Indonesian Journal of Pharmacy

VOL 34 (2) 2023: 193–204 | RESEARCH ARTICLE

B-Cell Conserved Epitope Screening and *In Silico* Cloning of Envelope Glycoprotein from Ebola Virus (EBOV) For Vaccine Candidate Construction

Tania Widjaja[1], Arif Nur Muhammad Ansori[2], Viol Dhea Kharisma[3], Irvan Faizal[1,4], Yulanda Antonius[5], Joko Pebrianto Trinugroho[6], Rasyadan Taufiq Probojati[3], Muhammad Hermawan Widyananda[3,7], Pavel Burkov[8], Pavel Scherbakov[9], Vera Gribkova[10], Natalia Nikolaeva[10], Natalia Vasilievich[10], Vikash Jakhmola[11], Md. Emdad Ullah[12], Arli Aditya Parikesit[13], Rahadian Zainul[14]*

1. Faculty of Biotechnology, Atma Jaya Catholic University of Indonesia, Jl. Jend. Sudirman No.51, RW.4, Karet Semanggi, 12930, Jakarta, Indonesia
2. Doctoral Program in Veterinary Science, Faculty of Veterinary Medicine, Universitas Airlangga, Faculty of Venetery Medicine, Universitas Airlangga, Jl. Mulyorejo, Kec. Mulyorejo, Surabaya, East Java 60115
3. Computational Virology Research Unit, Division of Molecular Biology and Genetics, Generasi Biologi Indonesia Foundation, Jl. Swadaya Barat No.24, 61171 Gresik, Indonesia
4. Center of Technology of Pharmaceutical and Medicinal, Agency for the Assessment and Application of Technology (BPPT), Jakarta, Indonesia
5. Faculty of Biotechnology, University of Surabaya, Baratajaya, Gubeng, 60284 Surabaya, Indonesia.
6. Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Jalan Kebon Jeruk Raya No. 27, Jakarta 11530 Indonesia
7. Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Jl. Veteran, Malang, Jawa Timur 65145 Indonesia
8. Center for Biotechnology of Animal Reproduction, South Ural State Agrarian University, Troitsk, Russian Federation, 13 Gagarin St., 457100 Troitsk, Russia
9. Department of Infectious Diseases, South Ural State Agrarian University, Troitsk, Russian Federation, 13 Gagarin St., 457100 Troitsk, Russia
10. Institute of Technology of Food Production and Technological Management, K.G. Razumovsky Moscow State University of Technologies and Management (The First Cossack University), Moscow, Russian Federation
11. Uttaranchal Institute of Pharmaceutical Sciences, Uttaranchal University, Dehradun, Chakrata Rd, Prem Nagar, Dehradun, Uttarakhand 248007, India
12. Department of Chemistry, Mississippi State University, Mississippi State, MS 39762-9573, United States
13. Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Jl. Pulomas Barat No.Kav. 88, RT.4/RW.9, Kayu Putih, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13210 Indonesia
14. Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Padang, Jalan Prof. Dr. Hamka, Air Tawar Padang, Sumatera Barat Indonesia

Article Info

Submitted: 10-03-2022
Revised: 02-03-2023
Accepted: 15-04-2023

*Corresponding author
Rahadian Zainul

Email:
rahadianzmsiphd@fmipa.unp.ac.id

ABSTRACT

Ebola virus (EBOV) is a type of RNA virus from the family of Filoviridae. The 2014-2016 Ebola outbreak in African countries Guinea, Liberia, and Sierra Leone has a total of 28,616 cases and 11,310 deaths. Death from Ebola is mainly caused by multi-organ failure due to internal bleeding and fluid loss. Another Ebola outbreak spiked this February 2021, suggesting the low effectiveness of the previous vaccines used. Zaire Ebola virus (EBOV) is known to be the species highly involved in the recent outbreak with a high mortality rate. This study is carried out to design a B-cell epitope Ebola vaccine based on the conserved region of Zaire EBOV glycoprotein. Reverse vaccinology and immunoinformatics approaches are used in this study. Samples of Zaire EBOV glycoprotein sequences were retrieved from GenBank, NCBI. The 3D modeling was done using the SWISS-MODEL web server and PyMol software. Phylogenetic tree analysis was also done using MEGA X. B-cell epitope prediction was done by BepiPred 2.0 and Emini Surface Accessibility using the IEDB web server. Epitopes were selected based on their conservancy by comparing the sequences with the MEGA X alignment result. Antigenicity, allergenicity, and toxicity properties of the peptides were predicted using VaxiJen 2.0, AllerTOP 2.0, and ToxinPred web servers. In silico cloning was done as the final step using the pET-24a(+) expression vector. This study revealed that peptides "LEIKKPD," "TGFGTNETEYLF," "PYFGPAA," "PYFGPA," and "KLSSTNQL" are the best candidate for the B-cell epitope vaccine. Phylogenetic tree and 3D modeling successfully showed the genetic and structural differences of Zaire EBOV GP originating from various countries. In silico cloning was also done using the pET28a(+) expression vector to design a clone vector map for the next vaccine development phase.

Keywords: B-Cell Epitope, Immunoinformatics, Ebola Virus, Vaccinology

INTRODUCTION

The Ebola virus (EBOV) is a type of RNA virus that belongs to the *Filoviridae* family. This virus is concerning because it is very infectious, with a high mortality rate (25-90%). Death caused by Ebola is primarily caused by hemorrhagic fever and cytokine storm that causes multi-organ failure. Six species of Ebola virus were identified, including Tai Forest EBOV, Zaire EBOV, Sudan EBOV, Bundibugyo EBOV, Reston EBOV, and Bombali EBOV. Among these species, Zaire EBOV is known to have the highest mortality rate and has caused several outbreaks (Kadanali and Karagoz, 2015; Su and Stahelin, 2020).

Seven years after the last Ebola epidemic in 2014, another outbreak spiked in February 2021. Phylogenetic analysis using the latest 2021 samples shows that the 2021 outbreak was caused by the same variants involved in the 2014 outbreak. It is also known that Zaire EBOV is the species responsible for the 2021 outbreak. This event of a recurrent outbreak is relatively concerning. Therefore, developing a new and more effective vaccine is needed to prevent possible future outbreaks (Gaurav *et al*., 2022; Mark *et al*., 2022; Keita *et al*., 2021; Raab *et al*., 2021).

