Sys Rev Pharm 2020;11(7):422-434
A multifaceted review journal in the field of pharmacy

Immunoinformatic Investigation of Three Structural Protein Genes in Indonesian SARS-CoV-2 Isolates

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

[1]Coronavirus and Vaccine Formulation Research Group, Professor Nidom Foundation, Surabaya, Indonesia.
[2]Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia.
[3]PADIA Lab, Surabaya, Indonesia.
[4]Faculty of Medicine Universitas Airlangga, Surabaya, Indonesia.
[5]PT. Riset AIRC Indonesia, Surabaya, Indonesia.
[#]Both authors contributed equally to this manuscript.

Coresponding Author: Chairul A. Nidom, Professor Nidom Foundation, Surabaya, Indonesia; 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 has crossed the species barrier to infect human. It is a rapidly spreading virus that has poses a significant public threat and is a considerable burden on the global economy and human health. **Objective:** We characterized the nucleocapsid phosphoprotein (N), membrane protein (M), and envelope protein (E) genes of Indonesian isolates to investigate genetic composition, predict B-cell epitopes, and construct a molecular phylogenetic tree. **Methods:** In the present work, we retrieved the sequences of 16 Indonesian isolates from the GISAID EpiCoV and the Wuhan-Hu-1 isolate (reference sequence) from GenBank, NCBI. We used MEGA X to identify mutations in the three structural protein genes and to construct a molecular phylogenetic tree. The IEDB was employed to reveal the linear B-cell epitopes and other parameters. Allergenicity prediction was evaluated using AllerTOP and ToxinPred was performed to analyze non-toxic antigen prediction. **Results:** In this study, we report the genetic composition of three structural protein genes in Indonesian SARS-CoV-2 isolates. Furthermore, we identified the peptide RRGPEQTQGNFGDQELIRQGTDYK from nucleocapsid phosphoprotein to generate a peptide-based vaccine contrary to SARS-CoV-2. **Conclusion:** In summary, we propose a candidate for a peptide-based vaccine contrary to SARS-CoV-2. However, advance trials such as *in vitro* and *in vivo* testing are involved for validation.

Keywords: COVID-19, SARS-CoV-2, structural protein genes, vaccine design

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

INTRODUCTION

SARS-CoV-2 is the seventh coronavirus that has crossed the species barrier to infect human. The virus was first declared in China in 2019 and appeared sporadically all over China and many other nations worldwide[1]. In March 2020, the WHO declared that the infection was a pandemic. The sudden outbreak and quick deployment of COVID-19 have endangered the global health and economy. This crisis has called for extensive scientific mobilization for studies on SARS-CoV-2 concerning its characteristics, mechanism of transmission, and clinical aspects, with the ultimate aim of counteracting the devastating outcomes[2,3]. At present, the virus has infected approximately 20 million people globally with more than 750,000 global deaths. In Indonesia alone, there are more than 120,000 cases and around 6,000 deaths. These data are based on the Johns Hopkins University online website that tracks COVID-19 cases globally in real-time[4].

Animals and humans can be infected by coronaviruses. The coronavirus family is composed of four different genera: *Alpha-*, *Gamma-*, *Beta-*, and *Deltacoronavirus*. *Alphacoronavirus* and *Betacoronavirus* infect animals and humans, whereas *Deltacoronavirus* and *Gammacoronavirus* infect only animals. SARS-CoV-2 belongs to *Coronaviridae* (*Betacoronavirus*), a family that has formerly caused epidemics, SARS-CoV-1 and MERS-CoV[5]. SARS-CoV-2 genome is single-stranded positive-sense RNA roughly 30,000 nucleotides. This genome encodes four structural proteins: spike (S), nucleocapsid (N), membrane (M), and envelope (E)[6]. Zeng *et al.* mentioned that the nucleocapsid phosphoprotein is a very immunogenic and plenteously expressed protein throughout infection[7]. The nucleocapsid phosphoprotein is regularly applied in serological evaluation or vaccine construction[8]. Today, a small number of researches concentrate on the nucleocapsid phosphoprotein, and an updated concept of the Indonesian isolates nucleocapsid phosphoprotein is in urgent need. However, there are no reports about the Indonesian SARS-CoV-2 envelope and membrane proteins.

The molecular epidemiological data of the SARS-CoV-2 in Indonesia remains unclear. Research on molecular epidemiology is a vital instrument for observing new emerging viruses. There is a need to establish further studies in molecular epidemiology to understand the probable impact of the disease[9]. In addition, Indonesia is one of the ASEAN that have informed the entire series of SARS-CoV-2 genomes in their respective areas. Currently, scientists are attempting to generate vaccines against SARS-CoV-2 worldwide, with protein-based vaccines becoming one of the most advanced types and the private sector is at the forefront of this study[10]. Although Jean *et al.* have published the treatment options against COVID-19, presently there are no approved drugs or vaccines against the virus[11]. In the present study, we characterized the nucleocapsid phosphoprotein, envelope protein, and membrane protein genes of Indonesian SARS-CoV-2

isolates to investigate genetic composition, predict the B-cell epitopes, and construct a molecular phylogenetic tree.

Materials and Methods

Indonesian SARS-CoV-2 isolates and preparation

We retrieved Indonesian virus isolates from the database (GISAID EpiCoV) and used the reference virus (Wuhan-Hu-1) extracted from GenBank, NCBI (Table 1). In this study, we characterized genes encoding envelope protein (228 bp), membrane protein (669 bp), and nucleocapsid phosphoprotein (1260 bp). The translation process and multiple sequence alignment were completed by performing MEGA X.

Table 1. Indonesian SARS-CoV-2 isolates extracted from the database.

No	Virus Name	Sample ID	Specimen Source	Country	GISAID Clade
1	Wuhan-Hu-1	NC_045512.2 (NCBI Reference Sequence)	Unknown	Wuhan (China)	-
2	EJ-ITD2766NT	EPI_ISL_458082	Nasopharyngeal and Oro-pharyngeal swab	Surabaya (Indonesia)	GH
3	EJ-ITD3590NT	EPI_ISL_437188	Nasopharyngeal and Oro-pharyngeal swab	Surabaya (Indonesia)	GH
4	EJ-ITD3101NT	EPI_ISL_458083	Nasopharyngeal and Oro-pharyngeal swab	Surabaya (Indonesia)	L
5	JKT-EIJK04	EPI_ISL_437192	Nasopharyngeal and Oro-pharyngeal swab	Jakarta (Indonesia)	L
6	JKT-EIJK0317	EPI_ISL_435282	Nasopharyngeal and Oro-pharyngeal swab	Jakarta (Indonesia)	L
7	JKT-EIJK07	EPI_ISL_467376	Nasopharyngeal and Oro-pharyngeal swab	Jakarta (Indonesia)	L
8	MND-EIJK06	EPI_ISL_467375	Nasopharyngeal and Oro-pharyngeal swab	Manado (Indonesia)	L
9	JKT-EIJK0141	EPI_ISL_435281	Nasopharyngeal and Oro-pharyngeal swab	Jakarta (Indonesia)	L
10	SMR-EIJK05	EPI_ISL_467374	Nasopharyngeal and Oro-pharyngeal swab	Samarinda (Indonesia)	L
11	EJ-ITD1273NT	EPI_ISL_458081	Nasopharyngeal and Oro-pharyngeal swab	Pasuruan (Indonesia)	L
12	EJ-ITD1238Sp	EPI_ISL_458079	Nasopharyngeal and Oro-pharyngeal swab	Surabaya (Indonesia)	L
13	JKT-EIJK03	EPI_ISL_437191	Nasopharyngeal swab	Jakarta (Indonesia)	L
14	EJ-ITD853Sp	EPI_ISL_437187	Sputum	Surabaya (Indonesia)	L
15	JKT-EIJK02	EPI_ISL_437190	Nasopharyngeal swab	Jakarta (Indonesia)	L
16	JKT-EIJK2444	EPI_ISL_435283	Nasopharyngeal swab	Jakarta (Indonesia)	O
17	JKT-EIJK01	EPI_ISL_437189	Nasopharyngeal and Oro-pharyngeal swab	Jakarta (Indonesia)	L

