Research article

Epitope-Based Peptide Vaccine against Bombali Ebolavirus Viral Protein 40: An Immunoinformatics Combined with Molecular Docking Studies

Mujahed I. Mustafa¹, Shaza W. Shantier², Miyssa I. Abdelmageed³, and Abdelrafie M. Makhawi¹.

¹Department of Biotechnology, University of Bahri, Khartoum, Sudan.

²Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.

³Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.

Correspondence should be addressed to Mujahed I. Mustafa;

mujahedimustafa@gmail.com

Abstract:

Background: Bombali Ebolavirus is RNA viruses belong to the Filoviridae family. They are causing lethal hemorrhagic fever with high mortality rate. Despite having available molecular knowledge of this virus, no approved vaccine or antiviral drugs have been developed yet for the eradication of Bombali Ebolavirus infections in humans.

Objective: the present study described a multi epitope-based peptide vaccine against Bombali Ebolavirus matrix protein VP40, using several immunoinformatics tools.

Materials and Methods: The six strains of Ebolavirus were retrieved from NCBI and Uniprot databases and submitted to VaxiJen to identify the most antigenic protein among all. Then PSIPRED, SOPMA, QMEAN, and PROCHECK tools were used to check the protein quality. T-cell prediction, population coverage, and molecular docking analysis were achieved to select peptides containing multiple Bombali VP40 epitopes showing interaction with multiple HLA molecules for expected immune response across the world. **Result:** Bombali Ebola (YP_009513276.1) was found to be the most antigenic protein among all. Which it has been used in all required analysis. For T cell three epitopes showed high affinity to MHC class I (YSFDSTTAA, VQLPQYFTF, and MVNVISGPK) and high population coverage against Africa and the world. Furthermore in MHC class II, six promising epitopes that associated with most common MHC class II alleles. **Conclusion:** The above result conclude that, these peptides capable of provoking T-cell response and being interacted with a wide range of HLA molecules have a strong potential to be a vaccine against Bombali Ebolavirus.

Keywords: Bombali Ebolavirus, matrix protein, Viral Protein (VP40), vaccine design, Immunoinformatics.

1- Introduction:

Ebola viruses are RNA viruses belong to the *Filoviridae* family [1, 2]. They are causing lethal hemorrhagic fever with high mortality rate [3-5]. The main source of the ebolavirus is unknown, but several mammal groups are the source from which it is transferred to the human species [6]. In addition to Africa, it has been reported in the United Kingdom, and Italy, and secondary cases (Health-care workers) have occurred in the United States and Spain [7, 8].

It 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].

So far, six different strains of EBOV have been identified: Sudan (SEBOV), Ivory Coast, Zaire (ZEBOV), Reston, Bundibugyo and the most recently identified strain Bombali virus. Among them, ZEBOV is considered as the most frequently occurring infection with the highest number of deaths. Ebola viruses infection is considered to be the most deadly one having mortality rate as high as 90% in human cases only [10]. Due to this high mortality rates, lack of treatment options and vaccination make Ebola virus a biothreat pathogen [11].

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]. The matrix protein 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 [13]. With the available knowledge, several attempts have been made to produce affective vaccine against Ebola viruses due to its deadly nature.

Vaccine design methods based on T-cell epitope can be characterized as the recognition of immune-dominant epitopes of the virus and synthesizing it to be utilized as vaccines to induce effective immune response. These assessments enhance the possibility of an ideal vaccine candidate. Computer-based prediction tools reduce the number of validation experiments and time for epitope prediction. The method of epitope-based

vaccine design has been performed against several life-threatening diseases [14-17]. 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 [18].

Development of an effective and safe Ebola virus vaccine has been hindered by a lack of knowledge. In this study, we attempted to identify major immunogenic epitopes on Ebola viral proteins and predict a vaccine. Simultaneously, we also achieved a genome wide search to recognize the most appropriate vaccine target site by using *in silico* tools. The study will enhance further upcoming laboratory-based approach to develop effective vaccine against Bombali Ebolavirus infection.