Previous studies have shown a moderate to high mutation rate on several EBOV proteins. The Zaire EBOV Makona strain, which is highly related to the recent outbreak, is known to undergo non-synonymous mutation in the nucleoprotein, polymerase, and glycoprotein genes (Maroney *et al*., 2021). It is also known that the mutation rate accelerated during the 2014 outbreak. It suggested that human-to-human transmission generated the new variants of the virus (Liu *et al*., 2015). This might explain the ineffectiveness of the vaccine over time. Therefore, new strategies need to be developed to overcome the problem. Identifying a highly conserved region of viral protein can be considered a strategy to overcome the high mutation rate (Rantam *et al*., 2021).

Epitope-based vaccines have been known to be very effective in inducing an immune response. An epitope is a specific site of a protein that can highly induce the immune response. The target protein that is usually used to develop epitope vaccines is the structural protein of the virus, especially the ones located on the outer surface of the virus (Chia *et al*., 2023; Parvizpour *et al*., 2020; Rantam *et al*., 2021). One of EBOV's surface proteins that is highly likely to interact with the host is the glycoprotein (GP) (Kadanali and Karagoz, 2015). GP facilitates the binding and fusion of viral to the host's cell membrane via viral entry (Agnolon *et al*., 2020).

The B-cell immune response is significant to avoid further infection by the Ebola virus. This is because once EBOV enters the host cell, its viral protein 24 can inhibit the further T-cell immune response by repressing IFN-induced gene, one of which is the presentation of a major histocompatibility complex (Kuhl and Pohlmann, 2012; Zhang *et al*., 2012). Once the infection has reached this stage, the virus may go further to latent infection. Latent infection is even harder to cure since the virus is able to stay viable inside the host cell for a long time and avoid an immune response (Bosworth *et al*., 2021).

Therefore, this study aimed to design a B-cell epitope vaccine based on Zaire EBOV's glycoprotein conserved region. Characteristics that determine the best candidate for B-cell epitope vaccine are the ability to induce B-cell immune response, highly conserved, non-allergenic, and non-toxic. The screening and prediction of the B-cell epitope vaccine candidate can be made through bioinformatics. Many web servers and software can be used, such as BepiPred, VaxiJen, AllerTOP, and Toxinpred. Bioinformatics is a potential approach to screen and predict vaccine candidates. It can reduce the cost, time, and effort to develop a new vaccine compared to *in vitro* or *in vivo* studies (Adianingsih and Kharisma, 2019; Ansori *et al*., 2021). This method has been used since early 2000, initially used to develop a vaccine against Meningococcus B. This method successfully develops a vaccine against Meningococcus B, which is known to have a wide variety of surface proteins with high similarity to human antigens. Since then, reverse vaccinology has been used to develop other vaccines, such as *S. pneumoniae*, Chlamydia, and even antibiotic-resistant *S. aureus* (Sette and Rappuoli, 2010). The construction of vaccines through conserved B cell epitopes on EBOV glycoproteins is expected to provide solutions related to preventing continued infection, increasing the coverage of protection in a person, and increasing effective vaccine production.

MATERIAL AND METHODS

Sample retrieval of Zaire EBOV GP

Zaire EBOV GP protein sequences are retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/) using the keyword "Zaire Ebola virus glycoprotein." Full amino acid sequences of EBOV GP were selected as a sample. Ebola virus glycoprotein has a full sequence of 676 amino acids (Lee and Saphire, 2009; Agnolon *et al*., 2020). Furthermore, samples were selected from humans (*Homo sapiens*) and two samples from *Macaca fascicularis* and *Cynomolgus macaque* to increase the variation. Bombali EBOV GP and Tai Forest EBOV GP were determined as outgroup samples for phylogenetic alignment (Kemenesi *et al*., 2018).

The 3D modeling of Zaire EBOV GP originated from various countries

Among the protein samples collected, one representative from each country's sample origin is modeled using SWISS-MODEL (https://swissmodel.expasy.org/) (Waterhouse *et al*., 2018). SWISS-MODEL is a server used to predict 3D models from a protein sequence based on homology modeling (Bienert *et al*., 2017). Models were selected based on the Ramachandran favored score of ≥80% and sequence identity of ≥95% (Maxwell and Popelier, 2017). The 3D models which had been generated were then visualized by using PyMol software v.2.x (Schrödinger, Inc., USA) (Bramucci *et al*., 2012).

Phylogenetic tree construction of Zaire EBOV GP originated from various countries

A phylogenetic tree was constructed by selecting five samples from each country originating from Africa, such as Sierra Leone, Zaire, Congo, Gabon, and Guinea. Samples with a full amino acid sequence and no undefined (X) amino acid were selected. The Bombali EBOV GP and Tai Forest EBOV GP were used as outgroups (Kemenesi *et al*., 2018). Alignment was conducted using MUSCLE alignment software (Pennsylvania State University, USA). Furthermore, the phylogenetic tree was constructed using MEGA X software with

a maximum likelihood approach and a bootstrap value of 1000. Jones-Taylor-Thomton (JTT) model was used as substitution matrices (Kyrychenko *et al*., 2017).

Multiple sequence alignment of Zaire EBOV GP

All collected Zaire EBOV GP samples were aligned using MEGA X software and MUSCLE alignment (Edgar, 2004; Kharisma and Ansori, 2020). Samples with long undefined (X) amino acid sequences were discarded. The screening of the conserved region was addressed after the alignment process.

Conserved B-cell epitope prediction

The B-cell epitope of Zaire EBOV GP was predicted using BepiPred 2.0 and Emini Surface Accessibility through the IEDB server (http://tools.iedb.org/bcell/). The BepiPred 2.0 epitope prediction was performed with a threshold value of ≥0.50. Emini Surface Accessibility epitope prediction was performed using the threshold value of ≥1.00. Among the epitope peptides, the conserved region was generated in MEGA X (Jespersen *et al*., 2017; Kharisma *et al*., 2021).

Antigenicity, allergenicity, and toxicity prediction

The epitope peptides selected from B-cell epitope prediction were then addressed for further analysis related to antigenic, allergenic, and toxicity properties. Antigenicity prediction was carried out by using VaxiJen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with a threshold value of 0.5 (Doytchinova and Flower, 2007). The antigenic peptides were categorized by VaxiJen 2.0, then predicted for allergenicity properties using AllerTOP 2.0 server (https://www.ddg-pharmfac.net/AllerTOP/) (Dimitrov *et al*., 2013). The peptides categorized as non-allergens were then addressed for toxicity prediction using ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) with an SVM threshold of -0.5 (Kharisma *et al*., 2021).