Genetic composition and mutation analysis

In addition, we analyzed the similarity of three structural protein genes (nucleocapsid phosphoprotein, membrane protein, and envelope protein) in the Indonesian isolates using the default threshold in the LALIGN web server[12]. Then, we evaluated the genetic and amino acid mutations of these genes by using MEGA X software.

Epitope prediction analysis

B-cell epitopes prediction of three structural protein genes of the Indonesian isolates was done by applying the IEDB. Then, we performed other parameters such as the Kolaskar and Tongaonkar antigenicity prediction, Karplus and Schulz flexibility prediction, Parker hydrophilicity prediction, Emini surface accessibility prediction, and Chou and Fasman beta-turn prediction using default thresholds[13]. Next, we submitted the predicted peptides to the VaxiJen to determine whether the predicted epitopes could be prospective protective antigens that would establish an immune response.

Allergenicity prediction and protective non-toxic antigen prediction analysis

In this study, an extensive analysis of the allergenicity prediction of the predicted peptides was conducted using AllerTOP with default settings. We submitted the predicted peptides to this web server as demonstrated by Peele *et al*.[14]. Then, we predicted protective non-toxic antigens performing ToxinPred. We used standard thresholds as reported by Gupta *et al*.[15].

Molecular phylogenetic construction

We constructed molecular phylogenetic designing and tree visualization by applying MEGA X on the three structural genes of the Indonesian SARS-CoV-2 isolates with the ML approach. The molecular phylogenetic construction was deduced by using 1000 bootstrapped

input datasets and cross-referenced with the Tamura–Nei substitution model[16].

Results

We generated a schematic diagram of a 3D visualization of three structural proteins in the Indonesian isolate (JKT-EIJK0317). The protein model was rendered using the SWISS-MODEL web server and edited using BioRender (Figure 1). In addition, we evaluated the nucleotide and amino acid mutation sites of the Indonesian isolates in Table 2. B-cell epitopes and protective antigens prediction of the Indonesian SARS-CoV-2 isolates and the allergenicity and toxin prediction analysis are shown in Table 3. In Figures 2–4, we demonstrate the B-cell epitope prediction of the Indonesian SARS-CoV-2 isolates using the IEDB web server. In addition, molecular phylogenetic analysis of the Indonesian SARS-CoV-2 isolates is shown in Figures 5–7.

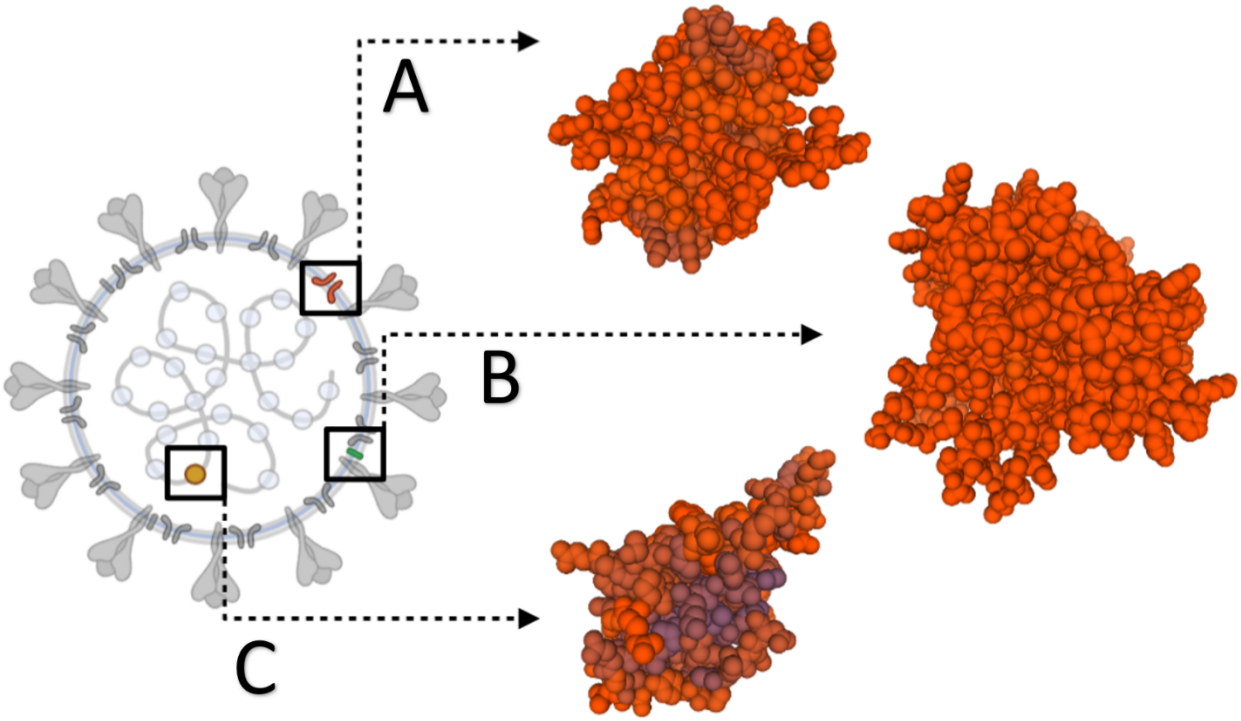

Figure 1. A diagram of the 3D structure visualization of three structural proteins of the Indonesian SARS-CoV-2 isolate (JKT-EIJK0317). A: Membrane protein, B: Envelope protein, and C: Nucleocapsid phosphoprotein. This protein model was rendered using the SWISS-MODEL web server and edited using BioRender.

Table 2. Mutation sites in membrane protein and nucleocapsid phosphoprotein of the Indonesian SARS-CoV-2 isolates.