2. Material and methods:

The immunoinformatics approaches used for vaccine design are illustrated in **Figure 1**.

2.1. Sequence retrieval:

The FASTA format of the reference sequence of 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), alongside 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 matrix protein (VP40) of Zaire ebolavirus (Q77DJ6), Tai Forest ebolavirus (B8XCN8), Reston ebolavirus (Q8JPX9), Bundibugyo ebolavirus (B8XCM9) and Sudan ebolavirus (Q5XX06) were retrieved from UniProt (<https://www.uniprot.org/>).

The 3D structures of MHC I allele HLA-A*02:01 and MHC II alleles HLA-DRB1*07:01, HLA-DPA1*02:01/DPB1*14:01 were retrieved from RCSB PDB (protein data bank) database. Swiss PDB viewer V.4.1.0 software [19] was used for structure optimization and energy minimization.

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 VaxiJen Server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) allows the classification of antigens solely

based on the physicochemical properties of proteins without referring to sequence alignment [20].

2.3. Conservation region analysis

Pairwise and multiple sequence alignment of the six strains of Ebola virus were performed using 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. Maximum likelihood phylogenetic tree was constructed with bootstrapping value of 300.

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 [21]. The analysis was performed using Molecular Evolutionary Genetics Analysis (MEGA-X) (version 10.1.18) (<https://www.megasoftware.net>). It is a software applied for sequence alignment, inferring phylogenetic trees, and testing evolutionary relationship. It can analyze DNA, RNA and Protein sequences [22].

2.5. Secondary structure prediction, homology modeling and protein validation:

The secondary structures of the vaccine design were predicted using PSIPRED [23] by submitting the amino acids sequence as an input. PSIPRED is a secondary structure prediction online tool used to predict the transmembrane topology, transmembrane helix, fold and domain recognition. Its result was confirmed by another secondary structures online tool, SOPMA [24]. While the tertiary structure of Bombali Ebolavirus matrix protein was generated using SWISS-Model server [19] because there is no structure available in Uniprot database [25]. Discovery studio 2020 [26] was used to visualize the most promising peptides for vaccine design. The confirmation of the predicted 3D structure was done using QMEAN [27], while the protein quality was tested using PROCHECK server by Ramachandran plot [28].

2.6. T-Cell epitope prediction tools:

2.6.1. Peptide Binding to MHC Class I Molecules:

To predict the interaction with different MHC I alleles, the Major Histocompatibility Complex class I (MHC I) binding prediction tool on the IEDB <http://tools.iedb.org/mhci> [29]. 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. All alleles having a binding affinity of IC₅₀ that are equal or less than 500 nM were selected for further analysis.

2.6.2. Peptide Binding to MHC Class II Molecules:

The MHC II prediction tool provided by the IEDB was used to predict the peptide binding to MHC class II molecules, consisting of human allele references sets was used. The ANN prediction method was selected to recognize the binding affinity of MHC II grooves and MHC II binding core epitopes [30]. All epitopes that bind to many alleles at a score equal to or less than 1000, IC₅₀ were chosen for additional investigation. The Human allele reference sets (HLA DR, DP, and DQ) were included in the prediction.

2.7. Allergenicity prediction

AllergenFP v.1.0 [31] and AllerTop v 2.0 servers [32] were used to evaluate the allergenicity of the predicted peptides.

2.8. 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 [33]. All epitopes and their binding to MHC I and MHC II molecules were assessed against population covering the World and Africa.

2.9. 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.

Molecular docking was then performed using AutoDock 4.0 software [34]. 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 [35]. 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 [36] at the ProteinPlus web portal [37], respectively.

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

3. Results:

3.1. Conservation and antigenicity analysis:

The multiple sequence alignment of the six strains of Ebola revealed eight conserved regions (**Table 1**). Analysis of those conserved regions by VaxiJen revealed five sequences are antigenic as they met the criteria of default threshold ≥ 0.5 in VaxiJen (**Table 1**).

