Indonesian Journal of Pharmacy

VOL 32 (3) 2021: 328–337 | RESEARCH ARTICLE

Potential Vaccine Targets for COVID-19 and Phylogenetic Analysis Based on the Nucleocapsid Phosphoprotein of Indonesian SARS-CoV-2 Isolates

Muhammad Aldino Hafidzhah[1], Renadya Maulani Wijaya[1], Rasyadan Taufiq Probojati[2], Viol Dhea Kharisma[2], Arif Nur Muhammad Ansori[3], Arli Aditya Parikesit[1]*

1. Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, East Jakarta, Indonesia.
2. Computational Virology and Complexity Sciences Research Unit, Division of Molecular Biology and Genetics, Generasi Biologi Indonesia Foundation, Gresik, Indonesia
3. Doctoral Program in Veterinary Science, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia

Info Article

Submitted: 05-05-2021
Revised: 06-07-2021
Accepted: 14-09-2021

*Corresponding author
Arli Aditya Parikesit

Email:
arli.parikesit@i3L.ac.id

ABSTRACT

Recently, the world is facing outbreaks of severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 and the number of infected patients is increasing every day. Researchers are doing their best to find the most effective treatment to tackle this deathly virus. Several approaches had been proposed to be tested in the lab for efficacy but none of them are qualified to be used as the treatment of the COVID-19. Therefore, this study aimed to design a vaccine based on epitope, which was obtained from the nucleocapsid phosphoprotein (N protein). 38 samples of SARS-CoV-2 isolates were retrieved from the GISAID Database and NCBI GenBank. These samples were used to check the evolutionary relationship of the SARS-CoV-2 and determine whether these nucleocapsid proteins are well-conserved with less or even no mutations occur at all, and whether there was any evolutionary relationship between the recent coronavirus with the previous coronavirus by conducting the phylogenetic analysis. Then, it is desirable to see the molecular interaction between the human BCR/FAB receptor with the predicted peptides through the molecular docking process. All of the peptides were generated by the IEDB analysis tools and have already been tested for antigenicity, so the one that was being docked is the peptide that has antigen properties. Based on the analysis that had been done, the PEP1 was recommended as an epitope-based peptide vaccine candidate to deal with the SARS-CoV-2 outbreaks.

Keywords: Bioinformatics, COVID-19, nucleocapsid protein, SARS-CoV-2, Vaccine Design

INTRODUCTION

COVID-19 was known as the disease that was caused by the newly emerging pandemic of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which originally came from the *Coronaviridae* family (WHO, 2020). The number of patients infected with COVID-19 continues to increase significantly every day since the first outbreak in December 2019 in Wuhan, China. According to the data from World Health Organization (WHO), since the first outbreak in every country, the number of confirmed cases had reached 226,000,000 cases worldwide per September 2021. Meanwhile, in Indonesia, the number of confirmed cases has reached 4.190.763 cases and 140,468 deaths, also per September 2021 (WHO, 2021; Dong *et al*., 2020). This condition is quite threatening for all the elements of society. Researchers around the world are racking up their brains to develop a way to solve this pandemic and there were tons of approaches had been proposed. One of the methods that had been proposed and might show a promising result is by utilizing the immunoinformatic approaches. Immunoinformatic could be very effective to tackle this problem because these methods utilized the study of immunology, which is about the immune system, and bioinformatics as the methods to determine

potential candidate compounds for COVID-19 vaccines.

Before going further deep into the vaccine study, basic understandings of bioinformatics, structural bioinformatics, immunology, and immunoinformatic are needed. Bioinformatics is defined as an interdisciplinary study that combines biology, computer science, and statistics to develops methodology and software tools to understand complex biological data available in this world (Lesk, 2019). This study is focused on using some of the bioinformatics "sub-components" which are structural bioinformatics and Immuno-Informatics. In the field of bioinformatics, there is an area of study, which focuses on the structure, and molecular interactions of a certain biological molecule. Since their main focus is the structure and molecular interactions, this area of study is commonly being used in drug discovery and drug designing. This field of study is known as structural bioinformatics (Brown and Tastan, 2017), and this area of study will leverage the most common tools/technique used for result validation, which is the molecular docking and dynamics.

The structural bioinformatics will be accompanied by another type of study, which combines the study of immunology and the study of informatics/bioinformatics, which are known as immunoinformatic. In scientific definition, immunoinformatic was defined as the interdisciplinary study between experimental immunology and computational approaches (bioinformatics) to help in defining new theories related to immune responses (Tomar and De, 2014). Immunology is a wide field of study, since our study is about designing an epitope-based vaccine, further understanding about epitopes and epitope-based peptide vaccines needs to be obtained. Epitopes were defined as the short and specific amino acid sequences in an antigen that are well recognized by the immune response (Kao and Hodges, 2009). Epitopes are further differentiated into three types, which are the B-cell epitopes, T-helper epitopes, and CTL epitopes. T-helper epitopes are defined as the epitopes that are available on the surface of antigen-presenting cells, which later on will bind with the MHC molecules (Steers *et al.*, 2014). Next is one of the antigen parts, which was shown to be binding with the antibodies, which are called the B-cell epitope (Sanchez-Trincado *et al*., 2017).