No	Virus Name	Nucleotide Position			Amino Acid Position
		Membrane Protein Gene	Nucleocapsid Phosphoprotein Gene		Nucleocapsid Phosphoprotein Gene
		213	945	1041	347
1	Wuhan-Hu-1 (Reference)	C	C	A	K
2	EJ-ITD2766NT	**T**	C	A	K
3	EJ-ITD3590NT	**T**	C	A	K
4	EJ-ITD3101NT	C	**T**	A	K
5	JKT-EIJK04	C	C	**T**	**N**
6	JKT-EIJK0317	C	C	A	K
7	JKT-EIJK07	C	C	A	K
8	MND-EIJK06	C	C	A	K
9	JKT-EIJK0141	C	C	A	K
10	SMR-EIJK05	C	C	A	K
11	EJ-ITD1273NT	C	C	A	K
12	EJ-ITD1238Sp	C	C	A	K
13	JKT-EIJK03	C	C	A	K
14	EJ-ITD853Sp	C	C	A	K
15	JKT-EIJK02	C	C	A	K
16	JKT-EIJK2444	C	C	A	K
17	JKT-EIJK01	C	C	A	K

Table 3. B-cell epitopes and other prediction analyses in the Indonesian SARS-CoV-2 isolates.

No	Gene	Peptide	Position	Length	Protective Antigens Prediction	Allergenicity Prediction	Toxigenicity Prediction
1	Envelope Protein	YVYSRVKNLNSSRVP	57-71	15	0.4492 (Probable Antigen)	Probable Non-Allergen	Probable Non-Toxin
2	Membrane Protein	NGTITVEELKKLLEQW	5-20	16	-0.1969 (Probable Non-Antigen)	Probable Allergen	Probable Non-Toxin
3		KLGASQRVAGDS	180-191	12	0.0439 (Probable Non-Antigen)	Probable Non-Allergen	Probable Non-Toxin
4		YRIGNYKLNTDHSSSSDNIA	199-218	20	0.2216 (Probable Non-Antigen)	Probable Allergen	Probable Non-Toxin
5	Nucleocapsid Phosphoprotein	NGPQNQRNAPRI	4-15	12	0.1648 (Probable Non-Antigen)	Probable Allergen	Probable Non-Toxin
6		FGGPSDSTGSNQNGERSGARSKQRRPQGLPNN	17-48	32	0.2916 (Probable Non-Antigen)	Probable Allergen	Probable Non-Toxin
7		HGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS	59-145	47	0.5773 (Probable Antigen)	Probable Non-Allergen	Probable Non-Toxin
8		AGLPYGANK	119-127	9	0.2631 (Probable Non-Antigen)	Probable Non-Allergen	Probable Non-Toxin
9		GALNTPKDHIGTRNPANNAAIVLQLPQ	137-163	27	-0.1089 (Probable Non-Antigen)	Probable Allergen	Probable Non-Toxin
10		TTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGD	165-216	52	0.5206 (Probable Antigen)	Probable Allergen	Probable Non-Toxin
11		RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA	226-267	42	0.5627 (Probable Antigen)	Probable Non-Allergen	Probable Non-Toxin
12		RRGPEQTQGNFGDQELIRQGTDYK	276-299	24	0.6277 (Probable Antigen)	Probable Allergen	Probable Non-Toxin
13		DAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDD	358-402	45	0.4968 (Probable Antigen)	Probable Allergen	Probable Non-Toxin
14		SKQLQQSMSSADS	404-416	13	0.3724 (Probable Non-Antigen)	Probable Non-Allergen	Probable Non-Toxin