Table 1: Predicted conserved regions with their antigenicity scores by VaxiJen:

Conserved peptide	start	End	VaxiJen score
LRPIADDNIDH	51	61	0.2337
KVLMKQIPIWLPLGV	86	100	0.3173
YSFDSTTAAIMLASYTITFGK	106	127	0.6932
SNPLVRINRLG	129	139	0.5018
HPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLITQPLPAAT	145	190	0.4827
LRPGISFHPKLRPILLP	203	219	1.5593
KNIMGIEVPE	256	265	-0.1386
IIPILLPKYIG	284	294	0.6551

3.2. Phylogenetic analysis

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

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

3.3. Secondary structure prediction:

Out of the 326 amino acids analyzed by PSIPRED, 59 amino acids formed Alpha helix, 175 amino acids were involved in random coil, and 20 amino acids formed Beta turn; This result was confirmed by another secondary structures online tool, SOPMA (Figures 3, 4).

Figure 3: PSIPRED analysis of the secondary structure prediction of Bombali Ebolavirus matrix protein.

Figure 6: Assessment of structural superiority by QMEAN valuation.

3.5. T-Cell epitope prediction tools:

The MHC I and MHC II binding prediction tools predicted 102 peptides and 313 peptides from the VP40 protein that could interact with different MHC I and MHC II alleles, respectively. Tables 1 and 2 summarizes the most promising peptides bound to MHC I and MHC II alleles along their allergenicity predication.

Table 2: Predicted epitopes binding with MHC I alleles along with their allergenicity prediction:

Peptide	Alleles	Position	Allergenicity
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
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

Table 3: Predicted epitopes binding with MHC II alleles along with their allergenicity prediction:

Peptide	Alleles	Allergenicity
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
LSFHPKLRP	HLA-DPA1*02:01/DPB1*05:01, HLA-DPA1*02:01/DPB1*14:01, HLA-DQA1*01:02/DQB1*05:02, HLA-DQA1*01:02/DQB1*06:02, HLA-DQA1*02:01/DQB1*02:02, HLA-DQA1*02:01/DQB1*03:03, HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*01:03, HLA-DRB1*04:03, HLA-DRB1*04:04, HLA-DRB1*08:01, HLA-DRB1*09:01	Non-allergen
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
VISGPKVLM	HLA-DPA1*02:01/DPB1*05:01, HLA-DPA1*02:01/DPB1*14:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DQA1*02:01/DQB1*04:02, HLA-DQA1*05:01/DQB1*02:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DQA1*05:01/DQB1*03:03, HLA-DQA1*05:01/DQB1*04:02, HLA-DRB1*01:03, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:03, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:01, HLA-DQA1*03:03/DQB1*04:02	Non-allergen
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*08:02	Non-allergen

3.6. Molecular docking:

Figure 7: 3D structures for the most promising peptides bound to MHC I and MHC II alleles.

Figure 8: 2D structure for the most promising peptides bound to MHC I alleles.

Figure 9: 2D structures for the most promising peptides bound to MHC II alleles.

4. Discussion:

Ebola is a very smart RNA virus which by somehow convince the cell to engulf it, then the cell take it inside a lysosome [39]. However, Ebola needs to get out of it for two reasons, the first one is that the lysosome is filled with enzymes that can eradicates it, and all what is needed for its coping is found in the cytoplasm. The way the virus escapes is by attaching its membrane to lysosome's membrane by receiving signals from the infected cell by identifying the right place (NPC1 protein) and time to carry out this fusion reaction and release its genetic material to the cytoplasm and then initiate the replication [40, 41]. So to outsmarting ebolavirus, there are three ways to eradicates it, the first one is to go after the infected cellular machinery evolve to exploit, this is done by drug that prevent the ebolavirus from binding to NPC1; which the virus eventually get destroyed inside the lysosome because it cannot get out to replicates.