Those three epitopes are related to each other especially the T-helper and B-cell because if T-helper cells do not start or trigger any immune mechanism then B-Cell would not be produced and there will be no immune response inside our body, which makes our body prone to be attacked by diseases. For this study our focus is towards the B-cell epitope since this epitope produces an immune response inside our body, and if the virus can be "introduced" with the immune system first by developing a vaccine that utilizes an epitope-based component. Hopefully, when the COVID-19 tries to attack our immune system, our immune system has "prepared" itself. Now, let us know more about the epitope-based peptide vaccine.

Epitope-based peptide vaccine was defined as a type of vaccine that utilizes the chunks of sequences that were taken from antigenic proteins of targeted pathogens which has high immunogenicity (TopuzoĞullari *et al*., 2020). Epitope-based peptide vaccines are used to overcome the problems that commonly appear on the other type of vaccines. Epitope-based peptide vaccine can overcome safety concerns. The vaccine provides us with a maximal therapeutic efficacy, it is cost and time-effective, it limits the allergenic or reactogenic complications, and also it can be further modified to obtain multi-epitope or conjugated structures (TopuzoĞullari *et al*., 2020).

That is why in this study, a new methodology was developed in designing a vaccine for viral diseases especially the COVID-19. The methodology that was used in this study was adapted from a methodology that has been developed previously by Kharisma *et al*., Chen *et al*., and Bhattacharya *et al*. in their study about designing an epitope-based peptide vaccine against SARS-CoV-2 with some changes in the parameter and samples that were being used (Muttaqin and Ansori, 2020; Bhattacharya *et al*., 2020; Chen *et al*., 2020; Kharisma and Ansori, 2020). This methodology mainly utilized the Immuno-informatic approaches along with bioinformatics software to help in designing and predicting specific epitope-based peptide vaccine, which targets the SARS-CoV-2.

MATERIAL AND METHODS

The main tools that were used in this research are software and databases that are freely available and accessible online. In this study, the GISAID EpiCoV database and NCBI GenBank were used to retrieved samples or isolates of the SARS-CoV-2 and previous coronavirus outbreaks. GISAID and NCBI GenBank are freely accessible online databases that store sequences related to living

organisms. In terms of the analysis, prediction, validation, and visualization, online analysis and prediction tools that were utilized in this study are the IEDB Analysis Tools and VaxiJen v2.0 which was used to help in conducting the B-Cell Epitope predictions (IEDB Analysis Tools) and by utilizing the results from the Epitope Prediction to conduct the Antigenicity test with the help of VaxiJen v2.0. These predictions need to be tested for their validity and to do those, molecular docking needs to be done. To support this molecular docking process, CLUSPRO 2.0 Web-based Docking tools were used. To help in visualization, PyMOL 2.4.0 Visualization software was used in this study. The whole process in this study had been summarized in form of a flowchart (Figure 1).

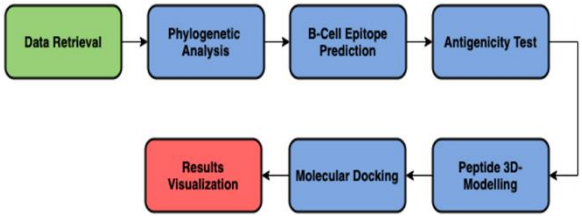

Figure 1. The brief summary of the analysis that had been conducted starting from the sample retrieval to the visualization of the molecular docking results.

Data Retrieval

The process starts by retrieving samples or isolates from the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database and also from NCBI GenBank. In total, 38 Nucleocapsid proteins of the SARS-CoV-2 isolates were successfully being retrieved from both the GISAID EpiCoV database and NCBI GenBank. Nucleocapsid protein was chosen because of its sequence quality. It was shown that nucleocapsid protein has a very well conserved sequence compared to the other proteins available inside the coronavirus structures (Thomas and Gorelick, 2008; Mcbride *et al*., 2014; Samrat *et al*., 2020). Global Initiative on Sharing Avian Influenza Data (GISAID) is a global initiative and resource for genomic data of influenza-related viruses (Shu and McCauley, 2017). To support GISAID in terms of the availability of sequences that came from the previous outbreak of coronavirus, NCBI GenBank was used. National Centre of Biotechnology GenBank (NCBI GenBank) is a comprehensive genetic sequence database that provides various types of genetic sequences that came from various types of organisms, including viruses (Benson *et al*., 2010). In this process, most of the hCoV-19 samples were obtained from the GISAID EpiCoV database and as stated before, the sample from the previous coronavirus outbreak was taken from the NCBI GenBank since most of the samples from the previous coronavirus outbreak are not available in the GISAID EpiCoV database.