***In silico* cloning simulation**

In this study, the pET-28a(+) expression vector was selected for cloning, and its nucleotide sequences were collected from the Addgene vector database (https://www.addgene.org/vector-database/) (Kamens, 2015). Then, SnapGene software v.6.0 (GSL Biotech LLC, USA) was utilized for in silico cloning simulation of peptide-based vaccine components (Le Bert *et al*., 2020) against EBOV (Figure 5).

RESULT AND DISCUSSION

The 3D modeling of Zaire EBOV GP from various country origin

About 170 samples of Zaire EBOV GP, amino acid sequence, were successfully retrieved from the NCBI database. The samples originated from various African countries, including Sierra Leone, Zaire, the Democratic Republic of Congo, Gabon, and Guinea. One representative from each country, Sierra Leone (ASV62152.1), Zaire (ALT19718.1), Democratic Republic of Congo (AZL57318.1), Gabon (AQA27610.1), and Guinea (ARV89893.1) were modeled using SWISS-MODEL and further visualized using PyMol software. All 3D model results of these samples have a Ramachandran favored value of ≥80% and sequence identity of ≥95% (Figure 1).

SWISS-MODEL is a server based on a homology modeling approach used to generate 3D models from amino acid sequences. This approach is carried out by comparing the similarities of sample amino acid sequences with a template protein sequence that has been modeled before. The homology modeling method has an accuracy of up to 90%. Therefore, there will be a sequence similarity score between the sample and template protein (Waterhouse *et al*., 2018; Rantam *et al*., 2021). Ramachandran plot visualizes the likeliness of the amino acid to be positioned at a certain positional angle. The dark green region of the Ramachandran plot showed the most energetically probable positional angle of an amino acid; light green shows less probable; and the white region shows the non-permitted positional angle of an amino acid. The higher the Ramachandran favored score, the more amino acid residue of the peptide positioned at the dark green region (Maxwell and Popelier, 2017).

All the 3D models of Zaire EBOV GP originating from various countries have Ramachandran favored by a score of more than 80%. The 3D model of the glycoprotein of Zaire EBOV originating from Congo (Figure 1A). The Ramachandran plot shows that 84.13% of the amino acids in this model were positioned at an energetically possible angle and torsion.

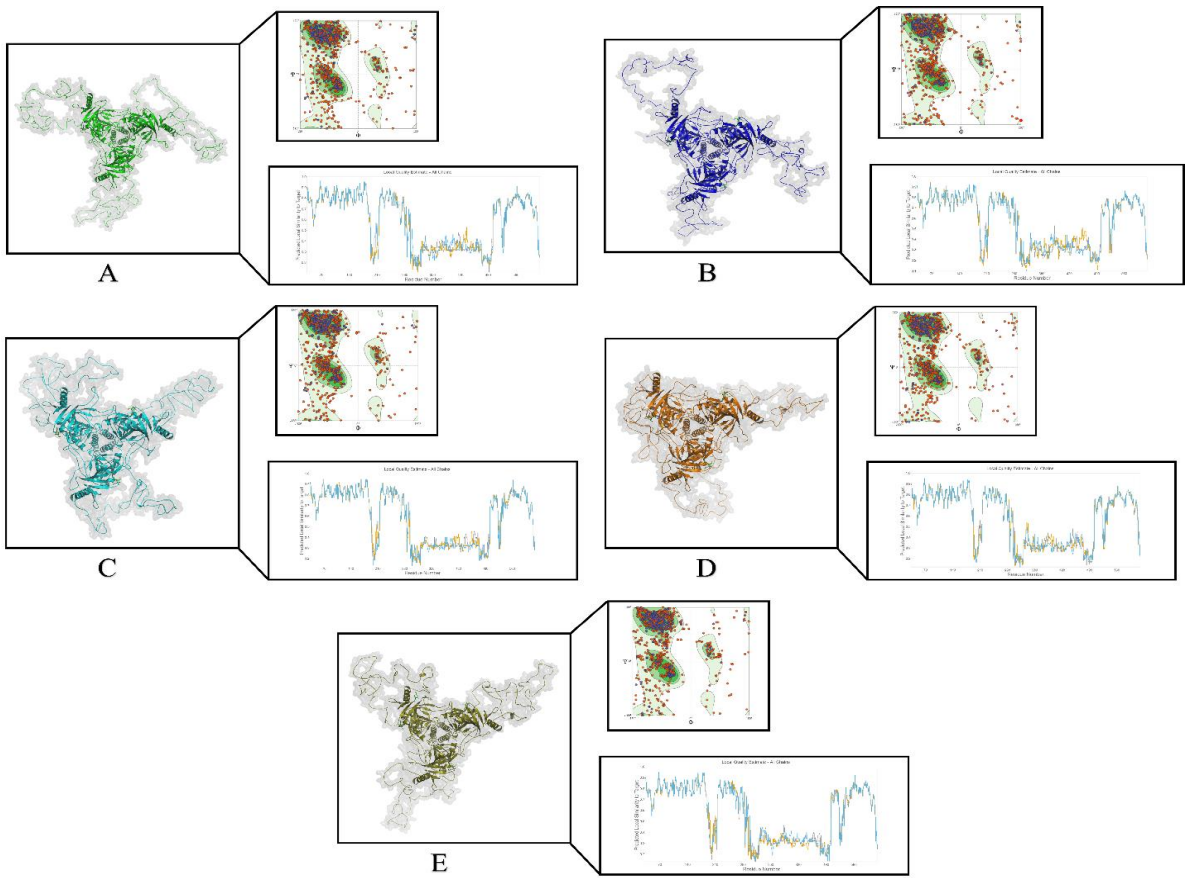

Figure 1. 3D models, Ramachandran plot, and QMean graph of Zaire EBOV GP originated from Congo (A), Gabon (B), Guinea (C), Sierra Leone (D), and Zaire (E)

And the 3D model of the glycoprotein of Zaire EBOV originating from Gabon (Figure 1B). The Ramachandran plot generated shows that 83.9% of the amino acids in this model were positioned at an energetically possible angle and torsion. The 3D model of the glycoprotein of Zaire EBOV originating from Guinea (Figure 1C). The Ramachandran plot of the model shows that 84.02% of the amino acids were positioned at an energetically possible angle and torsion. And the 3D model of the glycoprotein of Zaire EBOV originating from Sierra Leone (Figure 1D). The Ramachandran plot generated shows that 83.62% of the amino acids in this model are positioned at an energetically possible angle and torsion. The 3D model of the glycoprotein of Zaire EBOV originating from Zaire (Figure 1E). The Ramachandran plot generated shows that 84.13% of the amino acids from this model were positioned at an energetically possible angle and torsion. The QMean graph generated from each model also shows a very close similarity between the 3D model generated with its experimental reference structure; ID: 6VKM.1A from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RSCB PDB).