The second approach is to go after the virus itself. The glycoprotein is a highly flexible structure, thus, engineered antibody can be designed to attach to glycoprotein which will prevent the virus from binding to the host cell and eventually kill it [42-45].

The third approach is to design a vaccine which has been followed in our present study applying an immunoinformatics approach. In this study we used an immunoinformatics approach for vaccine design; we achieved protein structure-based approach for the development of multi- epitope-based peptide vaccine against Bombali Ebolavirus using matrix protein VP40 as a target.

Classifying of epitopes 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 *vitro* experiments for epitopes validation [18]. Peptide vaccines are more superior to traditional or single-epitope-based vaccines due to these factors: it contains multiple HLA epitopes than can be detected by many T-cell receptors; it contains multiple HLA epitopes than can be detected by many T-cell receptors; multiple epitopes provide a long-term memory, so increasing the virus range; additionally, designed vaccine can be associated with adjuvant thereby boosting its immunogenicity [18]. This 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 for the deadly Bombali virus infection using matrix protein VP40 as a target. It's well known that peptides can be recombinant or construct pf epitopes which are targeted against surface or intracellular proteins (such as our target) [46]. Some studies determine the ability of VP40 of Ebolavirus alone to evoke strong immune response against Ebolavirus [47, 48]. Nevertheless, there is no epitope-based peptide vaccine prediction has been done for Bombali Ebolavirus so far.

Indeed, Zaire Ebola is the strain which is responsible for most of the outbreaks with the highest case–fatality rates of all the ebolaviruses. But we are looking for the most antigenic Ebola proteins not the strain that responsible for most of outbreaks. The idea of reversed vaccinology is to replace a whole genome with small antigenic proteins [49-51], which have the ability to strongly evoke the immune system. So the antigenicity of VP40 and GP proteins from the different Ebola virus strains were analyzed by VaxiJen server (**Table 1**). Bombali ebolavirus VP40 was found to be the most antigenic (highest score = 0.5854), therefore it was used as our vaccine target. After performing multiple alignments of all the sequences, five conserved regions 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 [21]. (**Figure 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 [52], also Sudan Ebola, Reston Ebola on the same clade [53] and both Zaire and Bombali Ebola are in individual clades. Zaire strain is responsible for most of the outbreaks it's an ortholog with Bombali strain. Which supported by the high bootstrapping values on each node. As shown in (**Table 1**) the most antigenic protein among all the matrix protein of the Ebola virus, was from Bombali strains, that's why it's used as a target in our study.

The major histocompatibility complex (MHC) is a highly polymorphic group of cell-surface receptors found on all the cells of the body. 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) has the best potential to evoke a strong immune response [54, 55].

Among the analyzed antigenic sequences, two epitopes were predicted as most promising taking into consideration their conservancy and binding affinity to the highest numbers of MHC I alleles (**Table 2**). On the other hand, 5 conserved epitopes were predicted to interact with several MHC II alleles named HLA-D, and Q (**Table 3**), 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 [31] and AllerTop v 2.0 servers [32]. The result showed that all the selected peptides are non-allergen and non-toxin (**Tables 2, 3**).

Population coverage analysis was also predicted for the total peptides. 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 II alleles showed a 68.40% estimated population coverage in the world, and 85.51% in Africa.

For analyzing the stability of the vaccine, the reference sequence of Bombali Ebolavirus VP40 was used for the prediction of secondary and tertiary structure prediction by PSIPRED, and RaptorX servers respectively. A valuable insight was done from secondary structure prediction by PSIPRED. The plenty of alpha-helix and coiled region indicates high conserved regions and greater stability of VP40 protein (**Figure 3**). To validate PSIPRED's result, another secondary structure online tool (SOPMA) was used for this purpose (**Figure 4**).

