KnE Medicine

ICMEDH
The International Conference of Medicine and Health (ICMEDH)
Volume 2022

Knowledge E
enriching | engaging | empowering

Conference Paper

Developing an Epitope-Based Peptide Vaccine for the Hepatitis C Virus Using an in Silico Approach

Firasti AN Sumadi* Ahmad Shobrun Jamil, Nabilatul Auliyana, Melinia Nur H, Riana Nabilah K

Pharmacy Study Program, Faculty of Health Sciences, University of Muhammadiyah Malang, Malang City 65145, Indonesia

ORCID
Firasti AN Sumadi: https://orcid.org/0000-0002-0839-7941

Abstract.
A vaccine has still not been found for the hepatitis C virus (HCV) and the number of global cases of HCV is still high. The development of HCV vaccines is challenging due to the complex genetic diversity of the virus. Epitope-based vaccine design using in silico computational methods is an effective strategy that could lead to the development of vaccines with the ability to induce the required immunogenicity without the emergence of cytokine storms or immune tolerance. This study aimed to find an epitope candidate from the HCV E2 protein which has potential as a peptide vaccine. The research was observational descriptive and was carried out in silico on an HCV vaccine candidate. The software used was MEGA X, IEDB, VaxiJen 2.0, BLASTp. 3 conserved regions were obtained from 10 countries, namely VCGPVYCFTPSPVVVGTTD, CPTDCFRK, and YRLWHYPCT. The sequences between these countries still have phylogenetic relationships, even though they are in different branches, showing the evolution of the HCV subtypes. The VYCFTPSPVVVGTTD epitope became the candidate for the development of a peptide vaccine because of its antigenicity score and its ability to be used for effective epitope-based vaccines. This HCV vaccine candidate epitope does not cause an autoimmune response because it has been confirmed to be using BLASTp with the result that it shares no similarity with human cell surface proteins.

Keywords: vaccines, peptides, epitope, hepatitis C virus, in silico

Corresponding Author: Firasti AN Sumadi; email: firasti@umm.ac.id

Published 15 September 2022

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICMEDH Conference Committee.

1. INTRODUCTION

Hepatitis consists of A, B, C, D and E. Hepatitis A and B have found the vaccine, while for Hepatitis D and E cases are still rare. In this study, the hepatitis C virus was used, for which the vaccine has not yet been found and the number of cases is quite high in the world, which is estimated at 170 million people infected with hepatitis C [1].

HCV infection is one of the leading causes of chronic liver disease worldwide. HCV with chronic infection causes liver cirrhosis and is associated with the development of hepatocellular carcinoma (HCC) [1]. Transmission of HCV usually occurs through direct

interactions with blood exposure through blood transfusions, parenteral administration of health care, or injecting drug use [2].

New direct-acting antiviral (DAA) treatment has changed the treatment of HCV, with cure rates of over 90%. Although Direct Acting Antivirus (DAA) is highly effective, it is impossible to treat HCV alone with medication. In the use of DAA, there are also challenges including the presence of resistant variants, low efficacy in cirrhotic patients, drug-drug interactions [3] and if the Hepatitis C virus develops into cirrhosis and even becomes hepatocellular carcinoma (HCC) in the process. The higher treatment costs and complex requirements for genotyping are significant barriers to improving HCV treatment, so vaccines remain essential to prevent transmission and reinfection in at-risk groups.

However, the inability to enhance the immune response as an effective protector and the high variability of the HCV genotype (7 confirmed genotypes and 67 subtypes) have hindered progress in vaccine development. The development of HCV vaccines is challenging due to the complex genetic diversity of the virus [4].

Epitope-based vaccine design using in silico computational methods is an effective strategy that could lead to the development of vaccines with the ability to induce the required immunogenicity without the emergence of cytokine storms or immune tolerance. This in silico strategy also helps in selecting better molecules before testing conditions in vitro or in vivo. The use of the in silico method in this early stage greatly shortens subsequent experimental work. In addition, the proper use of in silico methods can replace, reduce and improve the often misleading and time-consuming use of animal experiments [5].

2. METHOD

2.1. Hepatitis C Virus Sequence Collection With NCBI

Sequence collected from NCBI website https://www.ncbi.nlm.nih.gov/labs/virus in the search field. Keyword that was used is “Hepatitis C Virus”. From each sequence, HCV E2 protein sequence extracted and downloaded in Fasta file type.