Phylogenetic analysis

The next process is to conduct a phylogenetic analysis by utilizing the isolates that had been retrieved from the databases (GISAID and NCBI GenBank). The maximum-likelihood tree was generated with the help of MEGA X tools, which is known as the software used for conducting Multiple-Sequence Alignment. The alignment results were then used to generate an accurate phylogenetic tree (Hall, 2013). The purpose of conducting this analysis is to address some issues related to viral research such as the evolutionary relationships and epidemiology of a certain virus (Ansori *et al*., 2020). In this study would like to address the genetic relationship or evolutionary relationships between the coronaviruses with the help of a phylogenetic tree. In terms of the parameter that are being used, the phylogenetic tree was generated by using a maximum-likelihood method, the Neighbor-Joining algorithms, and the Tamura-Nei model. In terms of the bootstrap method, using bootstrap around 1000 iteration iterated the tree.

B-cell epitope predictions

Moving on to the next process, which is the epitope prediction using IEDB Analysis tools. IEDB B-cell epitope prediction (IEDB BepiPred-2.0) was used to help in predicting epitopes that might be available in the samples of SARS-CoV-2 isolates with an accuracy of 75% (Jespersen *et al*., 2017). In this process, a default parameter was used which consist of the default threshold (threshold score = 0.5). From this prediction, eleven predicted peptides were acquired and those peptides were further tested for their antigenicity using the VaxiJen v2.0 tools.

Antigenicity test

After the predicted peptides were obtained, the peptides underwent an antigenicity test using the VaxiJen v2.0 online tools. VaxiJen v2.0 is an online web-based tool, which is used to determine the characteristics of immunogenicity or protective antigens (Doytchinova and Flower, 2007). From this test, only 6 peptides were considered as a protective antigen and qualified to move to the

next step which was the peptide 3D modeling. Meanwhile, the other five peptides were excluded for the next process since they did not have any antigen features, which were needed to develop an epitope-based peptide vaccine. Mostly default parameter was used in this process, only the 'target organism' menu was changed (originally the 'target organism' was in the 'bacteria' option).

Peptide 3D-modelling

In this process, the six peptides were modeled into 3D structures and the reason why those peptides need to model beforehand is that to conduct molecular docking, 3D structures of the peptides were needed. So, from this process, the 3D-Structure of the six qualified peptides was acquired and these 3D-Structures will be used for Peptide-Peptide docking. In this process, there was no specific parameter that needs to be chosen, so the assumption is that a default parameter provided by the tools was being used. In modeling the peptides, the structure prediction tool used was the PEPFOLD 3 which is a *de novo* approach aimed at predicting peptide structures from amino acid sequences (Shen *et al*., 2014).

Molecular docking

Now, to further validate and determine which peptides will be chosen as an epitope-based peptide vaccine candidate to deal with the SARS-CoV-2 outbreaks, molecular docking needs to be done. In this step, the peptides were docked along with the BCR/FAB receptor, and their binding scores were measured. A peptide with the most negative binding score was proposed as the epitope-based peptide vaccine candidate to deal with the SARS-CoV-2 disease. In this molecular docking process, the default parameter was used since there were no specific parameters provided by the docking tools. To conduct this molecular docking process, the CLUSPRO 2.0 docking tool was deployed. CLUSPRO 2.0 is known as a web-based tool for the protein-protein docking process, and it provides the user with a simple user interface for basic use. Also, it only requires two main files in the PDB format (Kozakov *et al*., 2017).

RESULTS AND DISCUSSION

Data and isolate retrieval

In this process, most of the hCoV-19 samples were obtained from the GISAID EpiCoV database and the sample from the previous coronavirus outbreak was taken from the NCBI GenBank since most of the samples were from the previous coronavirus outbreak are not available in the GISAID EpiCoV database. In total, 38 samples were acquired from those two databases (Table I) (appendix).