The tertiary structure of Bombali Ebolavirus VP40 was generated using RaptorX sever (**Figure 5, A**), While the Ramachandran plot created by the PROCHECK server showed that about 92% of the residues of VP40 protein are located in the most favored region (**Figure 5, B**). 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 [56]. while the verification of the quality of the predicted 3D structure was done using

QMEAN. QMEAN analysis, the protein model in our interest resulted in a Z-score of -1.20, and the total score was 0.640. This value denotes a higher quality of the model, where the acceptable score ranges between 0 and 1 [38] (**Figure 6**). On the basis of the results obtained from the mentioned structural validation tools, the 3D model showed much reliability and was considered for further study.

In order to stimulate immunological responses, the predicted peptides should interact effectively with the MHC I and MHC II molecules [57]. Therefore, molecular docking was performed to study their binding affinities selecting the alleles HLA-A*02:01, HLA-B*15:01 and HLA-DRB1 as targets due to their diversity and immunogenic association [58, 59].

H-bond formation is an important parameter to deduce the stability of the conformation through the simulation period in the protein ligand complex [60]. Thus, the PoseView at the ProteinPlus web portal was used to illustrate their 2D interactions and bonding with MHC molecules. In regards to MHC I, the obtained results showed that YSFDSTTAA and VQLPQYFTF bound to the groove of HLA-A*02:01 with binding energies of -9.3 kcal/mole and -10.2 kcal/mole, respectively. Additionally, the 2D interactions viewed that YSFDSTTAA 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 (**Figure 8**). In contrast to MHC II, the most promising peptides (LPQYFTFDL, LSFHPKLRP, QEFVLPPVQ, VISGPKVLM) were bound to MHC II molecules with binding energies of -8.4kcal/mole, -7.3 kcal/mole, -7.8kcal/mole and -6.6 kcal/mole, respectively. They formed a number of hydrogen bonds (3 to 6 hydrogen bonds) with different amino acids (**Figure 9**).

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.

Conclusion:

Bombali Ebolavirus is one of the most re-emerging viruses in certain regions scatter 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 design a epitope-based peptide vaccine against Bombali Ebolavirus viral Protein 40 using an immunoinformatics approaches combined with molecular docking studies. The results are originated from a systematic analysis, which suggest that, an effective vaccine candidates linking the best fit epitopes (YSFDSTTAA, VQLPQYFTF, LPQYFTFDL, QEFVLPPVQ, VISGPKVLM, and LSFHPKLRP) 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.

Conflicts of Interest:

The authors declare that they have no conflicts of interest.

Authors' Contributions:

MIM: conceptualization, formal analysis, methodology, validation, and writing (original draft); SWS: conceptualization, formal analysis, visualization, validation, and writing (original draft); 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.

References:

