Sys Rev Pharm 2020;11(7):404-413
A multifaceted review journal in the field of pharmacy

Characterization of the Spike Glycoprotein and Construction of an Epitope-Based Vaccine Candidate against Indonesian SARS-CoV-2: *In Silico* Study

Irine Normalina[1,2,3], Setyarina Indrasari[1,2,4,5], Reviany V. Nidom[1,2,4], Muhammad K. J. Kusala[1,5], M. Yusuf Alamudi[1,2], Kuncoro P. Santoso[1,5], Kadek Rachmawati[1,5], Chairul A. Nidom[1,2,4,5]*

[1]Professor Nidom Foundation, Surabaya, Indonesia.
[2]PADIA Lab, Surabaya, Indonesia.
[3]Faculty of Medicine Universitas Airlangga, Surabaya, Indonesia.
[4]PT. Riset AIRC, Surabaya, Indonesia.
[5]Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia.

Corresponding Author: Chairul A. Nidom, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia. Phone: +62 31 5030255. E-mail address: nidomca@fkh.unair.ac.id / nidomca@pnfinstitute.org.

ABSTRACT

Introduction: SARS-CoV-2, a new member of the coronavirus family that originated from Wuhan, China, is the agent of COVID-19 pandemic and has rapidly spread globally.
Objective: We characterized the spike (*S*) glycoprotein gene from the Indonesian SARS-CoV-2 isolates to investigate its genetic composition, predict the B cell epitopes, and construct the molecular phylogenetic among Indonesian SARS-CoV-2 isolates.
Methods: We employed Wuhan-Hu-1 isolate available in GenBank, NCBI and fourteen Indonesian SARS-CoV-2 isolates acquired from the database (GISAID EpiCoV). We performed using the MEGA X for genetic and amino acid mutations and construct molecular phylogenetic tree. We used IEDB web server to predict epitopes, evaluated allergenicity by applying AllerTOP v.2.0 web server, and non-toxic antigens applying the ToxinPred web server.
Results: Interestingly, we discovered that the Indonesian SARS-CoV-2 isolates genetic composition do not have significant changes compared with the reference sequence based on the S glycoprotein gene. In addition, we proposed NSASFSTFKCYGVSPTKLNDLCFTNV as a candidate for a peptide-based vaccine against COVID-19. Furthermore, we also presented the molecular phylogenetic of Indonesian SARS-CoV-2 isolates and other coronaviruses.
Conclusion: In summary, this study supplied data regarding mutations in the S glycoprotein and we proposed a candidate for peptide-based vaccine against COVID-19. However, this research still requires further genetic analysis and we recommend improvement in the molecular epidemiological surveillance on COVID-19 in Indonesia.

Keywords: COVID-19, *in silico*, SARS-CoV-2, vaccine design

Correspondence:
Chairul A. Nidom
Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia. Phone: +62 31 5030255. E-mail address: nidomca@fkh.unair.ac.id / nidomca@pnfinstitute.org.

INTRODUCTION

A virus that was causing a new pneumonia outbreak was reported in China in the late 2019[1]. Subsequently, SARS-CoV-2 was recognized as the agent[2]. Since then, the virus has appeared sporadically all over China and in many other nations worldwide[3]. COVID-19 is the illness lead by the new virus, named by the World Health Organization[4]. In addition, the Johns Hopkins University has created an online website that tracks COVID-19 cases reported in real-time every day from various affected countries[5]. Today, approximately 20 million people globally infected by the new virus, about 125,000 incidents in Indonesia. In the past, several types of coronaviruses have caused diseases in humans[6]. The coronavirus family consists of *Alphacoronavirus*, *Deltacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*[7]. They are a RNA genome of about 32 kb with 65 to 125 nm in diameter. The membrane, envelope, spike, and nucleocapsid are structural proteins encoded by the genome[8]. The S glycoprotein has also recently emerged as a prime prospective antigen in formulation of a SARS-CoV-2 vaccine. The two main reasons for this are that it (1) mediates interactions with host cells by binding to ACE2 receptors and (2) engages in surface exposure and is thus immediately recognized by the host immune system[9].

Molecular epidemiology research is an important tool for monitoring new and emerging viruses. Indonesia is one of the Association of Southeast Asian Nations that have reported the entire series of SARS-CoV-2 genomes in their respective regions, along with Brunei, Myanmar, Vietnam, Singapore, Timor Leste, Malaysia, Thailand, and Cambodia. Currently, research groups are racing to generate vaccines for SARS-CoV-2 worldwide and protein-based vaccines becoming one of the most advanced types of vaccines. The private sector is at the forefront of this work[10]. Simultaneously, other potential therapeutic alternatives against COVID-19 have been described by Al-Tawfiq, remdesivir and chloroquine are potent inhibit SARS-CoV-2[11]. Nonetheless, presently there are no potent medication contrary to COVID-19. Currently, we characterized the S glycoprotein gene from the fourteen Indonesian SARS-CoV-2 isolates to investigate its genetic composition, predict the B-cell epitopes, and construct the molecular phylogenetic.