Phylogenetic analysis

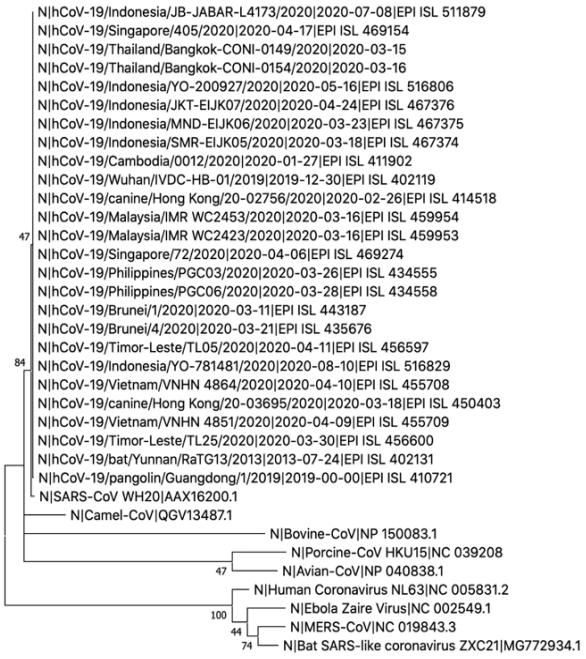

Figure 2. Phylogenetic tree generated to show the evolutionary relationship between the SARS-CoV-2 virus with the other types of coronaviruses that had been emerged in the past decades. the MEGA X software was utilized to generate the maximum-likelihood phylogenetic tree.

The phylogenetic tree (Figure 2) is actual proof that the SARS-CoV-2 still has a very close evolutionary relationship with the previous SARS Coronavirus. It is revealed that there were no significant differences shown in the virus isolates from various countries including those from the Southeast Asia region especially Indonesia. This statement can be proven by seeing the position of the isolates that were in the same clade, which means that they were the same. It does not show very significant differences in terms of their sequences and variants. This phylogenetic tree also shows us that the nucleocapsid protein of the SARS-CoV-2 isolates was very well conserved by seeing the result from the multiple-sequence alignment process right before the phylogenetic tree generation process occurred.

Table I. Samples used for this study. All of the samples were retrieved from GISAID EpiCoV Database and NCBI GenBank. Our study samples were specified to Southeast Asia regions only and some samples from previous coronavirus outbreaks. So, in total there are 38 samples of the coronavirus that will be used in this study.

NO.	ACCESSION ID	SAMPLE NAME	SOURCES / DATABASE
1.	EPI_ISL_435676	hCoV-19/Brunei/2020	GISAID EpiCoV
2.	EPI_ISL_456597	hCoV-19/Timor-Leste/2020	GISAID EpiCoV
3.	EPI_ISL_443187	hCoV-19/Brunei/2020	GISAID EpiCoV
4.	EPI_ISL_434558	hCoV-19/Philippines/2020	GISAID EpiCoV
5.	EPI_ISL_434555	hCoV-19/Philippines/2020	GISAID EpiCoV
6.	EPI_ISL_469274	hCoV-19/Singapore/2020	GISAID EpiCoV
7.	EPI_ISL_459953	hCoV-19/Malaysia/2020	GISAID EpiCoV
8.	EPI_ISL_459954	hCoV-19/Malaysia/2020	GISAID EpiCoV
9.	EPI_ISL_450403	hCoV-19/Hongkong/2020	GISAID EpiCoV
10.	EPI_ISL_455708	hCoV-19/Vietnam/2020	GISAID EpiCoV
11.	EPI_ISL_455709	hCoV-19/Vietnam/2020	GISAID EpiCoV
12.	EPI_ISL_516829	hCoV-19/Indonesia/2020	GISAID EpiCoV
13.	EPI_ISL_414518	hCoV-19/Hongkong/2020	GISAID EpiCoV
14.	EPI_ISL_402119	hCoV-19/Wuhan/2020	GISAID EpiCoV
15.	EPI_ISL_411902	hCoV-19/Cambodia/2020	GISAID EpiCoV
16.	EPI_ISL_467374	hCoV-19/Indonesia/2020	GISAID EpiCoV
17.	EPI_ISL_467375	hCoV-19/Indonesia/2020	GISAID EpiCoV
18.	EPI_ISL_516806	hCoV-19/Indonesia/2020	GISAID EpiCoV
19.	EPI_ISL_511879	hCoV-19/Indonesia/2020	GISAID EpiCoV
20.	EPI_ISL_469154	hCoV-19/Singapore/2020	GISAID EpiCoV
21.	EPI_ISL_467376	hCoV-19/Indonesia/2020	GISAID EpiCoV
22.	EPI_ISL_456600	hCoV-19/Timor-Leste/2020	GISAID EpiCoV
23.	EPI_ISL_402131	hCoV-19/Yunnan/2020	GISAID EpiCoV
24.	EPI_ISL_410721	hCoV-19/Guangdong/2020	GISAID EpiCoV
25.	AAX16200.1	SARS-CoV WH20	NCBI GenBank
26.	NP150083.1	Bovine-CoV	NCBI GenBank
27.	QGV13487.1	Camel-CoV	NCBI GenBank
28.	NC039208	Porcine-CoV	NCBI GenBank
29.	NP040838.1	Avian-CoV	NCBI GenBank
30.	NC043505	Yellow Head Virus	NCBI GenBank
31.	NP579881.1	Human Immunodeficiency Virus Type-1 (HIV-1)	NCBI GenBank
32.	NC016991	White-Eye Coronavirus HKU16	NCBI GenBank
33.	NC019843.3	MERS-CoV	NCBI GenBank
34.	MG772934.1	Bat SARS-like-CoV	NCBI GenBank
35.	NC005831.2	Human-CoV NL63	NCBI GenBank
36.	NC002549	Ebola Zaire Virus	NCBI GenBank
37.	EPI_ISL_455947	hCoV-19/Thailand/2020	GISAID EpiCoV
38	EPI_ISL_455943	hCoV-19/Thailand/2020	GISAID EpiCoV