- [1] B. Beer, R. Kurth, and A. Bukreyev, "Characteristics of Filoviridae: Marburg and Ebola Viruses," *Die Naturwissenschaften*, vol. 86, pp. 8-17, 02/01 1999.
- [2] F. Murphy, M. Kiley, and S. Fisher-hoch, "Filoviridae: Marburg and Ebola viruses," 08/01 2020.
- [3] H. Feldmann and T. W. Geisbert, "Ebola haemorrhagic fever," *Lancet*, vol. 377, pp. 849-862, 01/01 2011.
- [4] L. Pocock and M. Rezaeian, "Review : Ebola Haemorrhagic Fever," *Middle East Journal of Family Medicine*, vol. 12, 10/01 2014.
- [5] S. Otatume, "Ebola haemorrhagic fever in Zaire," *Journal of African Studies*, vol. 1996, pp. 95-97, 01/01 1996.
- [6] S. Badshah, A. Naeem, A. Farman, and K. Khan, "The Structure and Pathophysiology of the Ebola Virus," ed, 2020, pp. 1-14.
- [7] D. K. Evans, M. Goldstein, and A. Popova, "Health-care worker mortality and the legacy of the Ebola epidemic," *The Lancet Global Health*, vol. 3, pp. e439-e440, 2015/08/01/ 2015.
- [8] F. de la Calle-Prieto, M. Arsuaga-Vicente, M. Mora-Rillo, F. Arnalich-Fernandez, and J. R. Arribas, "[Ebola virus disease: Update]," *Enferm Infecc Microbiol Clin*, vol. 34, pp. 452-60, Aug-Sep 2016.
- [9] C. Genton, R. Cristescu, S. Gatti, F. Levréro, E. Bigot, D. Caillaud, *et al.*, "Recovery Potential of a Western Lowland Gorilla Population following a Major Ebola Outbreak: Results from a Ten Year Study," *PloS one*, vol. 7, p. e37106, 05/23 2012.
- [10] A. Sanchez, T. W. Geisbert, and H. Feldmann, "Filoviridae: Marburg and Ebola viruses," *Fields Virology*, pp. 1409-1448, 05/21 2013.
- [11] S. Jacob, I. Crozier, W. Fischer, A. Hewlett, C. Kraft, M.-A. Vega, *et al.*, "Ebola virus disease," vol. 6, p. 13, 02/20 2020.
- [12] J. J. Madara, Z. Han, G. Ruthel, B. D. Freedman, and R. N. Harty, "The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target," *Future Virol*, vol. 10, pp. 537-546, May 2015.
- [13] H. Feng, S. Nakatsu, T. Lopes, M. Imai, S. Yamayoshi, M. Yamashita, *et al.*, "Identification of Novel Adjuvants for Ebola Virus-Like Particle Vaccine," *Vaccines (Basel)*, vol. 8, May 10 2020.
- [14] M. I. Abdelmageed and A. H. Abdelmoneim, "Design of a Multiepitope-Based Peptide Vaccine against the E Protein of Human COVID-19: An Immunoinformatics Approach," vol. 2020, p. 2683286, 2020.
- [15] A. A. Mohammed and S. W. Shantier, "Epitope-Based Peptide Vaccine against Glycoprotein G of Nipah Henipavirus Using Immunoinformatics Approaches," vol. 2020, p. 2567957, 2020.
- [16] A. H. Abdelmoneim and M. I. Mustafa, "Immunoinformatics design of multiepitopes peptide-based universal cancer vaccine using matrix metalloproteinase-9 protein as a target," pp. 1-18, Jul 21 2020.
- [17] S. Anwar, J. T. Mourosi, M. F. Khan, and M. J. Hosen, "Prediction of Epitope-Based Peptide Vaccine Against the Chikungunya Virus by Immuno-informatics Approach," *Curr Pharm Biotechnol*, vol. 21, pp. 325-340, 2020.
- [18] A. Patronov and I. Doytchinova, "T-cell epitope vaccine design by immunoinformatics," *Open Biol*, vol. 3, p. 120139, Jan 8 2013.