Materials and Methods

Indonesian SARS-CoV-2 isolates and nucleotide sequence preparation

We obtained the S glycoprotein gene via the GISAID EpiCoV™. We used the reference virus (Wuhan-Hu-1)

available from GenBank, NCBI (Table 1) investigation was completed using the MEGA X.

Table 1. List of Indonesian SARS-CoV-2 isolates.

No	Sample ID	Name	GISAID Clade	Origin
1	NC_045512.2 (Reference)	Wuhan-Hu-1	-	China
2	EPI_ISL_435281	JKT-EIJK0141	L	Indonesia
3	EPI_ISL_435282	JKT-EIJK0317	L	Indonesia
4	EPI_ISL_435283	JKT-EIJK2444	O	Indonesia
5	EPI_ISL_437187	EJ-ITD853Sp	L	Indonesia
6	EPI_ISL_437188	EJ-ITD3590NT	GH	Indonesia
7	EPI_ISL_437189	JKT-EIJK01	L	Indonesia
8	EPI_ISL_437190	JKT-EIJK02	L	Indonesia
9	EPI_ISL_437191	JKT-EIJK03	L	Indonesia
10	EPI_ISL_437192	JKT-EIJK04	L	Indonesia
11	EPI_ISL_458079	EJ-ITD1238Sp	L	Indonesia
12	EPI_ISL_458081	EJ-ITD1273NT	L	Indonesia
13	EPI_ISL_467374	SMR-EIJK05	L	Indonesia
14	EPI_ISL_467375	MND-EIJK06	L	Indonesia
15	EPI_ISL_467376	JKT-EIJK07	L	Indonesia

Mutation and genetic similarity analyses

We analyzed the genetic and amino acid mutations of S glycoprotein genes of SARS-CoV-2 by using MEGA X software[12]. In addition, we investigated Indonesian SARS-CoV-2 isolates similarity by employing the LALIGN using default threshold (10.0)[13].

Epitope prediction analysis

BepiPred-2.0 web server employed to demonstrate epitopes of the S glycoprotein. In addition, we used IEDB to assess Chou and Fasman beta turn (threshold: 1.017), Kolaskar–Tongaonkar antigenicity (threshold: 1.015), Emini surface accessibility (threshold: 1.000), Karplus and Schulz flexibility (threshold: 1.002), and Parker hydrophilicity (threshold: 2.068). Then, we submitted the sequence to the VaxiJen v2.0 web to identify whether the predicted epitopes might be a potential antigen to generate an immune response.

Similarity analysis

We investigated the similarity of the area that was predicted as epitopes to prevent the autoimmune problems. We employed the BLASTp, NCBI and the result must be under 70%.

Allergenicity prediction analysis

We evaluated allergenicity performing AllerTOP v.2.0. We delivered the peptides to the web as described by Abraham Peele *et al*.[14].

Protective nontoxic antigens prediction analysis

We predicted protective nontoxic antigens with ToxinPred. We used all standard threshold described by Gupta *et al*.[15]

Molecular phylogenetic analysis

We rendered molecular phylogenetic modeling and tree visualization by applying MEGA X to the maximum likelihood method using the Indonesian and other coronavirus isolates from throughout the world. The molecular phylogenetic was tested by 1000 bootstrapped input datasets and cross-referencing it with the Tamura-Nei substitution model[12].

Results

We generated a schematic figure of the new virus using BioRender (Figure 1). To prepare a 3D visualization of the S glycoprotein (JKT-EIJK0141), we used SWISS-MODEL and PyMOL v2.4 and edited the figure using BioRender (Figure 2). In addition, we investigated mutation sites of nucleotides, amino acids, and genetic similarity from the Indonesian SARS-CoV-2 S glycoprotein isolates in Table 2, Table 3, and Table 4, respectively. However, the prediction of B-cell epitopes and protective antigens are shown in Table 5. The results of the similarity test, toxin, and allergenicity predictions of Indonesian isolates are shown in Table 6 and Table 7. In Figure 3, the results of the epitope prediction based on different parameters using the IEDB are shown. In Figure 4, we demonstrated the AllerTOP v. 2.0 result. Furthermore, molecular phylogenetic tree of Indonesian isolates is shown in Figure 5.

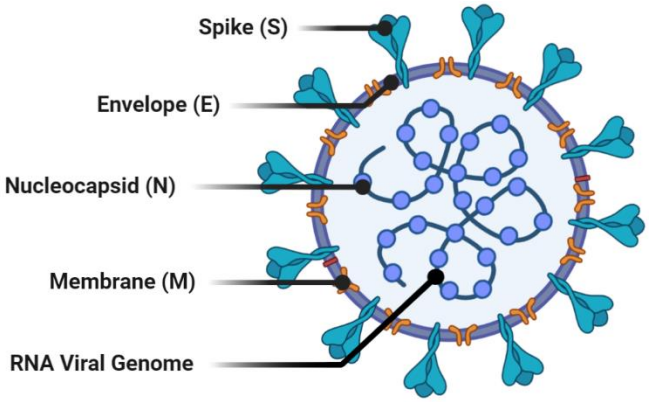

Figure 1. The SARS-CoV-2 structural proteins. N: Nucleocapsid, S: Spike, M: Membrane, and E: Envelope. This figure was prepared using BioRender.