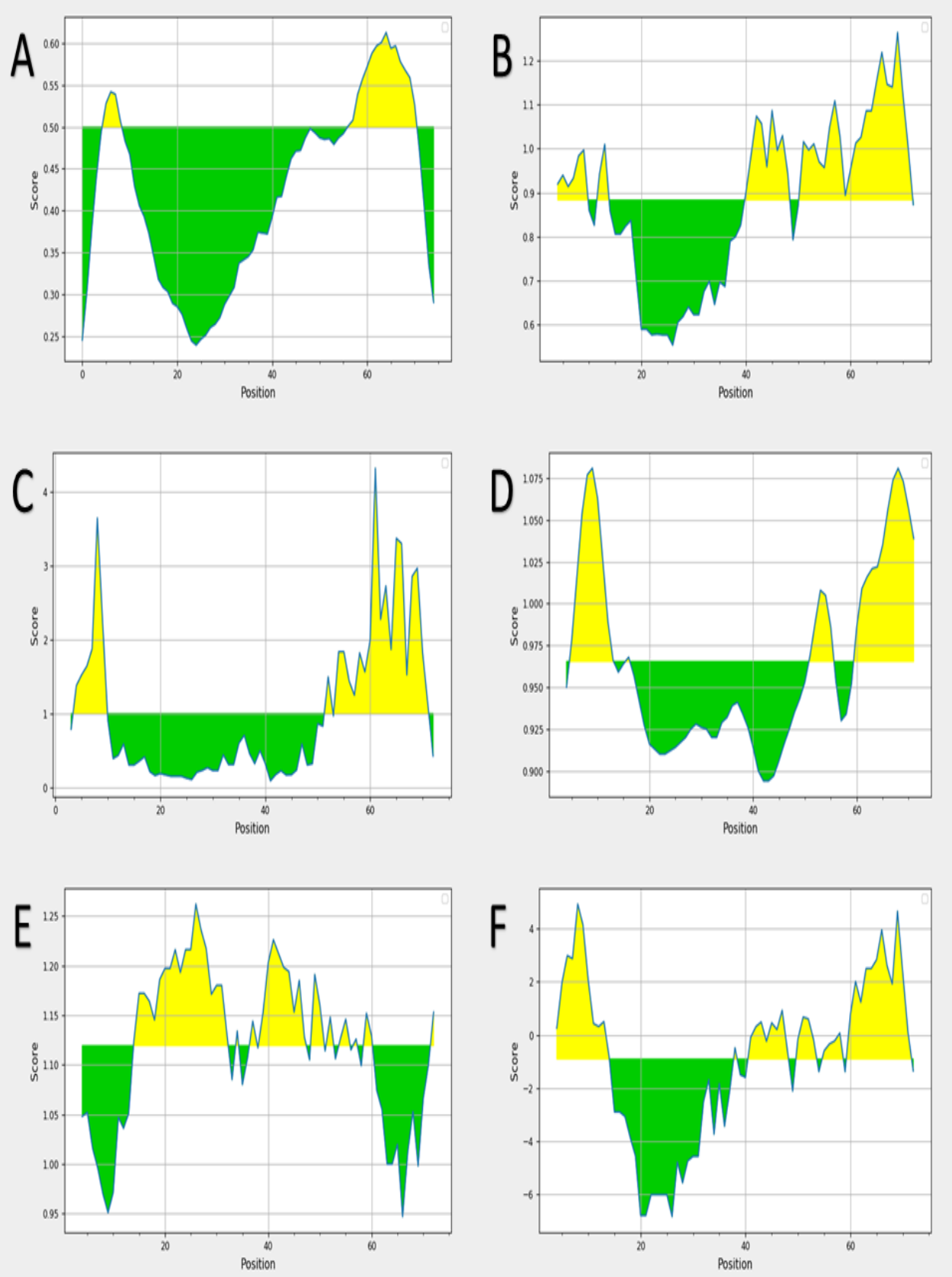

Figure 2. Prediction of B-cell epitopes in the Indonesian SARS-CoV-2 envelope protein using different parameters. A: BepiPred linear epitope prediction 2.0 (threshold: 0.500), B: Chou and Fasman beta-turn prediction (threshold: 0.883), C: Emini surface accessibility prediction (threshold: 1.000), D: Karplus and Schulz flexibility prediction (threshold: 0.965), E: Kolaskar and Tongaonkar antigenicity prediction (threshold: 1.119), and F: Parker hydrophilicity prediction (threshold: -0.911). The green area is a negative prediction of a B-cell epitope, whereas the yellow area is a positive prediction.

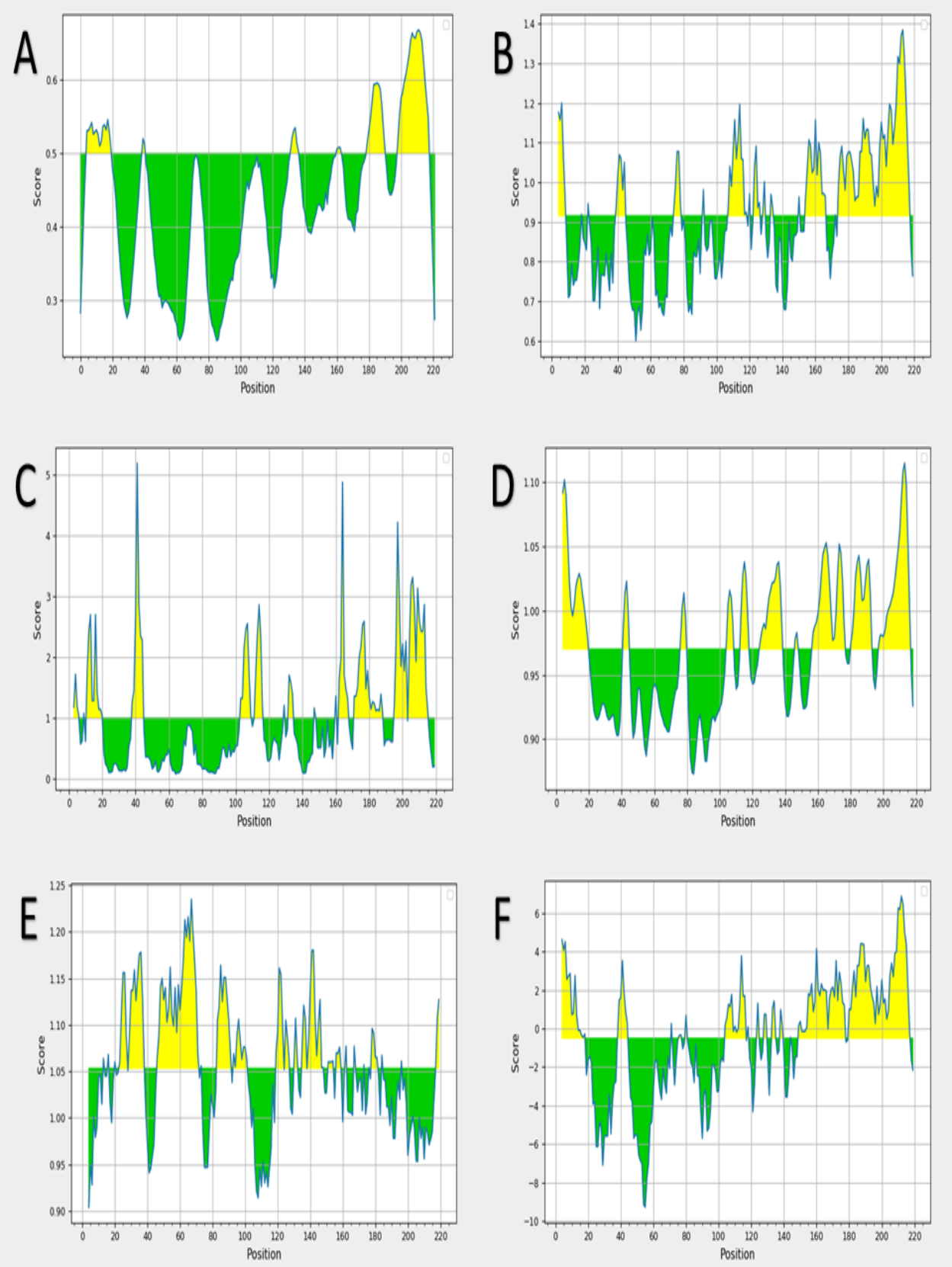

Figure 3. Prediction of B-cell epitopes in the Indonesian SARS-CoV-2 membrane protein using different parameters. A: BepiPred linear epitope prediction 2.0 (threshold: 0.500), B: Chou and Fasman beta-turn prediction (threshold: 0.915), C: Emini surface accessibility prediction (threshold: 1.000), D: Karplus and Schulz flexibility prediction (threshold: 0.970), E: Kolaskar and Tongaonkar antigenicity prediction (threshold: 1.053), and F: Parker hydrophilicity prediction (threshold: −0.499). The green area is a negative prediction of a B-cell epitope, whereas the yellow area is a positive prediction.