- [19] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, pp. 2714-23, Dec 1997.
- [20] I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics*, vol. 8, p. 4, Jan 5 2007.
- [21] G. Dudas and A. Rambaut, "Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak," *PLoS Curr*, vol. 6, May 2 2014.
- [22] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0," *Mol Biol Evol*, vol. 24, pp. 1596-9, Aug 2007.
- [23] L. McGuffin, K. Bryson, and D. Jones, "The PSIPRED protein structure prediction server," *Bioinformatics*, vol. 16, pp. 404-405, 04/01 2000.
- [24] C. Geourjon and G. Deleage, "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments," *Comput Appl Biosci*, vol. 11, pp. 681-4, Dec 1995.
- [25] "UniProt: the universal protein knowledgebase," *Nucleic Acids Res*, vol. 45, pp. D158-d169, Jan 4 2017.
- [26] "Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment (2017), San Diego: Dassault Systèmes."
- [27] P. Benkert, M. Künzli, and T. Schwede, "QMEAN Server for protein model quality estimation," *Nucleic acids research*, vol. 37, pp. W510-4, 06/01 2009.
- [28] K. Gopalakrishnan, G. Sowmiya, and S. Sheik, "Ramachandran Plot on The Web (2.0)," *Protein and peptide letters*, vol. 14, pp. 669-71, 02/01 2007.
- [29] R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, *et al.*, "The Immune Epitope Database (IEDB): 2018 update," *Nucleic Acids Res*, vol. 47, pp. D339-d343, Jan 8 2019.
- [30] M. Nielsen and O. Lund, "NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction," *BMC Bioinformatics*, vol. 10, p. 296, Sep 18 2009.
- [31] I. Dimitrov, L. Naneva, I. Doytchinova, and I. Bangov, "AllergenFP: allergenicity prediction by descriptor fingerprints," *Bioinformatics*, vol. 30, pp. 846-51, Mar 15 2014.
- [32] I. Dimitrov, D. R. Flower, and I. Doytchinova, "AllerTOP--a server for in silico prediction of allergens," *BMC Bioinformatics*, vol. 14 Suppl 6, p. S4, 2013.
- [33] H. H. Bui, J. Sidney, K. Dinh, S. Southwood, M. J. Newman, and A. Sette, "Predicting population coverage of T-cell epitope-based diagnostics and vaccines," *BMC Bioinformatics*, vol. 7, p. 153, Mar 17 2006.
- [34] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, *et al.*, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J Comput Chem*, vol. 30, pp. 2785-91, Dec 2009.
- [35] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, *et al.*, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *Journal of Computational Chemistry*, vol. 19, pp. 1639-1662, 1998.
- [36] K. Stierand and M. Rarey, "PoseView -- molecular interaction patterns at a glance," *Journal of Cheminformatics*, vol. 2, pp. P50-P50, 2010.
- [37] R. Fährrolfes, S. Bietz, F. Flachsenberg, A. Meyder, E. Nittinger, T. Otto, *et al.*, "ProteinsPlus: a web portal for structure analysis of macromolecules," *Nucleic Acids Res*, vol. 45, pp. W337-w343, Jul 3 2017.
- [38] P. Benkert, S. C. Tosatto, and D. Schomburg, "QMEAN: A comprehensive scoring function for model quality assessment," *Proteins*, vol. 71, pp. 261-77, Apr 2008.