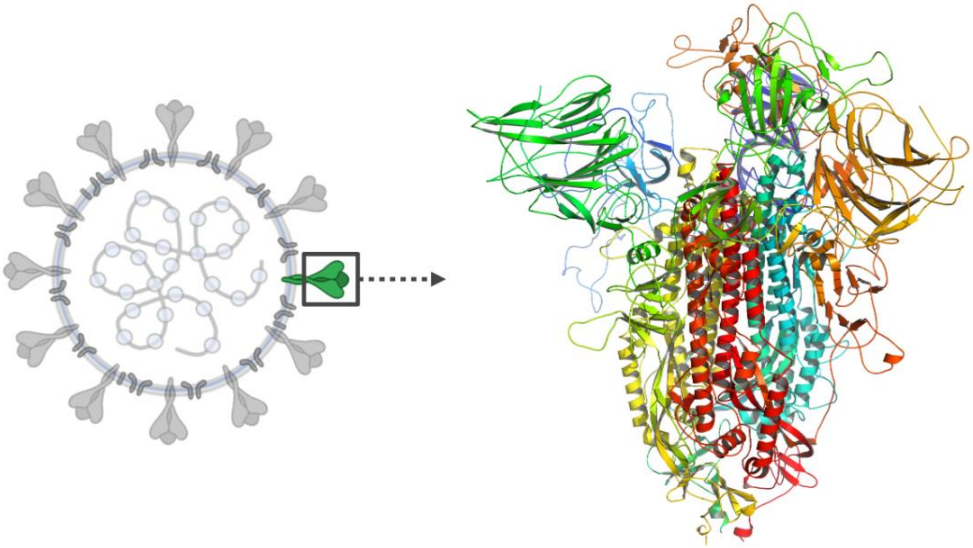

Figure 2. 3D visualization of the S glycoprotein (JKT-EIJK0141). The 3D visualization was conducted using SWISS-MODEL and PyMOL v2.4. This figure was edited using BioRender.

Table 2. Mutation sites of nucleotides in S glycoprotein.

No	Name	Nucleotide Position										
		224	347	414	1054	1715	1841	1864	2015	2031	2464	3761
1	Wuhan-Hu-1 (Reference)	C	C	T	C	C	A	G	C	G	C	G
2	JKT-EIJK0141	C	C	T	C	C	A	G	C	G	C	G
3	JKT-EIJK0317	C	C	T	C	C	A	G	C	G	C	G
4	JKT-EIJK2444	**T**	C	T	C	C	A	G	C	G	C	G
5	EJ-ITD853Sp	C	C	T	C	C	A	G	C	G	C	G
6	EJ-ITD3590NT	C	**G**	T	C	C	**G**	G	C	**T**	C	G
7	JKT-EIJK01	C	C	**C**	C	C	A	**T**	C	G	C	G
8	JKT-EIJK02	C	C	T	C	C	A	G	C	G	C	G
9	JKT-EIJK03	C	C	T	C	C	A	G	C	G	C	G
10	JKT-EIJK04	C	C	T	C	**T**	A	G	C	G	**T**	G
11	EJ-ITD1238Sp	C	C	T	C	C	A	G	C	G	C	G
12	EJ-ITD1273NT	C	C	T	**T**	C	A	G	C	G	C	G
13	SMR-EIJK05	C	C	T	C	C	A	G	**T**	G	C	G
14	MND-EIJK06	C	C	T	C	C	A	G	C	G	C	G
15	JKT-EIJK07	C	C	T	C	C	A	G	C	G	C	**T**

Table 3. Mutation sites of amino acids in S glycoprotein

No	Name	Amino Acid Position									
		76	116	352	572	614	622	672	677	822	1254
1	Wuhan-Hu-1 (Reference)	T	S	A	T	D	V	A	H	L	C
2	JKT-EIJK0141	T	S	A	T	D	V	A	H	L	C
3	JKT-EIJK0317	T	S	A	T	D	V	A	H	L	C
4	JKT-EIJK2444	I	S	A	T	D	V	A	H	L	C
5	EJ-ITD853Sp	T	S	A	T	G	V	A	Q	L	C
6	EJ-ITD3590NT	T	C	A	T	D	F	A	H	L	C
7	JKT-EIJK01	T	S	A	T	D	V	A	H	L	C
8	JKT-EIJK02	T	S	A	T	D	V	A	H	L	C
9	JKT-EIJK03	T	S	A	T	D	V	A	H	L	C
10	JKT-EIJK04	T	S	A	I	D	V	A	H	F	C
11	EJ-ITD1238Sp	T	S	A	T	D	V	A	H	L	C
12	EJ-ITD1273NT	T	S	S	T	D	V	A	H	L	C
13	SMR-EIJK05	T	S	A	T	D	V	V	H	L	C
14	MND-EIJK06	T	S	A	T	D	V	A	H	L	C
15	JKT-EIJK07	T	S	A	T	D	V	A	H	L	F

Table 4. Similarity of Indonesian SARS-CoV-2 S glycoprotein.