B-Cell epitope prediction and antigenicity test

The input for this B-cell epitope prediction is an N protein sequence which was retrieved from an Indonesian sample (N|hCoV-19/Indonesia/JKT-EIJK07/2020|EPI_ISL_467376). This sample was chosen based on the result from conducting a multiple-sequence alignment between five Indonesian samples, which consist of samples from West Java (1 sample), DKI Jakarta (1 sample), South-Sulawesi (1 sample), Central Java (2 samples). Based on the result of the alignment, the sequence of those 5 samples does

Table IIa. List of peptides generated from the B-cell epitope prediction using IEDB analysis tools, and their antigenicity prediction results which have been calculated by using the VaxiJen v2.0. Peptides that have antigen scores above the threshold proceeded to the molecular docking process.

Peptide	Antigen (Y/N)	Antigen Score (Threshold = 0.4)
NGPQNQRNAPRI	N	0.1648
FGGPSDSTGSNQNGERSGARSKQRRPQGLPNN	N	0.2916
HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS	Y	0.5773
AGLPYGANK	N	0.2631
GALNTPKDHIGTRNPANNAAIVLQLPQ	N	-0.1089
TTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGD	Y	0.5206
RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA	Y	0.5627
RRGPEQTQGNFGDQELIRQGTDYK	Y	0.6277
DPNFKD	Y	2.878
DAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDD	Y	0.4968

Table IIb. List of peptides that have the antigen properties and qualifies for the molecular docking process with human BCR/FAB Receptor.

Peptide	Antigen Score (Threshold = 0.4)
HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS	0.5773
TTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGD	0.5206
RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA	0.5627
RRGPEQTQGNFGDQELIRQGTDYK	0.6277
DPNFKD	2.878
DAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDD	0.4968

Table III. Binding affinity or binding score that were obtained from each of the molecular docking processes. The binding score had been sorted from the highest score (most negative) to the lowest score (least negative).

NO.	Peptide – BCR/FAB Interaction	Binding Score
1.	PEP1_BCR/FAB	-726.5 kJ
2.	PEP2_BCR/FAB	-701.2 kJ
3.	PEP3_BCR/FAB	-537.5 kJ
4.	PEP4_BCR/FAB	-502.0 kJ
5.	PEP5_BCR/FAB	-391.9 kJ
6.	PEP6_BCR/FAB	-607.0 kJ

not show any significant differences, and based on this result, the sequence that was chosen to be predicted in the B-cell epitope prediction is the sample from DKI Jakarta. From this B-Cell Epitope prediction that was conducted on the IEDB online webserver, a total of eleven predicted peptides were acquired (Figure 3). These peptides were generated into a list of peptides, which was inputted into a table based on their prediction score. To be included in the table, the peptide's prediction score needs to go above 0.5 (threshold).

The results after the peptides had gone through the antigenicity test (Table IIa). To determine whether the peptides have antigen properties or not, the peptide's antigenicity score needs to pass or go above the threshold score (threshold score: 0.4). After going through the antigenicity test, only six peptides that have the antigen properties and qualifies to be docked along with the human BCR/FAB receptor.