2.2. Sequence Analysis with Mega X App to Search Phylogenetic Trees Find Conserved Regions

In the Test of Phyogeny method, the Boostrap method was chosen, and No. of Boostrap Replication that was chosen is 1000. Meanhile,the phylogenetic tree analysis also conducted using the Mega X application. Sequence alignment was performed with the Align by ClustalW method.

2.3. Find Epitopes from the Conserved Region using IEDB Website

The analysis using the IEDB website aims to find candidate epitopes that will be used as candidate vaccines. There are several methods that can be used so that it can be adjusted to the needs. The steps taken for this analysis include opening the IEDB website by typing http://tools.iedb.org/bcell/ in the search field. After the main display appears, then enter the sequence to be analyzed, then select the method to be used, then submit. Results and graphs will appear immediately.

IEDB Website Analysis conducted with Kolaskar & Tongaonkar Method. A semi-empirical method exploiting the physicochemical properties of amino acid residues and their frequency of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants of proteins. The threshold used in this method is 1.024. This value was chosen because it represents the average antigenicity whose accuracy has been tested by application to a large number of proteins [6] Other method used is The Emini method. This method determine the surface accessibility of a protein. The threshold used is 1,000, the use of a threshold of 1,000 can indicate an increased probability of finding epitopes on the surface (7).

2.4. Antigenicity Analysis Using Vaxijen 2.0 Website

Vaxijen 2.0 analysis aims to validate the results of the epitope sequence found from the IEDB whether it is immunogenic or not [7] the Threshold value used is 0.5. This value is used because it has a better level of accuracy [5].

2.5. Epitope Protein Similarity Analysis with Human Surface Proteins with BLASTp

Analysis using BLASTp aims to determine the similarity with proteins on surface cells to prevent autoimmune. The acceptable value so that the epitope can be used is below

70% [8]. The analysis is done by entering the website address into the search field, then enter (https://blast.ncbi.nlm.nih.gov/blast.cgi?PAGE=Proteins). From the available display, enter the sequence or epitope that has been obtained in the ”Enter Query Sequence” column. Then in the ”Choose Search Set” column in the Database section, select ”Non Redundant Protein Sequence”. Next, the organism is selected Homo sapiens, and in the Algorithm in the Program Selection select blastp (protein blast). After all the required fields are filled in, click BLAST for the data to be processed immediately. When the data has been processed, the results to be obtained consist of 2 types, namely there is no similarity between the candidate epitope and the protein or receptor on the surface.

3. RESULT AND DISCUSSION

Hepatitis C Virus Sequence Collection with NCBI

TABLE 1: Hepatitis C Virus Protein Sequence Data.

No	Genotype	GenBank Accession Number	Original Isolate
1	1a	BBC62244.1	Australia
2	1a	BBC62261.1	Indonesia
3	1a	BBC62274.1	India
4	1b	BBH48832.1	Vietnam
5	1b	BBH48839.1	Kamboja
6	2a	BBC62224.1	Jepang
7	2a	BBC62248.1	Thailand
8	3b	AFR33829.1	Cina
9	3d	KJ470619.1	Nepal
10	4a	BBC62226.1	Jepang
11	4d	AHH29580.1	Cina
12	5a	AHH29581.1	Cina
13	6	BBC62291.1	Myanmar
14	6a	BBH48831.1	Vietnam
15	6e	BBH48833.1	Kamboja

From the results of data collection, a total of 15 sequences of Hepatitis C virus were obtained. For each genotype, 1-5 genomic sequence data were selected as a reference. This is done to avoid difficulties in determining sustainable regions, because one of the challenges faced in making HCV vaccines is the high genetic variation of HCV throughout the world, making it difficult to find a sustainable region from the entire

sequence. The genetic variation between one genotype and another is about 30%. So that the vaccine to be made must be able to reach all existing genotypes [9].