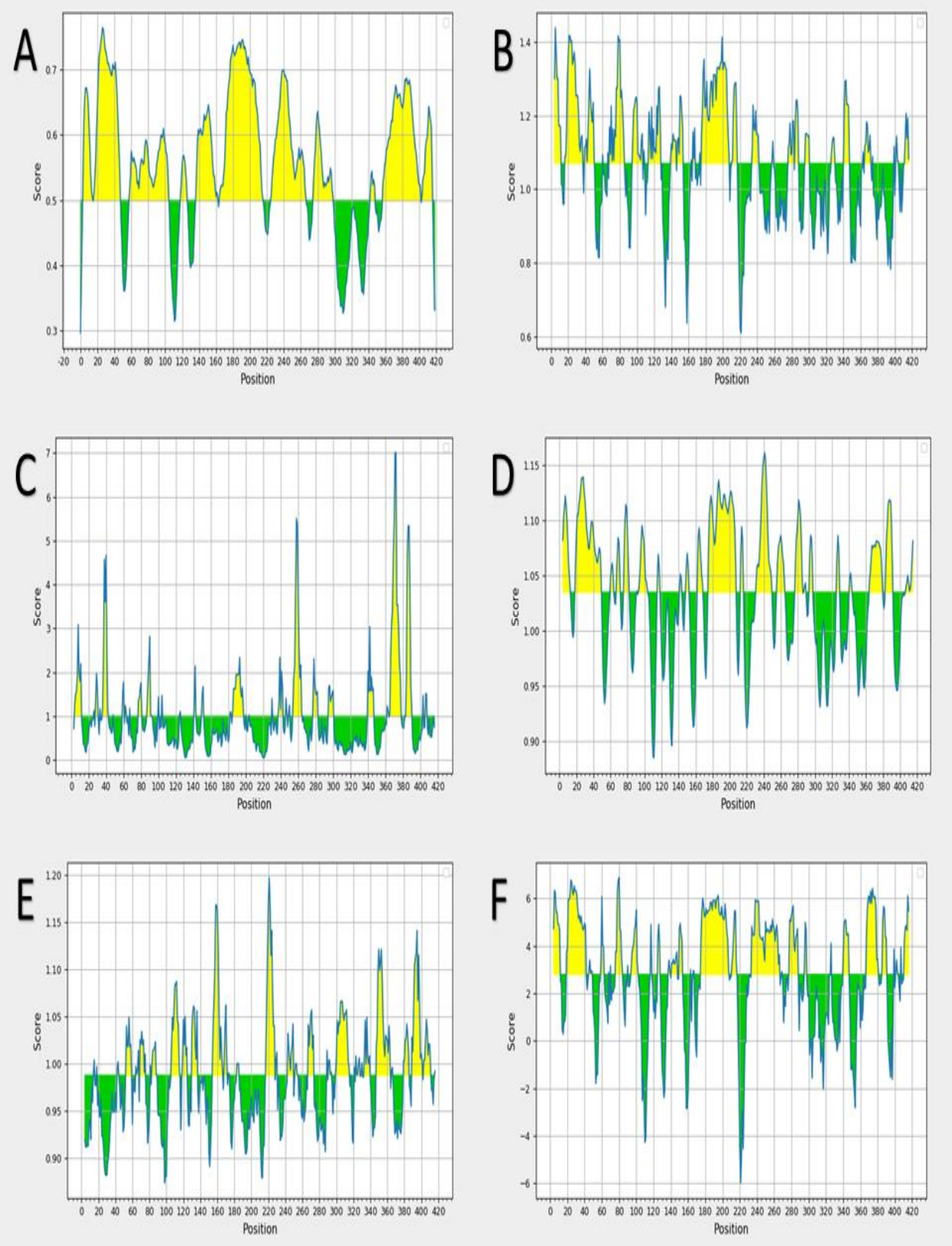

Figure 4. Prediction of B-cell epitopes in the Indonesian SARS-CoV-2 nucleocapsid phosphoprotein using different parameters. A: BepiPred linear epitope prediction 2.0 (threshold: 0.500), B: Chou and Fasman beta-turn prediction (threshold: 1.070), C: Emini surface accessibility prediction (threshold: 1.000), D: Karplus and Schulz flexibility prediction (threshold: 1.035), E: Kolaskar and Tongaonkar antigenicity prediction (threshold: 0.988), and F: Parker hydrophilicity prediction (threshold: 2.800). The green area is a negative prediction of a B-cell epitope, whereas the yellow area is a positive prediction.

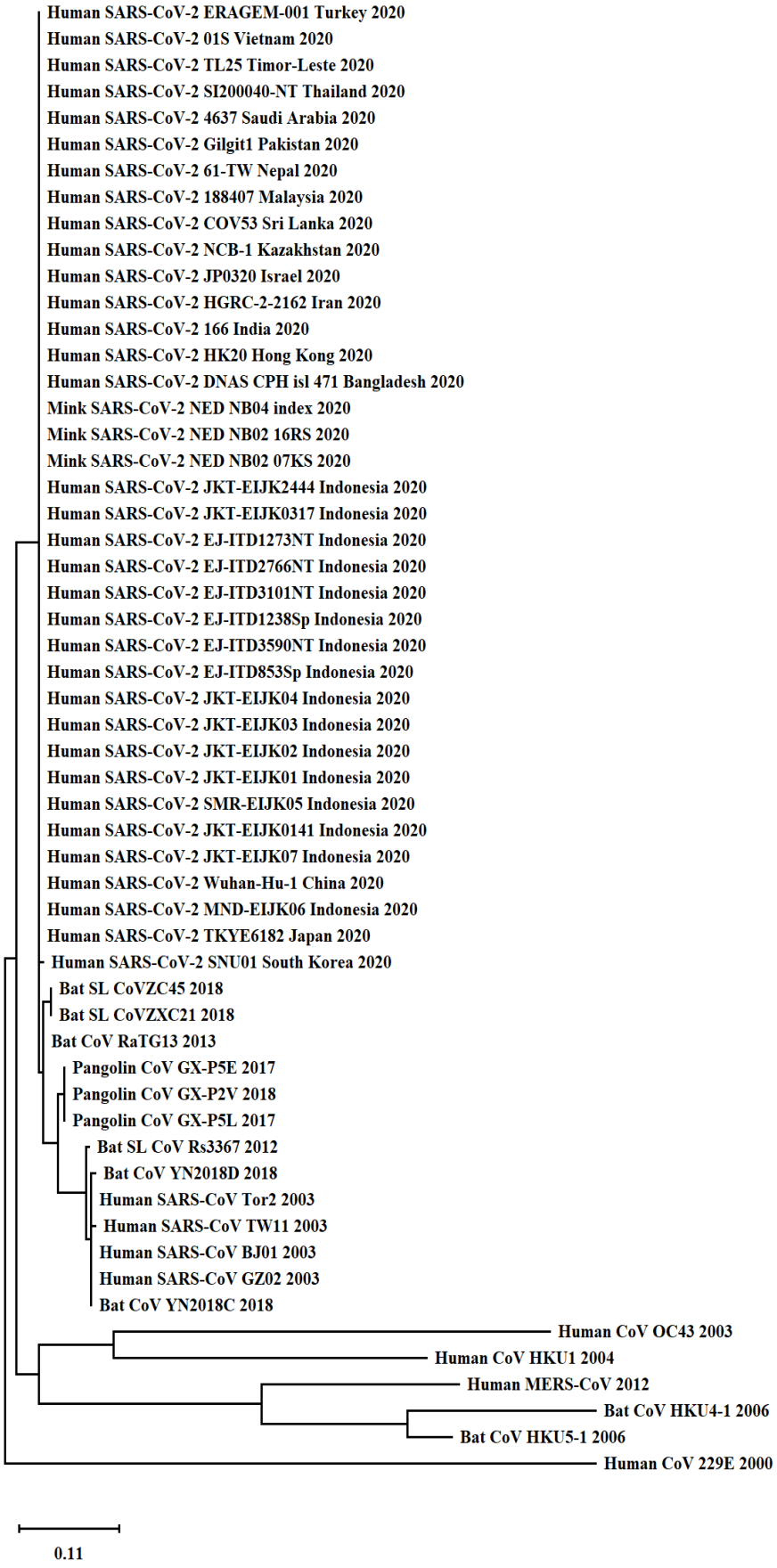

Figure 5. Molecular phylogenetic tree of the Indonesian SARS-CoV-2 isolates based on the envelope protein gene.