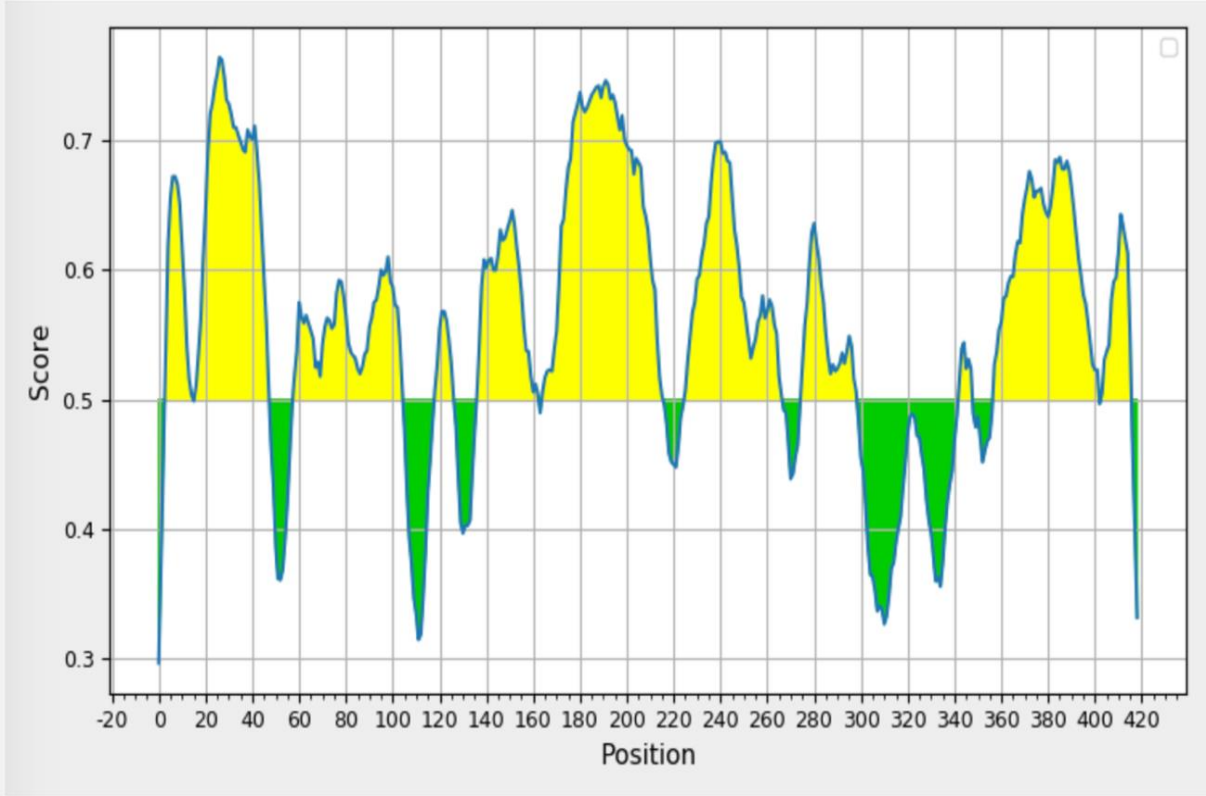

Figure 3. Prediction of B-cell epitopes from the amino acids of N protein of SARS-CoV-2. B-cell epitope prediction was performed using the IEDB online webserver. The region with yellow color indicates a positive prediction, meanwhile, the region with green color indicates a negative prediction of B-cell epitopes.

Meanwhile, the other five peptides will be excluded for the next process since they do not have any antigen features, which were needed to develop an epitope-based peptide vaccine (Table IIb).

But beforehand, those six peptides need to be modeled into 3D-Structures by using the PEP-FOL3 Software. The main reason why it is needed to model the peptides into 3D-Structure is that the molecular docking process can only use 3D-Structure (PDB file) as the input file and the expected result of the molecular docking process is a 3D-Structure of the peptides that bind with the receptor.

Molecular docking and visualizations

The first result shown from the molecular docking process is the visualization of the peptides that bind with the BCR/FAB fragment receptor (Figure 4). From the visualization can be seen that peptides bind perfectly with the BCR/FAB Fragment Receptor, however, this visualization needs to be further validated, and to validate the visualization, a binding affinity score/lowest binding score needs to be checked. To determine which peptides had the best binding affinity score, the most negative score in the results was chosen.

The binding scores of the six peptides after they are docked with the Human BCR/FAB receptors are shown (Table III). The first peptide (PEP1) after the docking process got the highest binding score (most negative) compared to the other six peptides with a score of -726.5. This means that the candidates for the Epitope-Based Vaccine were retrieved and based on the result (Table III) ***PEP1*** **(HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATR RIRGGDGKMKDLS)** was proposes as the epitope-base peptide vaccine candidates to deal with the SARS-CoV-2 virus. BLASTp result with E value cut off of 10^{-3} shows that the PEP1 still has high conservation with nucleoprotein of SARS-CoV-2. The BLASTp result shows that PEP1 still elicits