Results Analysis Using the Mega X Application

Results of Sequence Analysis with Mega X Application to Find Phylogenetic

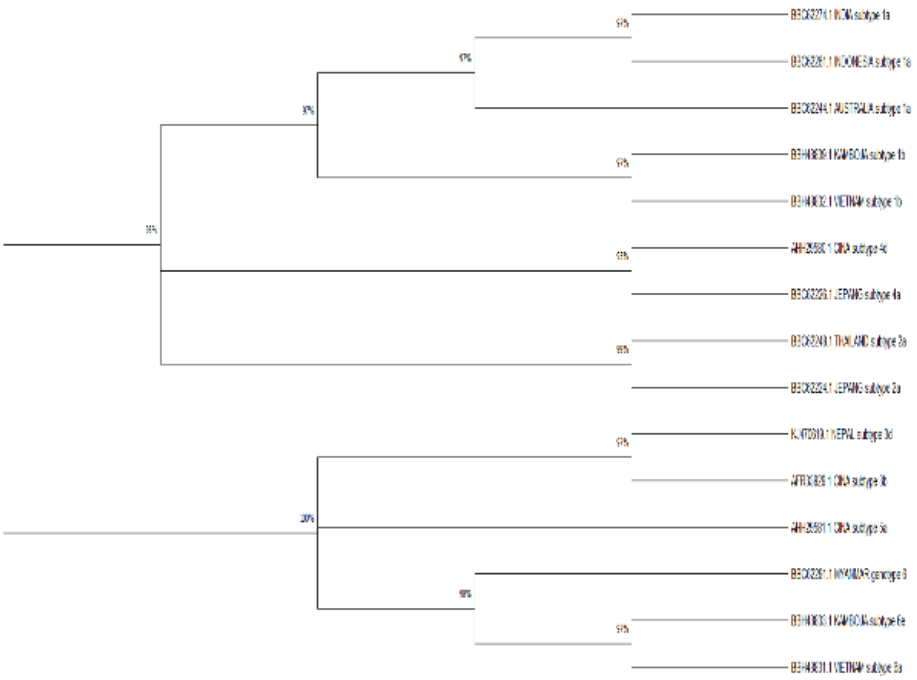

Figure 1: Results of Sequence Analysis with Mega X Application to Find Phylogenetic Trees.

In this study (Figure 1) the resulting bootstrap value was more than 70% which indicated that this phylogenetic analysis was reliable. In this phylogenetic reconstruction, the relationship between species is seen based on the length of branch lines. The different lengths of horizontal lines indicate the level of evolution of each species[10]. From the results of this study, the longer the horizontal line, the longer the line, the longer the evolutionary distance, while the shorter line indicates the close evolution of a species.

3.1. Results of Conserved Regions analysis

In this study, 15 sequences from 6 different genotypes were taken. After multiple sequence alignment, 3 conserved regions of the HCV E2 protein were obtained, namely VCGPVYCFTPSPVVVGTTD, CPTDCFRK and YRLWHYPCT.

The sequence sequence of the conserved Region of the obtained HCV E2 protein tends to be short because in the nucleotide composition of the genome, the six main

genotypes are about 30-35% different from each other, while about 20-25% are different from each other [9].

3.2. Results of Epitope Mapping using IEDB (Emini Method)

Table 2: IEDB Website Analysis Results with the Emini Method.

Emini Method (*Threshold* = 1)				
Sequence	**Start**	**End**	**Peptide**	**Score**
Sequence 1	8	13	FTPSPV	2.071
	9	14	TPSPVV	1.776
	6	11	YCFTPS	1.516
	7	12	CFTPSP	1.496
	14	19	VVGTTD	1.322
Sequence 2	3	8	TDCFRK	1,511
	2	7	PTDCFR	1,169
Sequence 3	1	6	YRLWHY	1.465
	2	7	RLWHYP	1.446

In the Emini method, 9 epitope B cells were found that had scores above the threshold with a maximum score of 2,071 and a minimum of 1.169. All values equal to or greater than the Threshold indicated that the epitope was on the surface. In the Emini graph the yellow area above the threshold (red line) is proposed to be part of the B cell epitope, while the green area is not. The entire sequence analyzed also shows that it can be used as a candidate epitope, which is shown in the yellow graph.

3.3. Results of Epitope Mapping using IEDB (Kolaskar & Tongaonkar Method)

In the Kolaskar and Tongaonkar methods, 17 epitopes of B cells were found that had scores above the threshold. In this study, the maximum score obtained was 1.241 and the minimum was 1.072. Values equal to or above the threshold have the potential to be antigenic determinants. In the Kolaskar and Tongaonkar graphs the yellow area above the threshold (red line) is proposed to be part of the B cell epitope, while the green area is not.