Values of more than 80% of Ramachandran's favored score indicate that the generated 3D models by SWISS-MODEL can be trusted (Maxwell and Popelier, 2017). All 3D modeling results of Zaire EBOV GP showed a sequence identity score of more than 95%, which means that the generated 3D models can be trusted (Waterhouse *et al*., 2018). The structural difference of the 3D models of Zaire EBOV GP from each country (Figure 1), indicating the occurrence of mutation in Zaire EBOV GP, corresponding to the findings of the previous studies (Liu *et al*., 2015; Maroney *et al*., 2021).

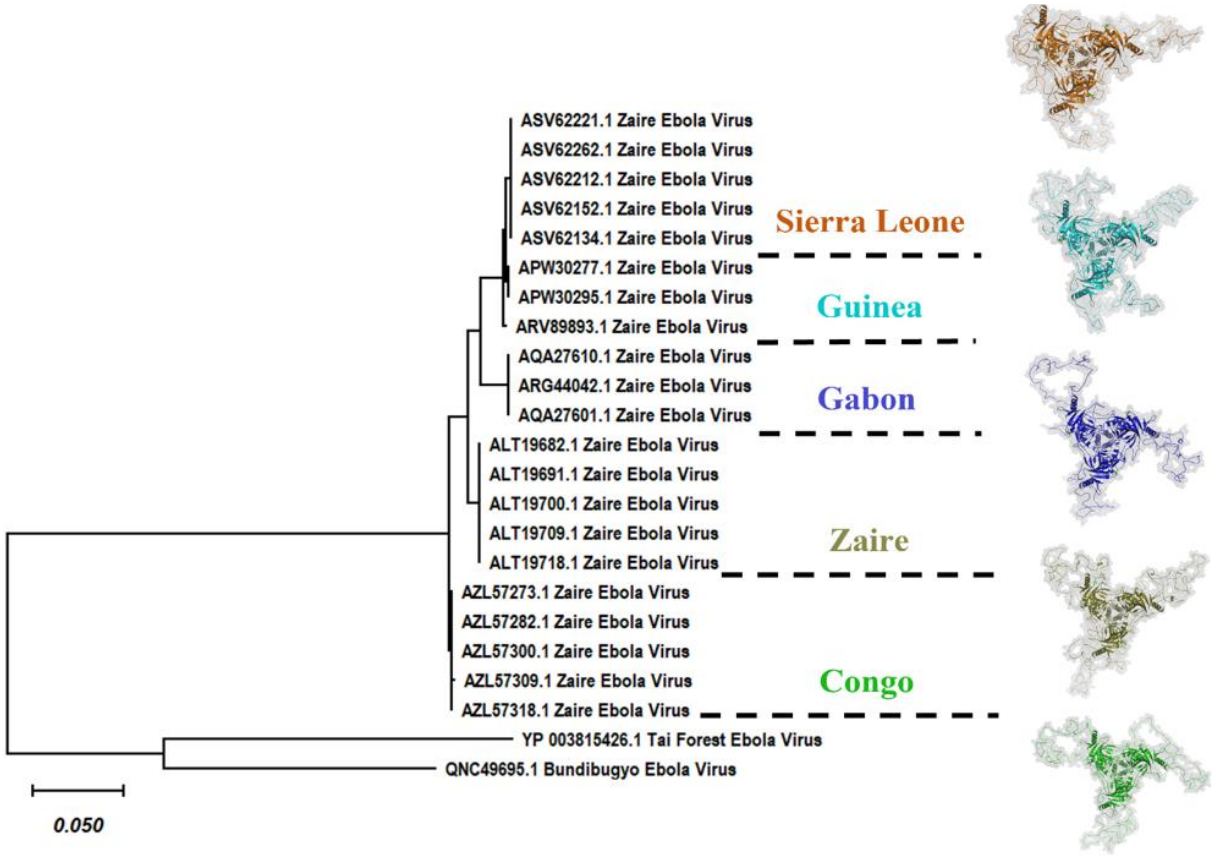

Figure 2. Phylogenetic tree result of Zaire EBOV GP originated from various countries.

Phylogenetic tree construction of Zaire EBOV GP originated from various countries

Five representative sequences of Zaire EBOV GP, originating from each country, were selected for phylogenetic tree construction. However, only three sequences each were used for Guinea and Gabon samples found in the NCBI database. Bombali EBOV and Tai Forest EBOV were included as the outgroup sample. These samples were then aligned by using MEGA X software and MUSCLE alignment. The phylogenetic tree was generated using the maximum likelihood approach, JTT substitution matrices, and 1,000 bootstrap values.

The phylogenetic tree result (Figure 2) shows that Zaire EBOV GP originated from Sierra Leone and is most related to the samples from Guinea. Zaire EBOV GP originated from Sierra Leone, and Guinea forms a clade most related to the samples from Gabon. The clade formed by Zaire EBOV GP from Sierra Leone, Guinea, and Gabon is related to the samples originating from Zaire. This clade is related to samples from Congo. Tai Forest EBOV and Bombali EBOV GP are mapped as the outgroup. It is known that there is a difference in the 3D structure of each Zaire EBOV GP from many countries. Phylogenetic tree construction here is used to further assess the genetic relationship of these samples. The bootstrap value is the number of times the phylogenetic tree construction is repeated (Ansori *et al*., 2020). The higher the Bootstrap value, the more accurate and trusted the phylogenetic tree result.

Conserved B-cell epitope prediction of Zaire EBOV GP

Multiple sequence alignment for the 170 samples was first carried out using the MEGA X software maximum likelihood approach and MUSCLE alignment. Multiple sequence alignment was performed to assess conserved regions found between nucleotide or amino acid sequences (Edgar, 2004).