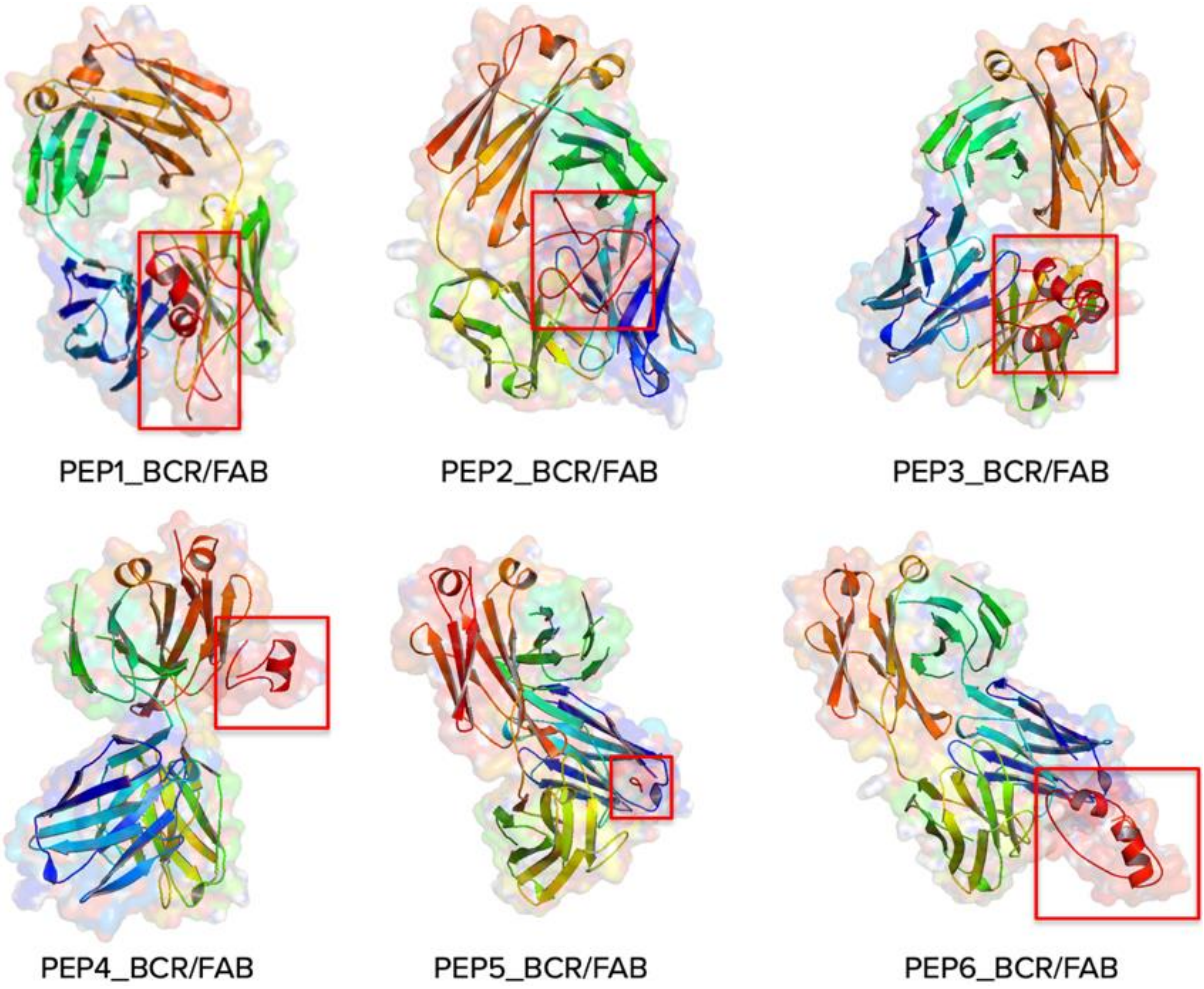

Figure 4. Visualization results from the molecular docking process between each of the peptides with BCR/FAB receptor in 3D structure visualization. It shows that each peptide binds perfectly with the BCR/FAB fragment receptor.

clearly defined properties of the SARS-CoV-2 nucleoprotein.

Then, more than 80% of the annotation SARS-CoV-2 mutations go to the spike protein, while the remaining was left mainly to the nucleoprotein (Troyano-Hernáez *et al,* 2021). Moreover, mutations in the SARS-CoV-2 protein only disrupt the drug-binding cavity, and there was no significant disruption in the epitope's elicitation (Azad, 2021). In this regard, PEP1 still has a higher probability to elicit an immune response as most of the mutation tendency was directed to the spike protein.

CONCLUSION

Based on the analysis that had been conducted and the results that had been retrieved, PEP1 was recommended as an epitope-based peptide vaccine candidate to deal with the SARS-CoV-2 outbreaks. PEP1 was chosen because it has the highest level of immunogenicity, and it can be confirmed that it would not trigger an autoimmune. Also, it was shown that PEP1 is capable of forming BCR molecular complexes with the lowest binding energy for activation of transduction signal in the direct B-Cell immune response. In the future, to further validate the results of this study and to test

the efficacy of this peptide vaccine candidate, molecular dynamics, *in vitro*, and *in vivo* testing could be done. Also, it is still possible to continue the study by developing a multi-epitope-based vaccine design in the future.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Bioinformatics, School of Life Sciences, Center of Student Development Department (CSD), Research and Community Service Department (LPPM) of Indonesia International Institute for Life Sciences (i3L), and Generasi Biologi Indonesia Foundation for supporting this research. The authors would like to thank also the Deputi Bidang Penguatan Riset dan Pengembangan, Kementerian Riset dan Teknologi for providing Hibah Penelitian Dasar 2021 No B/112/E3/RA.00/2021. The authors also gratefully acknowledge the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID

REFERENCES

Ansori, A., Kharisma, V., Antonius, Y., Tacharina, M., and Rantam, F. 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.