No	Similarity (%)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1															
2	100														
3	100	100													
4	100	100	100												
5	100	100	100	100											
6	99.9	99.9	99.9	99.9	99.9										
7	100	99.9	99.9	99.9	99.9	99.9									
8	100	100	100	100	100	99.9	99.9								
9	100	100	100	100	100	99.9	99.9	100							
10	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9						
11	100	100	100	100	100	99.9	99.9	100	100	99.9					
12	100	100	100	99.9	100	99.9	99.9	100	100	99.9	100				
13	100	100	100	99.9	100	99.9	99.9	100	100	99.9	100	99.9			
14	100	100	100	100	100	99.9	99.9	100	100	99.9	100	100	100		
15	100	100	100	99.9	100	99.9	99.9	100	100	99.9	100	99.9	99.9	100	

Table 5. Prediction of B cell epitopes and protective antigens of SARS-CoV-2 isolates from Indonesia.

No	Predicted Peptides	Length	Position	Prediction of Protective Antigens
1	QCVNLTTRTQLPPAYTNSFTRGV	23	14-36	0.7515 (Probable Antigen)
2	FSNVTWFHAIHVSGTNGTKRFDN	23	59-81	0.6767 (Probable Antigen)
3	LGVYYHKNNKSWMESEFRVYSSA	23	141-163	0.4829 (Probable Antigen)
4	DLEGKQGNFKNLRE	14	178-191	0.9256 (Probable Antigen)
5	HTPINLVRDLPQGFSA	16	207-222	0.3936 (Probable Non-Antigen)
6	YLTPGDSSSGWTA	13	248-260	0.6270 (Probable Antigen)
7	FGEVFNATRFASVYAWNRK	19	338-356	0.2386 (Probable Non-Antigen)
8	NSASFSTFKCYGVSPTKLNDLCFTNV	26	370-395	1.3609 (Probable Antigen)
9	GDEVRQIAPGQTGKIADYNYK	21	404-424	1.3212 (Probable Antigen)
10	NNLDSKVGGNYNY	13	439-451	0.9437 (Probable Antigen)
11	LFRKSNLKPFERDISTEIYQAGST	24	455-478	0.1290 (Probable Non-Antigen)
12	ELLHAPATVCGPKKSTNLVK	20	516-535	0.0205 (Probable Non-Antigen)
13	ADQLTPTWRVYSTGSNVFQT	20	626-645	0.3719 (Probable Non-Antigen)
14	SYQTQTNSPRRARSVASQS	19	673-691	0.2316 (Probable Non-Antigen)
15	AYTMSLGAENSVAYSN	16	694-709	0.6003 (Probable Antigen)
16	KQIYKTPPIKDFGGF	15	786-800	-0.3896 (Probable Non-Antigen)
17	LADAGFIKQYGDCLGD	16	828-844	0.0965 (Probable Non-Antigen)
18	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGD ISGI	42	1133-1172	0.1613 (Probable Non-Antigen)
19	SCCKFDEDDSEPVLKGVKL	19	1252-1270	0.6085 (Probable Antigen)

Table 6. Similarity test of SARS-CoV-2 isolates from Indonesia

No	Predicted B Cell Epitope	Target Organism	Similarity Score (BLASTp)
1	QCVNLTTRTQLPPAYTNSFTRGV	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
2	FSNVTWFHAIHVSGTNGTKRFDN	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
3	LGVYYHKNNKSWMESEFRVYSSA	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
4	DLEGKQGNFKNLRE	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
5	YLTPGDSSSGWTA	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
6	NSASFSTFKCYGVSPTKLNDLCFTNV	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
7	GDEVRQIAPGQTGKIADYNYK	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
8	NNLDSKVGGNYNY	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
9	AYTMSLGAENSVAYSN	*Homo sapiens* (taxid:9606)	< 40 (Very Low)
10	SCCKFDEDDSEPVLKGVKL	*Homo sapiens* (taxid:9606)	< 40 (Very Low)

Table 7. Allergenicity and toxin prediction analyses of SARS-CoV-2 isolates from Indonesia.

No	Predicted B Cell Epitope	Toxin Prediction Analysis	Allergenicity Analysis
1	QCVNLTTRTQLPPAYTNSFTRGV	Non-Toxin	Probable Allergen
2	FSNVTWFHAIHVSGTNGTKRFDN	Non-Toxin	Probable Allergen
3	LGVYYHKNNKSWMESEFRVYSSA	Non-Toxin	Probable Non-Allergen
4	DLEGKQGNFKNLRE	Non-Toxin	Probable Non-Allergen
5	YLTPGDSSSGWTA	Non-Toxin	Probable Non-Allergen
6	NSASFSTFKCYGVSPTKLNDLCFTNV	Non-Toxin	Probable Non-Allergen
7	GDEVRQIAPGQTGKIADYNYK	Non-Toxin	Probable Allergen
8	NNLDSKVGGNYNY	Non-Toxin	Probable Allergen
9	AYTMSLGAENSVAYSN	Non-Toxin	Probable Allergen
10	SCCKFDEDDSEPVLKGVKL	Toxin	Probable Allergen