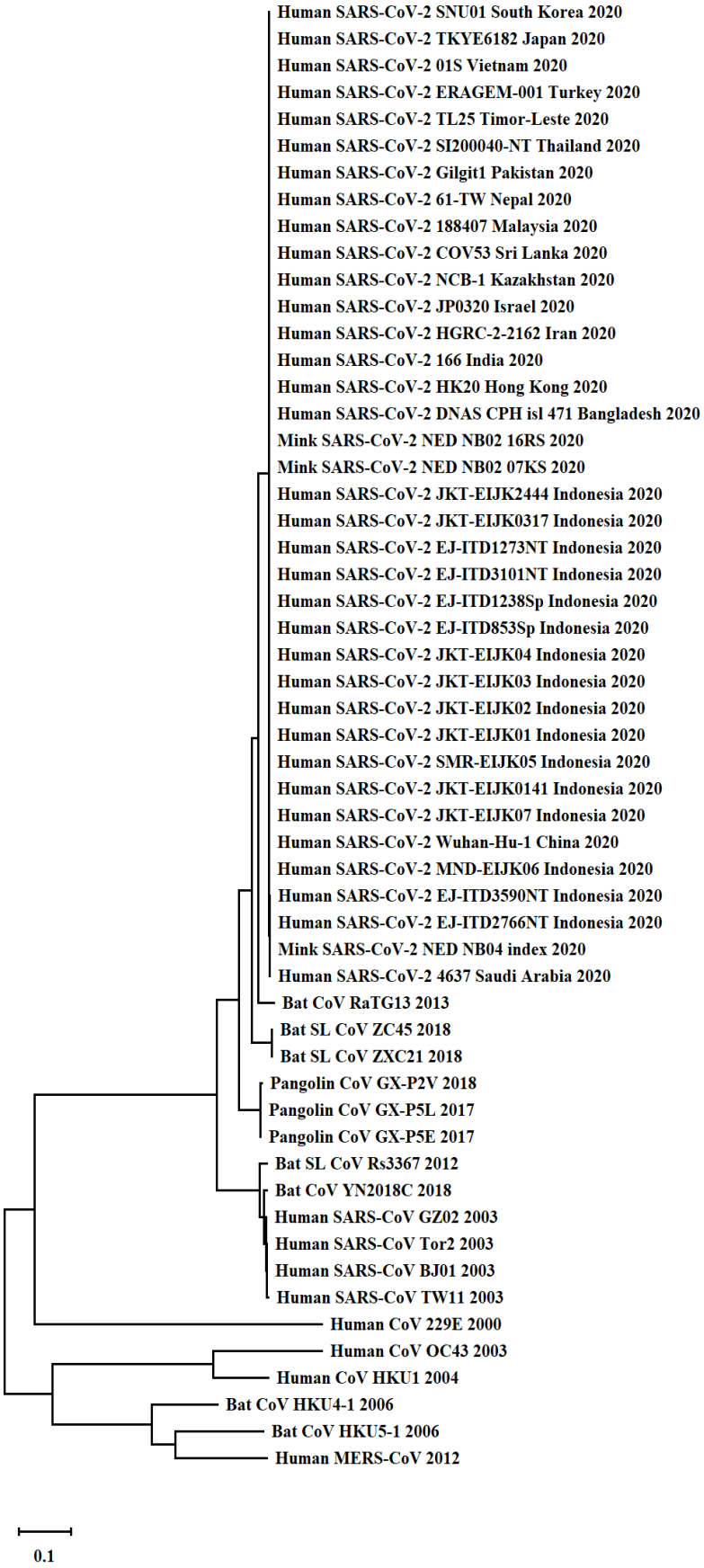

Figure 6. Molecular phylogenetic tree of the Indonesian SARS-CoV-2 isolates based on the membrane protein gene.