- [39] M. Saeed, A. Kolokoltsov, A. Freiberg, M. Holbrook, and R. Davey, "Phosphoinositide-3 Kinase-Akt Pathway Controls Cellular Entry of Ebola Virus," *PLoS pathogens*, vol. 4, p. e1000141, 09/01 2008.
- [40] M. Kuroda, D. Fujikura, A. Nanbo, A. Marzi, O. Noyori, M. Kajihara, *et al.*, "The Interaction between TIM-1 and NPC1 Is Important for the Cellular Entry of Ebola Virus," *Journal of virology*, vol. 89, 04/08 2015.
- [41] M. F. Saeed, A. A. Kolokoltsov, T. Albrecht, and R. A. Davey, "Cellular Entry of Ebola Virus Involves Uptake by a Macropinocytosis-Like Mechanism and Subsequent Trafficking through Early and Late Endosomes," *PLOS Pathogens*, vol. 6, p. e1001110, 2010.
- [42] R. Olivier and V. Volchkov, "Characterization of a Novel Neutralizing Monoclonal Antibody Against Ebola Virus GP," *The Journal of infectious diseases*, vol. 212 Suppl 2, 08/01 2015.
- [43] J. Dye, K. Chandran, Y. Kawaoka, L. Zeitlin, A. Ward, C. Nykiforuk, *et al.*, "E-105 Antibodies against ebola virus," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 71, p. 58, 01/01 2016.
- [44] E. Mendoza, T. Racine, and G. Kobinger, "The ongoing evolution of antibody-based treatments for Ebola virus infection," *Immunotherapy*, vol. 9, pp. 435–450, 04/30 2017.
- [45] E. Davidson, C. Bryan, R. Fong, T. Barnes, J. Pfaff, M. Mabila, *et al.*, "Mechanism of Binding to Ebola Virus Glycoprotein by the ZMapp, ZMAb, and MB-003 Cocktail Antibodies," *Journal of virology*, vol. 89, 08/26 2015.
- [46] T. Kumai, H. Kobayashi, Y. Harabuchi, and E. Celis, "Peptide vaccines in cancer-old concept revisited," *Curr Opin Immunol*, vol. 45, pp. 1-7, Apr 2017.
- [47] J. A. Wilson, M. Bray, R. Bakken, and M. K. Hart, "Vaccine potential of Ebola virus VP24, VP30, VP35, and VP40 proteins," *Virology*, vol. 286, pp. 384-90, Aug 1 2001.
- [48] E. Monreal-Escalante, A. A. Ramos-Vega, J. A. Salazar-González, B. Bañuelos-Hernández, C. Angulo, and S. Rosales-Mendoza, "Expression of the VP40 antigen from the Zaire ebolavirus in tobacco plants," *Planta*, vol. 246, pp. 123-132, Jul 2017.
- [49] R. Rappuoli, M. J. Bottomley, U. D'Oro, O. Finco, and E. De Gregorio, "Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design," *J Exp Med*, vol. 213, pp. 469-81, Apr 4 2016.
- [50] L. Bruno, M. Cortese, R. Rappuoli, and M. Merola, "Lessons from Reverse Vaccinology for viral vaccine design," *Curr Opin Virol*, vol. 11, pp. 89-97, Apr 2015.
- [51] K. L. Seib, X. Zhao, and R. Rappuoli, "Developing vaccines in the era of genomics: a decade of reverse vaccinology," *Clin Microbiol Infect*, vol. 18 Suppl 5, pp. 109-16, Oct 2012.
- [52] O. Oluwagbemi and O. Awe, "A comparative computational genomics of Ebola Virus Disease strains: In-silico Insight for Ebola control," *Informatics in Medicine Unlocked*, vol. 12, pp. 106-119, 2018/01/01/ 2018.
- [53] S. Baize, D. Pannetier, L. Oestereich, T. Rieger, L. Koivogui, N. Magassouba, *et al.*, "Emergence of Zaire Ebola virus disease in Guinea," *N Engl J Med*, vol. 371, pp. 1418-25, Oct 9 2014.
- [54] C. C. Wilson, D. McKinney, M. Anders, S. MaWhinney, J. Forster, C. Crimi, *et al.*, "Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1," *J Immunol*, vol. 171, pp. 5611-23, Nov 15 2003.
- [55] J. Vani, M. S. Shaila, N. R. Chandra, and R. Nayak, "A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of Mycobacterium tuberculosis," *Microbes Infect*, vol. 8, pp. 738-46, Mar 2006.

- [56] R. Dash, S. M. Hosen, T. Sultana, M. Junaid, M. Majumder, I. A. Ishaq, *et al.*, "Computational Analysis and Binding Site Identification of Type III Secretion System ATPase from *Pseudomonas aeruginosa*," *Interdiscip Sci*, vol. 8, pp. 403-411, Dec 2016.
- [57] K. L. Rock, E. Reits, and J. Neefjes, "Present Yourself! By MHC Class I and MHC Class II Molecules," *Trends Immunol*, vol. 37, pp. 724-737, Nov 2016.
- [58] W. C. Lim and A. M. Khan, "Mapping HLA-A2, -A3 and -B7 supertype-restricted T-cell epitopes in the ebolavirus proteome," *BMC Genomics*, vol. 19, p. 42, 2018/01/19 2018.
- [59] M. Tshabalala, J. Mellet, and M. S. Pepper, "Human Leukocyte Antigen Diversity: A Southern African Perspective," *J Immunol Res*, vol. 2015, p. 746151, 2015.
- [60] K. Falk, J. M. Lau, L. Santambrogio, V. M. Esteban, F. Puentes, O. Rotzschke, *et al.*, "Ligand exchange of major histocompatibility complex class II proteins is triggered by H-bond donor groups of small molecules," *J Biol Chem*, vol. 277, pp. 2709-15, Jan 25 2002.