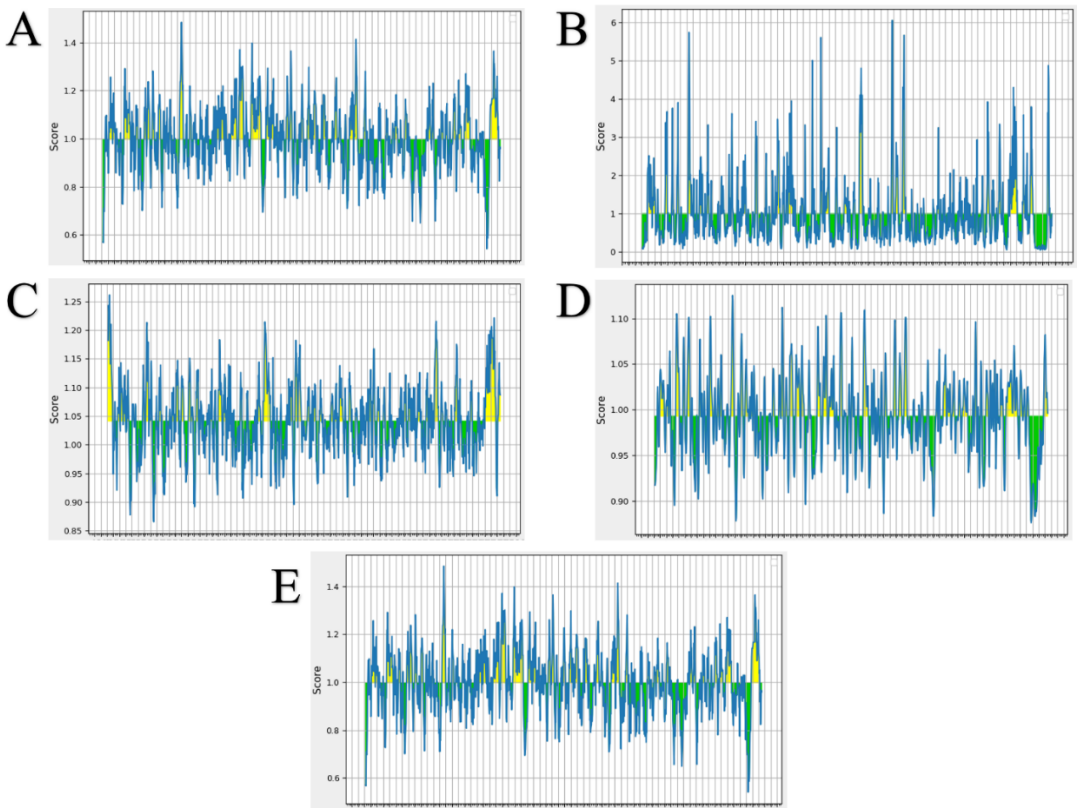

Figure 3. Prediction of epitopes using different parameters. A) Chou and Fasman beta-turn, B) Emini surface accessibility, C) Kolaskar–Tongaonkar antigenicity, D) Karplus and Schulz flexibility, and E) Parker hydrophilicity. The green area was a negative prediction of B cell epitope, whereas the yellow area was a positive.

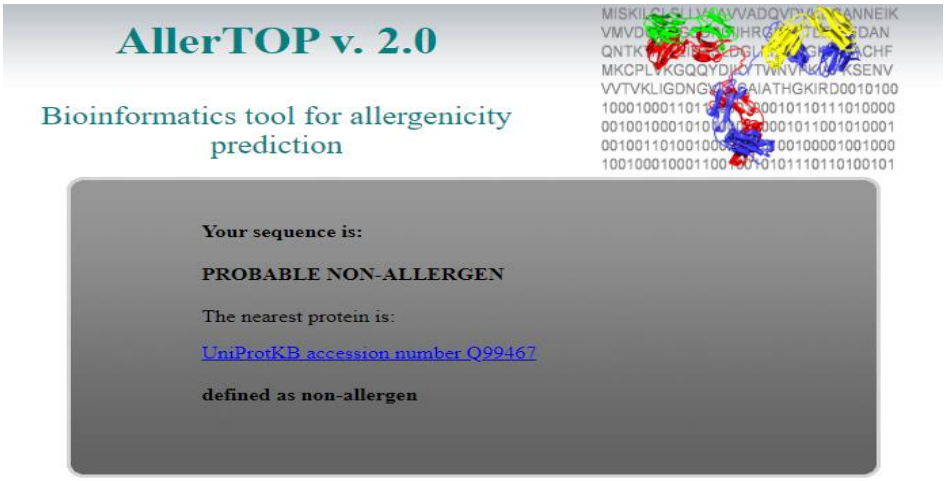

Figure 4. AllerTOP v. 2.0 results defined as non-allergen.