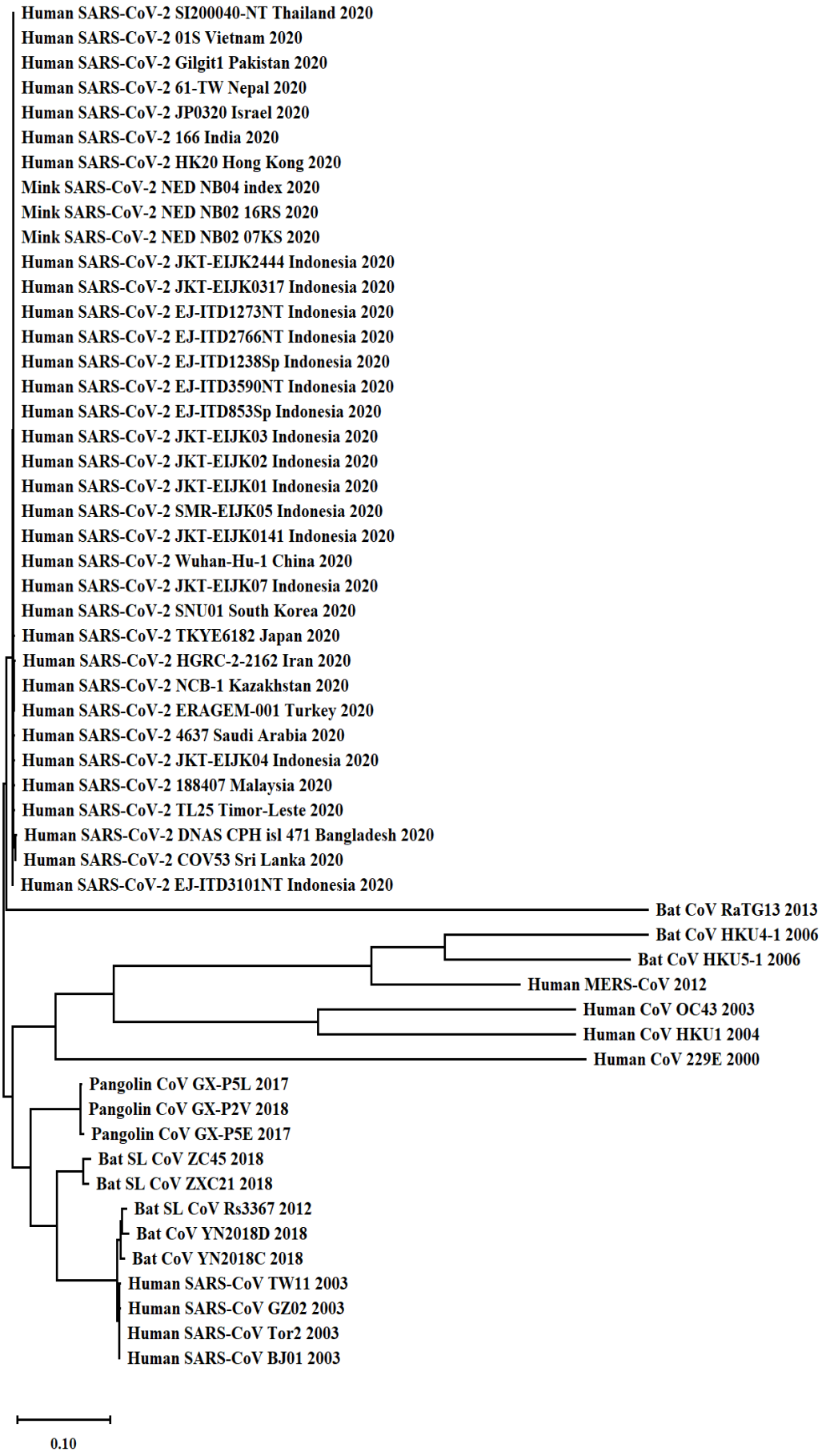

Figure 7. Molecular phylogenetic tree of the Indonesian SARS-CoV-2 isolates based on the nucleocapsid phosphoprotein gene.

Discussion

SARS-CoV-2, a new coronavirus, is deploying worldwide[17]. Since the virus appeared at the market at the end of 2019 in Wuhan, China[18], the number of cases has been increasing sporadically[19]. Furthermore, human-to-human transmission of the virus has been confirmed[17,18] and the possibility of aerosol transmission was noted by the WHO. The virus has been discovered in sputum, nasopharyngeal swabs, bronchoalveolar-lavage, throat, and saliva[17]. We retrieved 16 Indonesian isolates from the database (Table 1). These isolates were detected using a nasopharyngeal and oro-pharyngeal swab, nasopharyngeal swab, and sputum methods and submitted by the Institute of Tropical Disease, Universitas Airlangga, Surabaya, Indonesia, and the Eijkman Institute for Molecular Biology, Ministry of Research and Technology/National Agency for Research and Innovation, Jakarta, Indonesia. Thus, we conducted a novel investigation on the membrane, nucleocapsid, and envelope protein genes of Indonesian isolates.

Many studies have focused on the spike glycoprotein gene of the virus[20,21,22]. However, the other structural protein genes urgently need to be investigated in an extensive immunoinformatics study. Our genetic analysis has discovered many mutation sites in two genes (membrane protein and nucleocapsid phosphoprotein genes) and no mutation occurs in the envelope protein gene as shown in Table 2. However, there were no important alteration in the three structural protein genes of Indonesian SARS-CoV-2 isolates. Moreover, we identified that the genetic similarity is between 99.9 and 100% from the LALIGN web server analysis. In addition, there were four isolates with mutation sites, EJ-ITD2766NT (213C>T), EJ-ITD3590NT (213C>T), EJ-ITD3101NT (945C>T), JKT-EIJK04 (1041A>T). We only found amino acid changes in JKT-EIJK04 (K347N). Scientists have demonstrated that mutations occur in the virus genome globally. Previously, Phan *et al*. performed a genetic analysis in 86 virus genomes and reported many mutations. Furthermore, one of the most important mechanisms proposed for the evolution of viruses in nature is nucleotide substitution[17]. Garcés-Ayala *et al*. conducted a study with the reference sequence to identify the novel SARS-CoV-2 complete genome in Mexico[23]. In line with this, Yadav *et al*. also reported a study using Wuhan-Hu-1 isolates to analyze the first two virus isolates from India[24]. Khailany *et al*. retrieved 94 SARS-CoV-2 genomes and checked the molecular variation between them[25]. Kim *et al*. revealed that the quick transmission and infectivity of the virus correlated with specific mutations in the genome[26]. Furthermore, Zhang *et al*. stated that the spike glycoprotein mutation (D614G) is associated with the virulence of the virus[27]. Thus, to our knowledge, this study is the first to report the analysis of three structural protein genes in SARS-CoV-2 in Indonesia. These data might support further studies in the establishment of biological aspects of Indonesian SARS-CoV-2. In addition, without any experimental data, our interpretation using the limited data of Indonesian SARS-CoV-2 isolates will potentially benefit continuous and forthcoming research. At present, the GISAID EpiCoV database has established seven subtypes of SARS-CoV-2, specifically V, S, O, L, GR, GH, and G. Uniquely, all Indonesian isolates were categorized in the L clade except JKT-EIJK2444 (O), EJ-ITD2766NT (GH), and EJ-ITD3590NT (GH).

Duffy stated that the mutation rates in RNA viruses are much higher than in most other microorganisms. An elevated mutation rate can lead to an increase in virulence and a high potential for adaptive evolution. This capability boosts the chance of zoonotic viral pathogens to establish human-to-human transmission and permits them to enhance their virulence[28]. Moreover, our study provides fundamental data for accomplished studies into the medication and prevention of COVID-19. Furthermore, the Indonesian SARS-CoV-2 genomic data extraction would be valuable in vaccine construction and options in medication. In fact, mining the data of the Indonesian SARS-CoV-2 variants and molecular epidemiology could enable the mapping of its origin and the tracking of its transmission. In line with this, sequence investigation performs an important role in viral surveillance, public health policy problems, and host identification. Thus, high-speed detection of mutations from the Indonesian SARS-CoV-2 is mandatory in the answer to the COVID-19 pandemic in Indonesia.

Epitope prediction studies have been accomplished for some viruses, such as the influenza[29], dengue[30], and Ebola[31] viruses. In the present study, we developed a peptide-based vaccine using B-cell epitopes prediction with various parameters following analyses such as protective antigens prediction, allergenicity prediction, and protective non-toxic antigens prediction. We have revealed the epitope prediction of three structural proteins in Indonesian SARS-CoV-2 isolates based on the IEDB with various parameters such as Parker hydrophilicity prediction, Chou and Fasman beta-turn prediction, Karplus and Schulz flexibility prediction, Kolaskar and Tongaonkar antigenicity prediction, and Emini surface accessibility prediction using default thresholds (Figures 2–4). We used the VaxiJen v.2.0 for predicting the peptides that could be prospective protective antigens to establish an immune response (Table 3). We used AllerTOP to analyze the predicted peptides allergenicity and performed a protective non-toxic antigen prediction by using ToxinPred (Table 3).

We propose that the peptide RRGPEQTQGNFGDQELIRQGTDYK from nucleocapsid phosphoprotein can be used to generate a peptide-based vaccine contrary to SARS-CoV-2. Notable epitope prediction methods are primary significance in some biotechnological and clinical applications, such as therapeutic antibody and vaccine initiation, or theoretical studies of immune systems[10]. Presently, research groups are racing to develop vaccines against SARS-CoV-2 worldwide, with protein-based vaccines becoming one of the most developed vaccines[10]. Thereto, Dimitrov *et al*. specified that AllerTOP is an updated effective allergen prediction. It analyzes peptides based on the physicochemical composition of the protein sequences[33]. In addition, Gupta *et al*. reported the use of the ToxinPred web server for predicting the toxicity of peptides[15]. Thus, we used this to reveal the safety of the predicted peptides in this study.