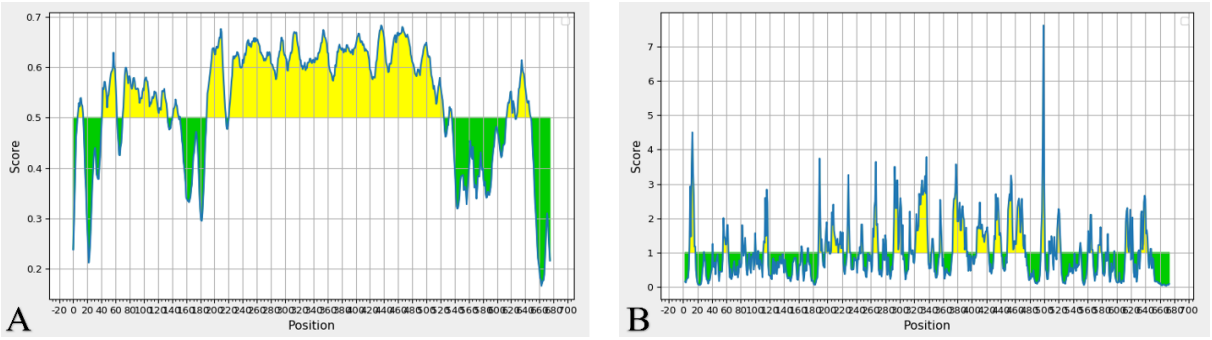

Figure 3. B-cell epitope prediction of Zaire EBOV sample originated from Sierra Leone (ASV62152.1); BepiPred 2.0 result (A) and Emini Surface Accessibility result (B).

B-cell epitope prediction was then carried out for the whole 170 samples using BepiPred 2.0 and Emini Surface Accessibility prediction algorithm. The generated epitope peptides were then observed for conservancy by comparing them with the alignment result. Peptides with less than six amino acid sequences were eliminated. In addition, peptides with one or more non-conserved amino acids were removed. The results revealed that 36 peptides were successfully identified as conserved B-cell epitopes based on BepiPred 2.0 and Emini Surface Accessibility prediction and its conservancy (Figure 3).

The B-cell epitope is a region of a protein that is most likely to bind to the antigen-binding site of an antibody. The B-cell immune response is induced by the binding of the epitope to the Fab region of the B-cell receptor, inducing the proliferation of plasma cells and the production of antibodies. This step revealed 36 conserved B-cell epitope peptides. The 36 peptides were then further analyzed for their antigenicity, allergenicity, and toxicity properties (Rantam *et al*., 2021).

Antigenicity, allergenicity, and toxicity prediction

Analysis of epitope peptide properties was carried out to select the best candidates for epitope-based vaccines among 36 selected peptides by using VaxiJen 2.0 server. VaxiJen predicted the peptide antigenicity based on the physicochemical properties of the protein (Doytchinova and Flower, 2007). Nine peptides with a VaxiJen score of ≥ 0.5 were identified as antigenic and selected for further analysis. Allergenicity prediction of the nine selected peptides was made using AllerTOP 2.0 server. AllerTOP is a server that can predict the allergenicity of a peptide according to its physicochemical properties (Dimitrov*et al*., 2013). Six peptides were identified as non-allergen and were selected for toxicity prediction using ToxinPred. It is used to predict peptide toxicity. The SVM score above the threshold is categorized as toxic (Rantam *et al*., 2021). Toxinpred is based on a dataset of 1805 toxic peptides. Its algorithm observes Cys, His, Asn, and Pro residues and their position (Gupta *et al*., 2013). After these three steps of selection of peptide properties, five peptides were found to be the best candidates for the B-cell epitope vaccine of Zaire EBOV, which has antigenic, non-allergenic, and non-toxic properties.

The list of conserved B-cell epitope peptides predicted and selected through each stage (Table I). Cells with yellow labels indicated that the peptide properties are favored and could be selected for further selection. The antigenic peptide is considered favored since it can induce the production of antibodies in the host, protecting the host from future infections. Allergenic peptides are not favored because they may induce undesired vigorous immune responses, such as severe inflammation that may harm the host (Parvizpour *et al*., 2020). Toxic peptides are not favored since they may cause harmful effects such as tissue disruption, alteration in red and white blood cell function, or increased free radical production in hosts (Khan *et al*., 2018) (Table II).

Table I. Predicted conserved B-cell epitope antigenicity, allergenicity, and toxicity properties. Cells labeled yellow indicate the selected peptides for further analysis (N = Frequency of peptide formed during epitope prediction stage).

No.	Peptide	N	Start	End	VaxiJen score	AllerTOP	Toxinpred (SVM score)
1.	QLPRDRFK	34	8	15	-1.16		
2.	SGTGPCAGD	25	142	150	-0.0309		
3.	KKDFFSSHPLREPVNATEDPSSGYYSTT	9	190	217	0.2625		
4.	PYFGP	6	533	537			
5.	WTKNITDKIDQI	14	615	626	-0.3588		
6.	KLSSTNQ	42	56	62	0.9936	Allergen	
7.	LEIKKPD	42	111	117	1.005	Non-allergen	-0.89
8.	HPLREP	39	197	202	-0.8441		
9.	ATEDPSSGYY	40	205	214	0.3053		
10.	TTIRYQ	45	216	221	1.4202	Non-allergen	-0.24
11.	TNETEY	3	227	232	0.6535	Allergen	
12.	LANETTQ	3	561	567	0.4374		
13.	HDWTKNI	30	613	619	-0.9592		
14.	TLPDQGDNDN	45	634	643	0.2449		
15.	GP	2	536	537			
16.	GTC	11	599	601			
17.	DWTKNITDKIDQIIHDFVDKTLPDQGDNDNW	7	614	644	-0.0865		
18.	LQLPRDRFK	7	7	15	-0.9687		
19.	SGTGPCAG	19	142	149	-0.1069		
20.	YFGP	8	534	537			
21.	DWTKNITDKIDQI	13	614	626	-0.1629		
22.	VDKTLPDQGDNDNWWTG	27	631	647	-0.0847		
23.	KKDFFSSHPLREPVNATEDPSSGYYS	2	190	215	0.2382		
24.	YFGPA	9	534	538			
25.	DWTKNITDKIDQIIHDFVDKTLPDQGDNDNWWT	7	614	646	-0.1380		
26.	KKDFFSSHPLREPVNATEDPSSGYYST	1	190	216	0.2218		
27.	YFGPAA	1	534	539	1.3027	Allergen	
28.	ILQLPRDRFK	3	6	15	-0.7355		
29.	TEDPSSGYY	5	206	214	0.322		
30.	TC	4	600	601			
31.	DWTKNITDKIDQII	4	614	627	-0.2577		
32.	DFVDKTLPDQGDNDNW	1	629	644	0.4294		
33.	KKDFFSSHPLREPVNATEDPSSGYY	1	190	214	0.1904		
34.	TGFGTNETEYLF	1	223	234	0.9518	Non-allergen	-0.91
35.	PYFGPAA	2	533	539	1.0874	Non-allergen	-1.06
36.	3EIKKPD	3	112	117	0.4369		
37.	QLPRDRF	1	8	14	-1.9067		
38.	SGTGPCA	1	142	148	0.0869		
39.	VDKTLPDQGDNDNWWT	1	631	646	-0.0763		
40.	KDFFSSHPLREPVNATEDPSSGYYSTT	1	191	217	0.1338		
41.	PYFGPA	3	533	538	0.9202	Non-allergen	-0.93
42.	KLSSTNQL	3	56	63	0.5835	Non-allergen	-0.94

Table II. Final peptide candidates of Zaire EBOV B-cell epitope vaccine (N= Frequency of peptide formed during epitope prediction stage).