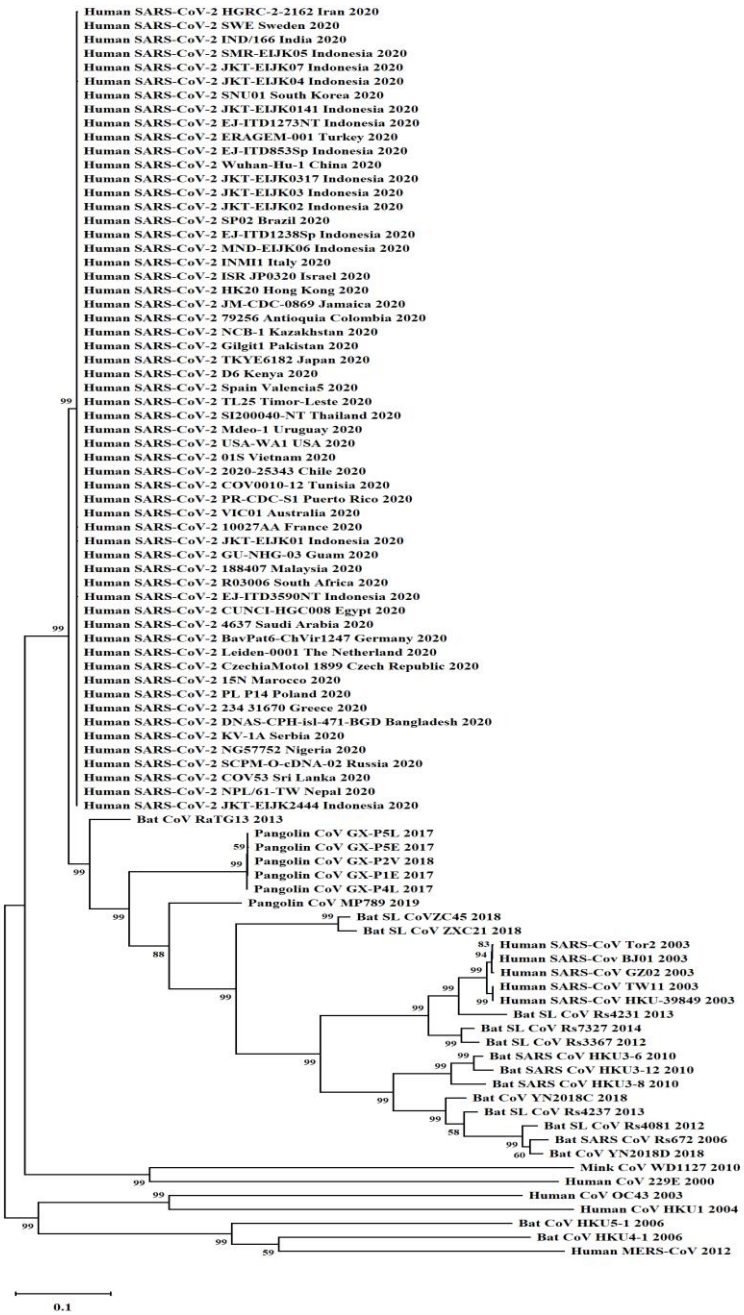

Figure 5. Molecular phylogenetic of Indonesian SARS-CoV-2 isolates.

Discussion

The coronavirus family is arranged of four different genera: *Alpha*-, *Gamma*-, *Beta*-, and *Deltacoronavirus*. Animals and humans might be infected by coronaviruses. *Alphacoronavirus* and *Betacoronavirus* infect animals and humans, whereas *Deltacoronavirus* and *Gammacoronavirus* infect only animals[7]. Until at the end of 2019, there were six coronaviruses known to be causative agents of infection in humans[8]. A seventh, SARS-CoV-2, emerged in China. To date, according to the CSSE at Johns Hopkins University, approximately 12 million people have been infected with the virus globally[5].

The S glycoprotein mediates entry and the new virus membrane fusion and is the main target for many studies of antiviral drugs and vaccines (Figure 1). S1 and S2 are the two domains of the virus S glycoprotein[16]. S1 is conscientious for binding to host cellular receptors (Figure 2). In addition, the efficacy of several therapies, including disrupting protease inhibitors, small RNAs, neutralizing antibodies, fusion blockers, S glycoprotein inhibitors, ACE2 blockers, however, the *in vitro* studies on S glycoprotein inhibitors have been unsatisfactory[17]. Many methods have been employed to produce vaccines using S glycoprotein as an antigen. Thus, it is very important to investigate the S glycoprotein from Indonesian SARS-CoV-2 isolates.

Interestingly, we studied the mutation sites and report no significant changes of the S glycoprotein genes from Indonesian isolates. We identified nucleotide variants of the S glycoprotein genes described in the present research. There were seven isolates with many mutation sites, such as JKT-EIJK01 (414T>C; 1864G>T), EJ-ITD1273NT (1054C>T), SMR-EIJK05 (2015C>T), EJ-ITD3590NT (347C>G; 1841A>G; 2031G>T), JKT-EIJK2444 (224C>T), JKT-EIJK07 (3761G>T), and JKT-EIJK04 (1715C>T; 2464C>T) (Table 2). Furthermore, we investigated amino acid alteration in JKT-EIJK2444 (T76I), EJ-ITD3590NT (S116C; V622F), EJ-ITD1273NT (A352S), JKT-EIJK04 (T572I; L822F), EJ-ITD853Sp (D614G; H677Q), SMR-EIJK05 (A672V), and JKT-EIJK07 (C1254F) (Table 3). Therefore, the genetic similarity score of the S glycoprotein gene from Indonesian isolates was between 99.9% and 100% (Table 4).