Elucidating the transmission routes, relationships, and origin of the causative agents of emerging infections is crucial to understanding the possible approach of interference and their biological action. Many scientists have established the molecular phylogenetic analysis of SARS-CoV-2 with that of previous coronaviruses using specific or whole-genome sequences to comprehend recombination events and an evolutionary chronicle[34,35,36]. According to the genome isolates of coronaviruses presently available, the whole-genome phylogenetic tree designates that SARS-CoV-2 is closest to *Rhinolophus affinis* coronavirus RaTG13, followed by pangolin coronavirus[37]. In line with this, Tu *et al*. stated

that SARS-CoV-2 is accepted to have issued in bats and may have used pangolins as an intermediate host before transmission to humans[38]. In addition, Zhou *et al*. stated that the SARS-CoV-2 genome shares about 80% of its genome to SARS-CoV[39]. Andersen *et al.* reported that the two probable frameworks that might illustrate the emergence of SARS-CoV-2[37].

We promoted the molecular phylogenetic tree and revealed the relationship between Indonesian SARS-CoV-2 isolates, other Asian isolates, and other groups of coronaviruses originating from humans, pangolins, and bats (Figures 5–7). Here, we report the first molecular phylogenetic tree of 16 Indonesian isolates based on the envelope, nucleocapsid, and membrane protein genes. Molecular phylogenetic analysis is used for applied and fundamental virus research, including studies of the taxonomy, evolution, and origin[1]. It is fundamental to investigate the likelihood of SARS-CoV-2 intermediate hosts to understand and contain the transmission of COVID-19[39]. In addition, Lam *et al*. found that coronaviruses currently occur in many wild mammals in Asia[1]. Based on this study, we advise that further surveillance investigations be carried out on various mammals in their natural environment, including pangolins and bats, especially in Asia, to contain the chance of future zoonotic transmissions.

Conclusion

In summary, we propose that the peptide RRGPEQTQGNFGDQELIRQGTDYK from the nucleocapsid phosphoprotein can be used to generate a peptide-based vaccine contrary to SARS-CoV-2. However, further trials such as *in vitro* and *in vivo* testing are involved for validation.

Financial Support

The authors are thankful for financial support via the PMDSU Grant (Batch III), provided by the Ministry of Education and Culture, Republic of Indonesia. This study was additionally supported by the Professor Nidom Foundation, Indonesia.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgement

We thank International Science Editing for the help in editing the manuscript. We would like to declare our heartfelt sympathy to the victims of COVID-19. Tribute goes to the frontliners globally, notably in Indonesia.

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. Li X, Giorgi EE, Marichann MH, et al. (2020) Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci Adv. 6(27): eabb9153.
3. Huang C, Wang Y, Li X, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 395(10223): 497-506.
4. 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.
5. 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.
6. 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.
7. Zeng W, Liu G, Ma H, et al. (2020) Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem Biophys Res Commun. 527(3): 618-623.
8. Ahmed SF, Quadeer AA, McKay MR. (2020) Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 12(3): 254.
9. Raza H, Wahid B, Rubi G, et al. (2020) Molecular epidemiology of SARS-CoV-2 in Faisalabad, Pakistan: A real-world clinical experience. Infect Genet Evol. 84: 104374.
10. Callaway E. (2020) The race for coronavirus vaccines: A graphical guide. Nature. 580: 576-577.
11. Jean SS, Lee PI, Hsueh PR. (2020) Treatment options for COVID-19: The reality and challenges. J Microbiol Immunol Infect. 53(3): 436-443.
12. Joob B, Wiwanitkit V. (2020) Genetic variant severe acute respiratory syndrome coronavirus 2 isolates in Thailand. J Pure Appl Microbiol. 14: 6314.
13. Dhanda SK, Mahajan S, Paul S, et al. (2019) IEDB-AR: immune epitope database-analysis resource in 2019. Nucleic Acids Res. 47(W1): W502-W506.
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. Kumar S, Stecher G, Li M, et al. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35: 1547-1549.
17. Phan T. (2020) Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 81: 104260.
18. Zhu N, Zhang D, Wang W, et al. (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 382(8): 727-733.
19. Velavan TP, Meyer CG. (2020) The COVID-19 epidemic. Trop Med Int Health. 25(3): 278-280.
20. Yin C. (2020) Genotyping coronavirus SARS-CoV-2: Methods and implications. Genomics. S0888-7543(20): 30318-30319.
21. 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.
22. Watanabe Y, Allen JD, Wrapp D, et al. (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 369(6501): 330-333.
23. Garcés-Ayala F, Araiza-Rodríguez A, Mendieta-Condado E, et al. (2020) Full genome sequence of the first SARS-CoV-2 detected in Mexico. Arch Virol.
24. Yadav PD, Potdar VA, Choudhary ML, et al. (2020) Full-genome sequences of the first two SARS-CoV-2 viruses from India. Indian J Med Res. 151(2&3): 200-209.
25. Khailany RA, Safdar M, Ozaslan M. (2020) Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19: 100682.

26. Kim JS, Jang JH, Kim JM, et al. (2020) Genome-wide identification and characterization of point mutations in the SARS-CoV-2 genome. Osong Public Health Res Perspect. 11(3): 101-111.
27. Zhang L, Jackson CB, Mou H, et al. (2020) The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv.
28. Duffy S. (2018) Why are RNA virus mutation rates so damn high? PLoS Biol. 2018; 16: e3000003.
29. Lohia N, Baranwal M. (2020) An immunoinformatics approach in design of synthetic peptide vaccine against influenza virus. Methods Mol Biol. 2131: 229-243.
30. Himmah K, Fitriyah, Ardyati T, et al. (2017) Designing a polytope for use in a broad-spectrum dengue virus vaccine. J Taibah Univ Med Sci. 13(2): 156-161.
31. Kadam A, Sasidharan S, Saudagar P. (2020) Computational design of a potential multi-epitope subunit vaccine using immunoinformatics to fight Ebola virus. Infect Genet Evol. 85: 104464.
32. 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.
33. 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.
34. Wu F, Zhao S, Yu B, et al. (2020) A new coronavirus associated with human respiratory disease in China. Nature. 579(7798): 265-269.
35. Lu R, Zhao X, Li J, et al. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395(10224): 565-574.
36. Tang X, Wu C, Li X, et al. (2020) On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev.
37. Andersen KG, Rambaut A, Lipkin WI, et al. (2020) The proximal origin of SARS-CoV-2. Nat Med. 26: 450-452.
38. Tu YF, Chien CS, Yarmishyn AA, et al. (2020) A Review of SARS-CoV-2 and the ongoing clinical trials. Int J Mol Sci. 21(7): 2657.
39. 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.