No.	Peptide	N	Start	End	VaxiJen score	AllerTOP	Toxinpred	SVM score
7.	LEIKKPD	42	111	117	1.005 (Antigen)	Non-allergen	-0.89	Non-toxin
34.	TGFGTNETEYLF	1	223	234	0.9518 (Antigen)	Non-allergen	-0.91	Non-toxin
35.	PYFGPAA	2	533	539	1.0874 (Antigen)	Non-allergen	-1.06	Non-toxin
41.	PYFGPA	3	533	538	0.9202 (Antigen)	Non-allergen	-0.93	Non-toxin
42.	KLSSTNQL	3	56	63	0.5835 (Antigen)	Non-allergen	-0.94	Non-toxin

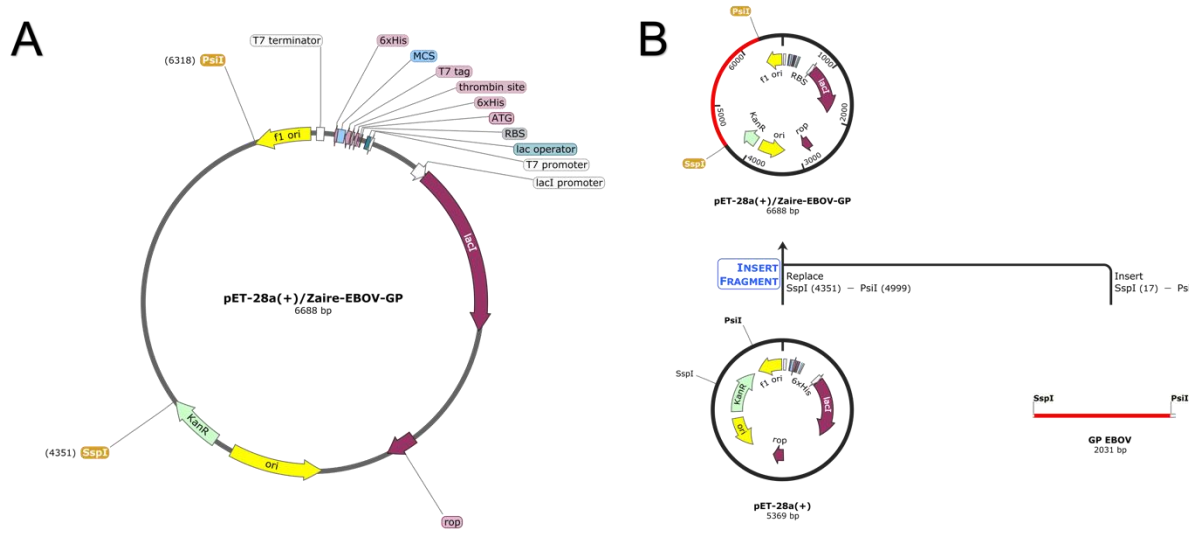

Figure 4. *In silico* cloning of Zaire EBOV GP protein gene in pET-28a(+) expression vector. (A) Cloned map of pET-28a(+)/Zaire-EBOV-GP; (B) The process of pET-28a(+)/Zaire-EBOV-GP construction. All figures were generated by using SnapGene software v.6.0.

In silico cloning

Simulation of in silico cloning was performed to develop a design of an expression vector containing the vaccine epitope of Zaire EBOV GP as the insert. The pET28a(+) has been used in many studies as an expression vector since it can be expressed in the *Escherichia coli* system. The pET28a(+) plasmid also had a kanamycin resistance gene and *LacI* gene to facilitate the selection and differentiation of *E.coli* transformants (Khatoon *et al.*, 2017; Pandey *et al.*, 2018). SnapGene is a software that facilitates *in silico* restriction cloning (Khatoon *et al.*, 2017; Ghosh *et al.*, 2021). *Ssp*I and *Psi*I restriction sites were first digested in the pET28a(+) vector, and the inserted gene was then cloned into the vector. In this study, we successfully inserted the Zaire EBOV GP protein gene (2031 bp) into the pET-28a(+) expression vector (5369 bp) (considering *Ssp*I and *Psi*I restriction sites). Thus, the total size of the plasmid expression vector (pET-28a(+)/Zaire-EBOV-GP) is 6688 bp. Figure 4 displays the in silico cloning results.

CONCLUSION

In silico approach for vaccine development could increase the efficiency of speed, cost, and safety aspects. It also can narrow down the number of candidate peptides before the in vitro and in vivo studies. In terms of safety, in silico vaccine design can reduce the risks of exposure to highly dangerous pathogens due to the need for pathogen cultivation when using traditional approaches. Therefore, it is true that a vaccine design based on a highly conserved region is needed. This study reveals that peptides 7, 34, 35, 41, and 42 are the best peptide candidates from Zaire EBOV GP to be used as conserved B-cell epitope vaccines.

Table II. Final peptide candidates of Zaire EBOV B-cell epitope vaccine (N= Frequency of peptide formed during epitope prediction stage).

