25.
- [62] Wilson CC, McKinney D, Anders M, MaWhinney S, Forster J, Crimi C, et al. Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1. *J Immunol* Nov 15 2003;171: 5611–23.
- [63] Vani J, Shaila MS, Chandra NR, Nayak R. A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of *Mycobacterium tuberculosis*. *Microb Infect* Mar 2006;8:738–46.
- [65] Lim WC, Khan AM. Mapping HLA-A2, -A3 and -B7 supertype-restricted T-cell epitopes in the ebolavirus proteome. *BMC Genom* 2018;19(42).
- [66] Tshabalala M, Mellet J, Pepper MS. Human leukocyte antigen diversity: a southern African perspective. *J Immunol Res* 2015:746151.
- [67] Falk K, Lau JM, Santambrogio L, Esteban VM, Puentes F, Rotzsche O, et al. Ligand exchange of major histocompatibility complex class II proteins is triggered by H-bond donor groups of small molecules. *J Biol Chem* Jan 25 2002;277:2709–15.