Azad, G. K. (2021). The molecular assessment of SARS-CoV-2 Nucleocapsid Phosphoprotein variants among Indian isolates. Heliyon, 7(2), e06167. https://doi.org/10.1016/J.HELIYON.2021.E06167

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., *and* Sayers, E. W. 2010. GenBank. *Nucleic Acids Research*, *38* (Database issue), D46–D51.

Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., Lee, S. S., *and* Chakraborty, C. 2020. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *Journal of Medical Virology*, *92*(6), 618–631.

Brown, D. K., *and* Tastan Bishop, Ö. 2017. Role of structural bioinformatics in drug discovery by computational SNP analysis: Analyzing variation at the protein level. *Global Heart*, *12*(2), 151–161.

Chen, H. Z., Tang, L. L., Yu, X. L., Zhou, J., Chang, Y. F., *and* Wu, X. 2020. Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. *Infectious Diseases of Poverty*, *9*, 88.

Dong, E., Du, H., *and* Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. The Lancet. *Infectious Diseases*, *20*(5), 533–534.

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.

Hall, B. G. 2013. Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229–1235.

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.

Kao, D. J., *and* Hodges, R. S. 2009. Advantages of a synthetic peptide immunogen over a protein immunogen in the development of an anti-pilus vaccine for *Pseudomonas aeruginosa*. *Chemical Biology & Drug Design*, *74*(1), 33–42.

Kharisma, V. D., *and* Ansori, A. N. 2020. Construction of Epitope-Based Peptide Vaccine Against SARS-CoV-2: Immunoinformatics Study. *Journal of Pure and Applied Microbiology, 14*, 999-1005.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., *and* Vajda, S. 2017. The ClusPro web server for protein-protein docking. *Nature Protocols*, *12*(2), 255–278.

Lesk, A. M., 2019. Bioinformatics. Retrieved November 14, 2020. https://www.britannica.com/science/bioinformatics.

Mcbride, R., Zyl, M. V., *and* Fielding, B. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses, 6*(8), 2991-3018.

Muttaqin, S. S., *and* Ansori, A. N. M. Candidate inhibitors of SARS-CoV-2 main protease with 3D structures similar to N3. *Research Journal of Biotechnology*, *15*(11), 55–59.

Samrat, S. K., Tharappel, A. M., Li, Z., *and* Li, H. 2020. Prospect of SARS-CoV-2 spike protein: Potential role in vaccine and therapeutic development. *Virus research*, *288*, 198141.

Sanchez-Trincado, J. L., Gomez-Perosanz, M., *and* Reche, P. A. 2017. Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research*, *2017*, 2680160.

Shen, Y., Maupetit, J., Derreumaux, P., *and* Tufféry, P. 2014. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of Chemical Theory and Computation*, *10*(10), 4745–4758.

Shu, Y., *and* McCauley, J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveillance: Bulletin Europeen sur les maladies transmissibles = European Communicable Disease Bulletin*, *22*(13), 30494.

Steers, N. J., Currier, J. R., Jobe, O., Tovanabutra, S., Ratto-Kim, S., Marovich, M. A., Kim, J. H., Michael, N. L., Alving, C. R., *and* Rao, M. 2014. Designing the epitope flanking regions for optimal generation of CTL epitopes. *Vaccine*, *32*(28), 3509–3516.

Thomas, J. A., *and* Gorelick, R. J. 2008. Nucleocapsid protein function in early infection processes. *Virus Research*, *134*(1-2), 39–63.

Troyano-Hernáez, P., Reinosa, R., & Holguín, Á. (2021). Evolution of sars-cov-2 envelope, membrane, nucleocapsid, and spike structural proteins from the beginning of the pandemic to september 2020: A global and regional approach by epidemiological week. Viruses, 13(2). https://doi.org/10.3390/v13020243

Tomar, N., *and* De, R. K. 2014. Immunoinformatics: A brief review. *Methods in Molecular Biology*, *1184*, 23–55.

TopuzoĞullari, M., Acar, T., Pelİt Arayici, P., UÇar, B., UĞurel, E., Abamor, E. Ş., ArasoĞlu, T., Turgut-Balik, D., *and* Derman, S. 2020. An insight into the epitope-based peptide vaccine design strategy and studies against COVID-19. *Turkish Journal of Biology*, *44*(3), 215–227.

WHO. 2020. Coronavirus. Retrieved September 16, 2021. https://www.who.int/health-topics/coronavirus/coronavirus.