No.	Peptide	N	Start	End	VaxiJen score	AllerTOP	Toxinpred	SVM score
7.	LEIKKPD	42	111	117	1.005 (Antigen)	Non-allergen	-0.89	Non-toxin
34.	TGFGTNETEYLF	1	223	234	0.9518 (Antigen)	Non-allergen	-0.91	Non-toxin
35.	PYFGPAA	2	533	539	1.0874 (Antigen)	Non-allergen	-1.06	Non-toxin
41.	PYFGPA	3	533	538	0.9202 (Antigen)	Non-allergen	-0.93	Non-toxin
42.	KLSSTNQL	3	56	63	0.5835 (Antigen)	Non-allergen	-0.94	Non-toxin

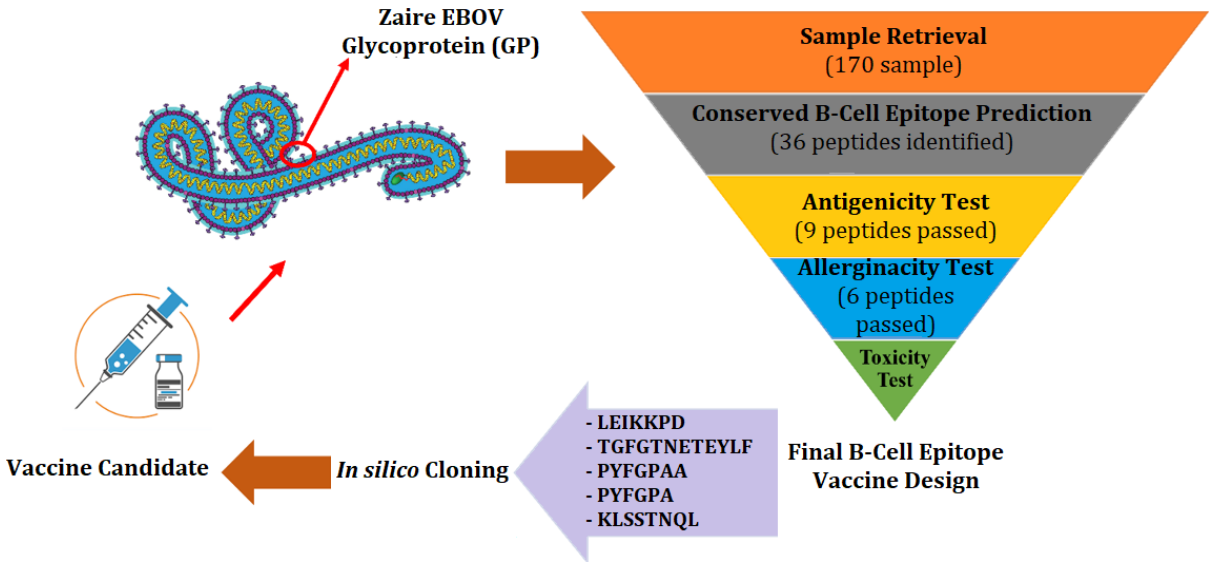

Figure 5. Conserved B-cell epitope vaccine selection stages.

In silico cloning was also successfully performed to be further used for in vitro and in vivo studies to develop an Ebola virus vaccine that is hopefully more effective in preventing future outbreaks (Figure 5).

ACKNOWLEDGEMENT

This study is supported by the Universitas Negeri Padang, Padang, Indonesia.

REFERENCES

Adianingsih, O. R., *and* Kharisma, V. D. 2021. Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. *Journal of Applied Pharmaceutical Science, 9*(1), 98-103.

Agnolon, V., Kiseljak, D., Wurm, M. J., Wurm, F. M., Foissard, C., Gallais, F., Wehrle, S., Munoz-Fontela, C., Bellanger, L., Correia, B. E., Corradin, G., *and* Spertini, F. 2020. Designs and characterization of subunit Ebola GP vaccine candidates: implications for immunogenicity. *Frontiers in Immunology, 11*(586595).

Ansori, A. N. M., Kharisma, V. D., Antonius, A., Tacharina, M. R., *and* Rantam, F. A. 2020. Immunobioinformatics analysis and phylogenetic tree construction of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Indonesia: spike glycoprotein gene. *Jurnal Teknologi Laboratorium*, *9*(1), 13-20.

Ansori, A. N. M., Kharisma, V. D., Fadholly, A., Tacharina, M. R., Antonius, Y., *and* Parikesit, A. A. 2021. Severe acute respiratory syndrome coronavirus-2 emergence and its treatment with alternative medicines: A review. *Research Journal of Pharmacy and Technology*, *14*(10), 5551-5557.

Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., *and* Schwede, T. 2017. The SWISS-MODEL Repository-new

features and functionality. *Nucleic Acids Research, 45*(D1), D313–D319.

Bosworth, A., Rickett, N. Y., Dong, X. F., Ng, L. F. P., Dorival, I. G., Matthews, D. A., Fletcher, T., Jacobs, M., Thomson, E. C., Carroll, M. W., *and* Hiscox, J. A. 2021. Analysis of an Ebola virus disease survivor whose host and viral markers were predictive of death indicates the effectiveness of medical countermeasures and supportive care. *Genome Medicine, 13*(5).

Bramucci, E., Paiardini, A., Bossa, F., *and* Pascarella, S. 2012. PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics, 13*(4).

Chia, L. Y., Kumar, P. V., Maki, M. A. A., Ravichandran, G., Thilagar, S. 2023. A review: The antiviral activity of cyclic peptides. International Journal of Peptide Research and Therapeutics. 29(1): 7.

Chukwudozie, O. S., Duru, V. C., Ndiribe, C. C., Aborode, A. T., Oyebanji, V. O., *and* Emikpe, B. O. 2021. The relevance of bioinformatics applications in the discovery of vaccine candidates and potential drugs for COVID-19 treatment. *Bioinformatics and Biology Insights, 15*(1), 1-8.

Cox, R. J., *and* Brokstad, K. A. 2020. Not just antibodies: B cells and T cells mediate immunity to COVID-19. *Nature Reviews Immunology, 20*(1), 581-582.

Dimitrov, I., Flower, D. R., *and* Doytchinova, I. 2013. AllerTOP - a server for *in silico* prediction of allergens. *BMC Bioinformatics, 14*(6).

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

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics, 5*(113).

Gaurav, A., Agrawal, N., Al-Nema, M., Gautam, V. 2022. Computational approaches in the discovery and development of therapeutic and prophylactic agents for viral diseases. Current Topics in Medicinal Chemistry. 22(26): 2190-2206.

Ghosh, P., Bhakta, S., Bhattacharya, M., Sharma, A. R., Sharma, G., Lee, S. S., *and* Chakraborty, C. 2021. A novel multi-epitopic peptide vaccine candidate against *Helicobacter pylori: In-Silico* identification, design, cloning and validation through molecular dynamics. *International Journal of Peptide Research and Therapeutics, 20*(1), 1-18.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., *and* Raghava, G. P. S. 2013. *In silico* approach of predicting toxicity of peptides and proteins. *PLoS One, 8*(9), e73957.

Jespersen, M. C., Peters, B., Nielsen, M., *and* Marcatili, P. 2017. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research, 45*(W1), W24-W29.

Kadanali, A., *and* Karagoz, G. 2015. An overview of Ebola virus disease. *Northern Clinics of Istanbul, 2*(1), 81-86.