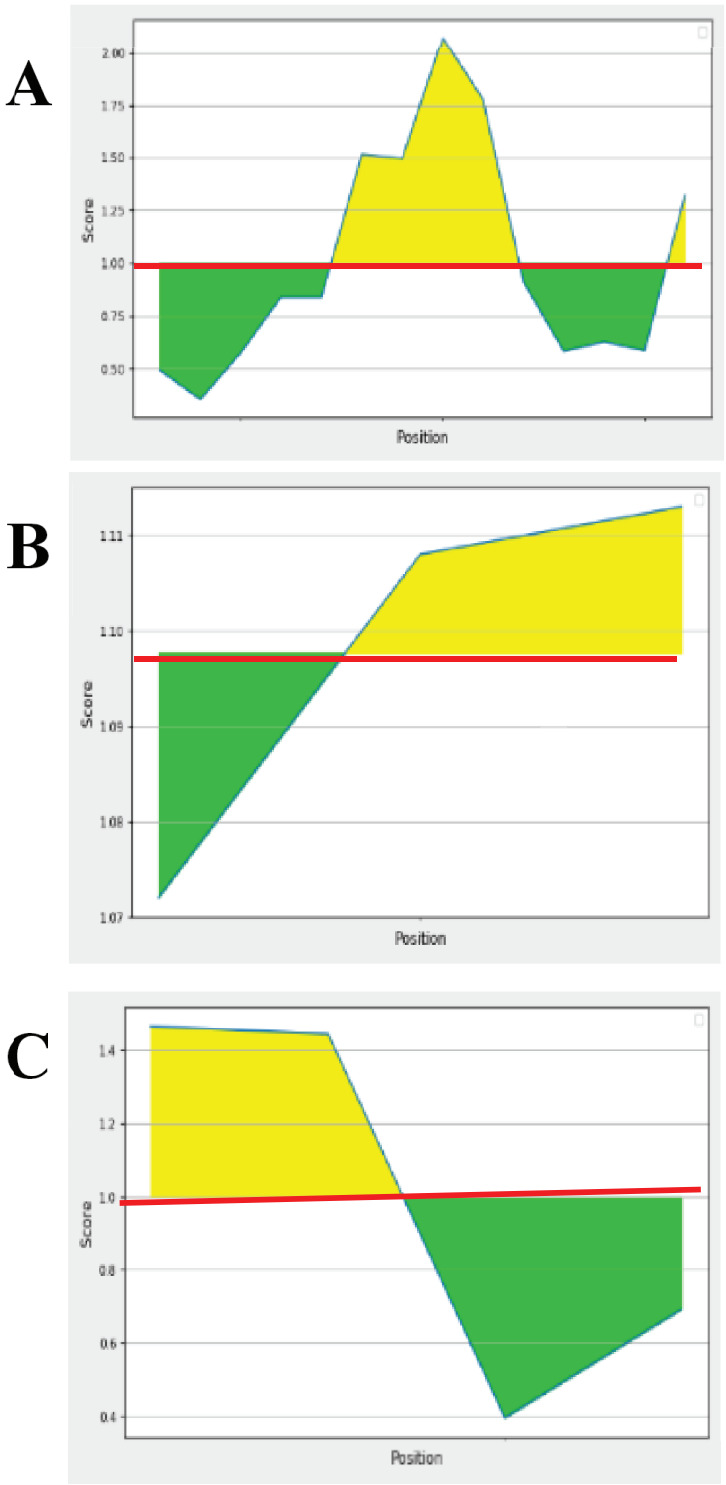

Figure 2: Graph of B cell epitope mapping using the Emini method. (A) Sequence 1, (B) Sequence 2, (C) Sequence 3.

3.4. Results of Antigenicity Analysis Using Vaxijen 2.0

In this study, 5 epitopes were produced which were antigenic from the Emini method and 10 epitopes which were antigenic from the Kolaskar & Toangaonkar method. Epitope

Table 3: Analysis of the IEDB Website with the Kolaskar & Tongaonkar Method.