Worldwide, scientists have demonstrated mutations in the new virus genome. Sekizuka *et al*. conducted a research and analyze variations of the new virus from Japanese travelers on the Nile River Cruise[18]. Eaaswarkhanth *et al*. speculated that the D614G in the new virus is related with virulence in SARS-CoV-2[19]. Interestingly, we first reported the same mutation pattern in the EJ-ITD853Sp (Table 3). In the lack of wet lab researches, our analysis employing inadequate data regarding the SARS-CoV-2 isolates from Indonesia will potentially benefit ongoing and future studies. Maitra *et al*. showed alteration in the virus isolated from Eastern India and their possible impact on host susceptibility and viral structure[20]. In Chile, a group of scientists demonstrated molecular phylogenetic analysis of the new virus with respect to amino acid and evaluation of nucleotide variants[21]. Khailany *et al*. extracted 94 genomes from GenBank and investigated the molecular variation between them[22].

Currently, the GISAID EpiCoV™ database has established seven subtypes of SARS-CoV-2 namely S, L, V, G, GR, GH, and O. Interestingly, all Indonesian isolates belonged to the L clade except EJ-ITD3590NT (GH) and JKT-EIJK2444 (O). Viral mutation rates are much higher than those of most other microorganisms. RNA viruses have the highest tendency for mutation. A high mutation rate causes increased virulence and a greater capability for adaptive evolution[23]. This ability elevates the possibility of zoonotic viral pathogens to accomplish efficient human-to-human spreading, allowing them to become more virulent. In line with this, SARS-CoV-2 displays the rapid mutation of a virus. This study is vital for successful study into the precaution and medication of COVID-19. Genome cultivation should be valuable in facilitating vaccine design and evaluation of treatment options[24]. The characterization of genotypes linked to infectious clusters and geographic regions prompts that the genome might be useful for observing and tracking the spreading of the virus in Indonesia. Likewise, understanding the virus variants and a molecular epidemiology might enable scientists to establish the origin and observe its transmission. Therefore, the high-speed invention of the changes of the virus is compulsory to an efficient reaction of COVID-19.

In the present research, we constructed the peptide-based vaccines using B cell epitopes prediction (Table 5) following analyses such as the protective antigens prediction (Table 5), similarity test (Table 6), allergenicity prediction (Table 7), and protective nontoxic antigens prediction (Table 7). Previously, B cell epitope prediction researches have been performed for other viruses such as avian influenza virus and dengue virus. We predicted the epitopes of the virus S glycoprotein perfroming BepiPred-2.0 as reported by Jespersen *et al*. (2020) along with Chou and Fasman beta-turn, Kolaskar–Tongaonkar antigenicity, Karplus and Schulz flexibility, Emini surface accessibility, and Parker hydrophilicity on the IEDB (Figure 3) using the default threshold. These techniques were performed to forecast the B cell receptor to bind with the specific areas in proteins. We utilized the AllerTOP v. 2.0 web server to analyze the allergenicity of predicted epitopes (Figure 4) and performed a protective nontoxic antigens prediction by using the ToxinPred web server. In this study, we successfully identified four predicted peptides and proposed that NSASFSTFKCYGVSPTKLNDLCFTNV as a candidate for peptide-based vaccine against COVID-19. Recently, Ansori *et al*. stated the immunoinformatic study of Indonesian SARS-CoV-2 isolates. However, the study has many limitations, such as the number of isolates analyzed and lack of accuracy of the prediction results[26].

Reputable epitope prediction instruments are of major importance in several biotechnological and clinical practices, i.e. therapeutic antibody and vaccine development or concepts of the immune systems[27]. Currently, research groups are racing to generate vaccines contrary to the new coronavirus worldwide and protein-based vaccines is the most developed vaccines. Gupta *et al*. demonstrated the utility of the ToxinPred platform for predicting the toxicity of peptides. Therefore, we used this to generate the safety prediction of peptides in the present study[15]. In addition, AllerTOP v.2.0 is a new powerful allergen prediction method. This web server analyzed peptides established on the physicochemical characteristics of the protein sequences[28]. Furthermore, in this study, the similarity test aims to predict the level of similarity of peptides with protein sequences of somatic cell surface receptors in *Homo sapiens* using the BLASTp server at NCBI. The similarity score of the peptide must be lower in order to prevent an autoimmune reaction if the vaccine is injected into humans. When aligning with the target *Homo sapiens* (taxid: 9606) using the BLASTp server at NCBI, the entire sample has a score < 40 (very low),

therefore, it can be predicted that the peptide from the whole sample of virus isolates cannot trigger autoimmunity because it has a low similarity score when compared with somatic cell surface receptors in *Homo sapiens*.

Molecular phylogenetic analysis is used for fundamental and applied virus research, including taxonomy, origin, and evolution[1]. In this study, we developed a molecular phylogenetic tree and revealed the relationship between Indonesian SARS-CoV-2 isolates, many isolates from various nations around the world, and coronaviruses originating from humans, mink, bats, and pangolins (Figure 5). Here, we report an advanced study to construct a molecular phylogenetic of 14 Indonesian virus isolates from GISAID. Andersen *et al*. demonstrated that the new virus is clearly not a manipulated virus[3]. Previously, an investigation conducted by Ansori *et al*. informed the molecular phylogenetic of Indonesian SARS-CoV-2 isolates. However, that study has many limitations such as the number of isolates, comparison to the circulating the new coronavirus worldwide, and the accuracy of the phylogenetic tree[26].