Kamens, J. 2015. The Addgene repository: An international nonprofit plasmid and data resource. *Nucleic Acids Research, 43*(1), D1152-D115.

Keita, A. K., Koundouno, F. R., Faye, M., Düx, A., Hinzmann, J., Diallo, H., Ayouba, A., Le Marcis, F., Soropogui, B., Ifono, K., Diagne, M. M., Sow, M. S., Bore, J. A., Calvignac-Spencer, S., Vidal, N., Camara, J., Keita, M. B., Renevey, A., Diallo, A., Soumah, A. K., Millimono, S. L., Mari-Saez, A., Diop, M., Doré, A., Soumah, F. Y., Kourouma, K., Vielle, N. J., Loucoubar, C., Camara, I., Kourouma, K., Annibaldis, G., Bah, A., Thielebein, A., Pahlmann, M., Pullan, S. T., Carroll, M. W., Quick, J., Formenty, P., Legand, A., Pietro, K., Wiley, M. R., Tordo, N., Peyrefitte, C., McCrone, J. T., Rambaut, A., Sidibé, Y., Barry, M. D., Kourouma, M., Saouromou, C. D., Condé, M., Baldé, M., Povogui, M., Keita, S., Diakite, M., Bah, M. S., Sidibe, A., Diakite, D., Sako, F. B., Traore, F. A., Ki-Zerbo, G. A., Lemey, P., Günther, S., Kafetzopoulou, L. E., Sall, A. A., Delaporte, E., Duraffour, S., Faye, O., Leendertz, F. H., Peeters, M., Toure, A., *and* Magassouba, N. F. 2021. Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. *Nature, 597*(7877), 539-543.

Kemenesi, G., Kurucz, K., Dallos, B., Zana, B., Földes, F., Boldogh, S., Görföl, T., Carroll, M. W., *and* Jakab, F. 2018. Re-emergence of Lloviu virus in *Miniopterus schreibersii* bats. *Emerging Microbes and Infections, 7*(66).

Khan, F., Niaz, K., *and* Abdollahi, M. 2018. Toxicity of biologically active peptides and future

safety aspects: an update. *Current Drug Discovery Technologies*, *15*(3), 236–242.

Kharisma, V. D., *and* Ansori, A. N. M. 2020. Construction of epitope-based peptide vaccine against SARS-CoV-2: Immunoinformatics study. *Journal of Pure and Applied Microbiology, 4*(1), 999–1005.

Kharisma, V. D., Ansori, A. N. M., Posa, A. N. V., Rizky, W. C., Permana, S., *and* Parikesit, A. A. 2021. Conserved B-cell epitope identification of envelope glycoprotein (GP120) HIV-1 to develop multi-strain vaccine candidate through bioinformatics approach. *Jurnal Teknologi Laboratorium, 10*(1), 6-13.

Khatoon, N., Pandey, R. K., *and* Prajapati, V. K. 2017. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Scientific Reports, 7*(8285).

Kuhl, A., *and* Pohlmann, S. 2012. How Ebola virus counters the interferon system. *Zoonoses and Public Health, 59*(2), 116-131.

Kyrychenko, A. M., Antipov, I. O., *and* Hrynchuk, K. V. 2017. Phylogenetic analysis of Ukranian BYMV isolates from soybeans and beans. *Cytology and Genetics, 51*(3), 173-177.

Le Bert, N., Tan, A. T., Kunasegaran, K., Tham, C. Y. L., Hafezi, M., Chia, A., Chng, M. H. Y., Lin, M., Tan, N., Linster, M., Chia, W. N., Chen, M. I., Wang, L. F., Ooi, E. E., Kalimuddin, S., Tambyah, P. A., Low, J. G., Tan, Y. J., *and* Bertoletti, A. 2020. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature, 584*(7821), 457-462.

Lee, J. E., *and* Saphire, E. O. 2009. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virology, 4*(6), 621-635.

Liu, S. Q., Deng, C. L., Yuan, Z. M., Rayner, S., *and* Zhang, B. 2015. Identifying the molecular pattern of evolution for Zaire ebolavirus in the 2014 outbreak in West Africa. *Infection, Genetics and Evolution, 32*(1), 51-59.

Mark, J. K. K., Lim, C. S. Y., Nordin, F., Tye, G. J. 2022. Expression of mammalian proteins for diagnostics and therapeutics: a review. Molecular Biology Reports. 49(11): 10593-10608.

Maroney, K. J., Pinski, A. N., Marzi, A., *and* Messaoudi, I. 2021. Transcriptional analysis of infection with early or late isolates from the 2013-2018 West Africa Ebola virus epidemic does not suggest attenuated pathogenicity as a result of genetic variation. *Frontiers in Microbiology, 12*(714817).

Maxwell, P. I., *and* Popelier, P. L. A. 2017. Unfavorable regions in the ramachandran plot: Is it really steric hindrance? The interacting quantum atoms perspective. *Journal of Computational Chemistry, 38*(29), 2459-2474.

Mitra, M. 2018. Infectious characteristics of Ebola virus. *International Journal of Virology and AIDS, 5*(44).

Pandey, R. K., Bhatt, T. K., *and* Prajapati, V. K. 2018. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for Malaria by investigating *Anopheles* salivary protein. *Scientific Reports, 8*(1125).

Parvizpour, S., Pouseif, M. M., Razmara, J., Rafi, M. A., *and* Omidi, Y. 2020. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discovery Today, 25*(6), 1034-1042.

Sette, A., *and* Rappuoli, R. 2010. Reverse vaccinology: developing vaccines in the era of genomics. *Immunity, 33*(4), 530-541.

Singh, R. J., Dhama, K., Malik, Y. S., Ramakrishnan, M. A., Karthik, K., Khandia, R., Tiwari, R., Munjal, A., Saminathan, M., Sachan, S., Desingu, P. A., Kattoor, J. J., Iqbal, H. M. N., *and* Joshi, S. K. 2017. Ebola virus – epidemiology, diagnosis, and control: threat to humans, lessons learnt, and preparedness plans – an update on its 40 year's journey. *Veterinary Quarterly, 37*(1), 98-135.

Su, Y., *and* Stahelin, R. V. 2020. The minor protein matrix VP24 from ebola virus lacks direct lipid binding properties. *Viruses, 12*(869), 1-14.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., *and* Schwede, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296–W303.

Zhang, A. P. P., Abelson, D. M., Bornholdt, Z. A., Liu, T., Woods, V. L. Jr., *and* Saphire, E. O. 2012. The ebolavirus VP24 interferon antagonist: know your enemy. *Virulence, 3*(5), 440-445.