Kolaskar dan Tongaonkar Method (Threshold = 1.024)				
Sequence	**Start**	**End**	**Peptide**	**Score**
Sequence 1	1	7	VCGPVYC	1.241
	2	8	CGPVYCF	1.2
	9	15	TPSPVVV	1.171
	10	16	PSPVVVG	1.166
	4	10	PVYCFTP	1.155
	5	11	VYCFTPS	1.147
	11	17	SPVVVGT	1.144
	7	13	CFTPSPV	1.134
	8	14	FTPSPVV	1.129
	12	18	PVVVGTT	1.129
	3	9	GPVYCFT	1.128
	6	12	YCFTPSP	1.102
	13	19	VVVGTTD	1.101
Sequence 2	1	7	CPTDFR	1.09
Sequence 3	3	9	LWHYPCT	1.113
	2	8	RLWHYPC	1.108
	1	7	YRLWHYP	1.072

above the threshold (0.5) is an antigenic epitope and below the threshold value is a non-antigen epitope. VaxiJen uses Wold's z-scale to describe the main physicochemical properties of the amino acids that make up the tested protein, then converts the derived strings into uniform vectors by auto cross covariance (ACC), selects relevant variables by genetic algorithm or stepwise regression and finally, classifies proteins as protective antigens or non-antigens by least squares-based discriminant analysis [5]. According to Kolaskar & Tongaonkar, the greater the score of an epitope, the greater its antigenicity.

3.5. Epitope Merger From Vaxijen 2.0 Website Analysis

Table 4: Epitope Merger Results From Website Analysis Vaxijen 2.0.

Method	Sequence	Start	End	Peptide	Result
Emini and Kolaskar & Tongaonkar	**1**	**5**	**19**	**VYCFTPSPVVVGTTD**	ANTIGENIC
	3	**2**	**8**	**RLWHYPC**	ANTIGENIC

Epitopes that show antigen results will be combined into one, the Emini method sequence will be combined with the Kolaskar and Tongaonkar method sequences, resulting in 2 epitopes, namely VYCFTPSPVVVGTTD and RLWHYPC (Table 4.). The

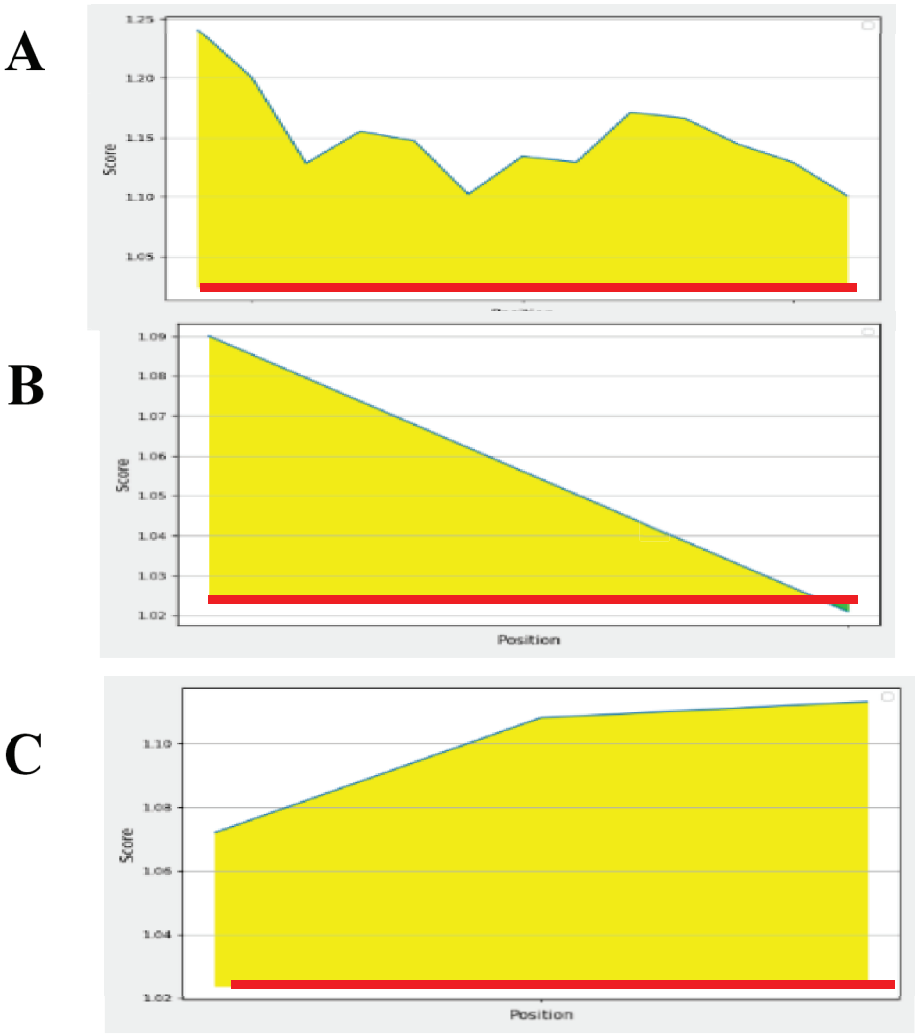

Figure 3: Graph of B cell epitope mapping using the Kolaskar & Tongaonkar method. (A) Sequence 1, (B) Sequence 2, (C) Sequence 3.

combination of epitope aims to validate an epitope so that it can become an effective vaccine.