Lam *et al*. (2020) concluded that coronaviruses are present in various wild mammals in Asia[1]. In addition, Zhou et al. demonstrated the new virus shares 96% of its whole genome with a coronavirus isolated from *Rhinolophus affinis* or the intermediate horseshoe bat, isolated in China. It is crucial to investigate the likelihood of intermediate hosts of the virus to recognize and contain a transmission[29]. Furthermore, pangolin coronavirus which is isolated from Malayan pangolin or *Manis javanica* was 91.02% identical to SARS-CoV-2 at the whole-genome level[1]. The present study advises that promote surveillance researches be implemented on many mammals in their native habitat, including bats and pangolins, especially in East Asia, so that we can be predicted the risk of the forthcoming zoonotic diseases.

Conclusion

In summary, this study supplied data regarding mutations in S glycoprotein of Indonesian SARS-CoV-2 isolates. In addition, we proposed that NSASFSTFKCYGVSPTKLNDLCFTNV might be a candidate for peptide-based vaccine against COVID-19. However, this research still requires further genetic analysis and we recommend improvement in the molecular epidemiological surveillance on COVID-19 pandemic.

Financial Support

This study funded by PADIA Laboratory, Professor Nidom Foundation, Surabaya, Indonesia.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgement

We thank Arif Nur Muhammad Ansori and International Science Editing for the help in editing the manuscript.

REFERENCES

1. Lam TT, Shum MH, Zhu HC, et al. (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature. 583(7815): 282-285.
2. Gorbalenya AE, Baker SC, Baric RS, et al. (2020) The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 5: 536-544.
3. Andersen KG, Rambaut A, Lipkin WI, et al. (2020) The proximal origin of SARS-CoV-2. Nat Med. 26: 450-452.
4. Lai CC, Shih TP, Ko WC, et al. (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents. 55: 105924.
5. Dong E, Du H, Gardner L. (2020) An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. S1473-3099(20): 30120-30121.
6. Walls AC, Park YJ, Tortorici MA, et al. (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 181: 281-292.e6.
7. Ou X, Liu Y, Lei X, et al. (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 11: 1620.
8. Shereen MA, Khan S, Kazmi A, et al. (2020) COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. 24: 91-98.
9. Phan T. (2020) Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 81: 104260.
10. Callaway E. (2020) The race for coronavirus vaccines: A graphical guide. Nature. 580: 576-577.
11. Al-Tawfiq JA, Al-Homoud AH, Memish ZA. (2020) Remdesivir as a possible therapeutic option for the COVID-19. Travel Med Infect Dis. 34: 101615.
12. Kumar S, Stecher G, Li M, et al. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35: 1547-1549.
13. Joob B, Wiwanitkit V. (2020) Genetic variant severe acute respiratory syndrome coronavirus 2 isolates in Thailand. J Pure Appl Microbiol. 14: 6314.
14. Abraham Peele K, Srihansa T, Krupanidhi S, et al. (2020) Design of multi-epitope vaccine candidate against SARS-CoV-2: A in-silico study. J Biomol Struct Dyn. 1: 1-9.
15. Gupta S, Kapoor P, Chaudhary K, et al. (2013) *In silico* approach for predicting toxicity of peptides and proteins. PLoS One. 8(9): e73957.
16. Yin C. (2020) Genotyping coronavirus SARS-CoV-2: Methods and implications. Genomics. S0888-7543(20): 30318-30319.
17. Du L, He Y, Zhou Y, et al. (2009) The spike protein of SARS-CoV–a target for vaccine and therapeutic development. Nat Rev Microbiol. 7: 226e36.
18. Sekizuka T, Itokawa K, Kageyama T, et al. (2020) Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. medRxiv.
19. Eaaswarkhanth M, Al Madhoun A, Al-Mulla F. (2020) Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? Int J Infect Dis. 96: 459-460.
20. Maitra A, Sarkar MC, Raheja H, et al. (2020). Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. J Biosci. 45(1): 76.
21. Castillo AE, Parra B, Tapia P, et al. (2020) Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. J Med Virol.

22. Khailany RA, Safdar M, Ozaslan M. (2020) Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19: 100682.
23. Duffy S. (2018) Why are RNA virus mutation rates so damn high? PLoS Biol. 2018; 16: e3000003.
24. Wang C, Liu Z, Chen Z, et al. (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol. 92(6): 667-674.
25. Jespersen MC, Peters B, Nielsen M, et al. (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 45(W1): W24-W29.
26. Ansori ANM, Kharisma VD, Antonius Y, et al. Immunobioinformatics analysis and phylogenetic tree construction of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Indonesia: spike glycoprotein gene. J Teknol Lab. 9(1): 13-20.
27. Kazi A, Chuah C, Majeed ABA, et al. (2018) Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. Pathog Glob Health. 112(3): 123-131.
28. Dimitrov I, Bangov I, Flower DR, et al. (2014) AllerTOP v.2--a server for *in silico* prediction of allergens. J Mol Model. 20(6): 2278.
29. Zhou P, Yang XL, Wang XG, et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579: 270-273.