3.6. Epitope Protein Similarity Analysis with Human Cell Surface Proteins with BLASTp

TABLE 5: Results of Protein Similarity Analysis Using BLASTp.

Peptide	Results BLASTP
VYCFTPSPVVVGTTD	Has no similarity to cell surface proteins
RLWHYPC	Similar to surface cell proteins

Of the 2 epitopes of the HCV vaccine candidate, there is 1 epitope that has similarities to surface proteins, namely the RLWHYPC epitope sequence, the sequence has similarities to receptors in the human body, namely, anaplastic lymphoma receptor tyrosine kinase and Feline leukemia virus subgroup C receptor-related protein 2 with Per(centage) of identity (a value that shows the percentage of similarity between the sequences owned and the target sequence) of 71.43% and 100%, respectively. Meanwhile, the epitope sequence VYCFTPSPVVVGTTD has been confirmed to have no similarity to cell surface proteins in humans.

4. CONCLUSION

In this study, it can be concluded that there are three conserved regions of HCV E2 protein were found, namely VCGPVYCFTPSPVVVGTTD, CPTDCFRK and YRLWHYPCT where the sequences between these countries still have kinship relationships, even though they are in different branches, it shows the evolution of the hepatitis C virus subtypes. Epitope mapping with IEDB obtained 9 B cell epitopes from the Emini method, and 17 B cell epitopes from the Kolaskar method. Meanwhile, the VaxiJen analysis obtained 15 epitope antigens from both Emini and Kolaskar & Tongaonkar methods. The VYCFTPSPVVGTTD epitope is the basis for the development of a peptide vaccine for the Hepatitis C virus because it has a good antigenicity score and can be used for effective epitope-based vaccines. This hepatitis C virus vaccine candidate epitope does not cause an autoimmune response because this epitope has been confirmed using BLASTp with the result that it has no similarity with human cell surface proteins.

References

[1] Duncan JD, Urbanowicz RA, Tarr AW, Ball JK. Hepatitis C virus vaccine: Challenges and prospects. Vaccines (Basel). 2020 Feb;8(1):90.

[2] Spearman CW, Dusheiko GM, Hellard M, Sonderup M. Hepatitis C. Lancet. 2019 Oct;394(10207):1451–66.

[3] Geddawy A, Ibrahim YF, Elbahie NM, Ibrahim MA. Direct acting anti-hepatitis C virus drugs: Clinical pharmacology and future direction. J Transl Int Med. 2017 Mar;5(1):8–17.

[4] Lanini S, Easterbrook PJ, Zumla A, Ippolito G. Hepatitis C: Global epidemiology and strategies for control. Clin Microbiol Infect. 2016 Oct;22(10):833–8.

[5] [5] Flower DR, Doytchinova I, Zaharieva N, Dimitrov I. Immunogenicity prediction by vaxiJen: A ten year overview. Journal of Proteomics & Bioinformatics. 2017;10(11). https://doi.org/10.4172/jpb.1000454.

[6] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990 Dec;276(1-2):172–4.

[7] Rachmawati H, Nugraheni RW, Sumadi FA. In-Silico Approach in the Development of Salmonella Epitope Vaccine. IntechOpen; https://doi.org/10.5772/intechopen.96313.

[8] A. OR. Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. J Appl Pharm Sci. 2019;9(1):98–103.

[9] Nouroz F, Shaheen S, Mujtaba G, Noreen S. An overview on hepatitis C virus genotypes and its control. Egypt J Med Hum Genet. 2015;16(4):291–8.

[10] Anafarida O, Badruzsaufari B. ANALISIS FILOGENETIK MANGGA (MANGIFERA SPP.) BERDASARKAN GEN 5,8S RRNA. ZIRAA'AH MAJALAH ILMIAH PERTANIAN. 2020;45(2):120–6.