© 2022 Journal of Pharmacy & Pharmacognosy Research, 10 (3), 429-444, 2022
ISSN 0719-4250
https://jppres.com

DOI: https://doi.org/10.56499/jppres21.1262_10.3.429

Original Article

Molecular characterization and prediction of B-cell epitopes for the development of SARS-CoV-2 vaccine through bioinformatics approach

[Caracterización molecular y predicción de epítopos de células B para el desarrollo de una vacuna contra el SARS-CoV-2 mediante un enfoque bioinformático]

Aamir Shehzad[1], Martia Rani Tacharina[1], Suryo Kuncorojakti[2], Hafiz Ishfaq Ahmad[3], Rofiqul A'la[1], Andi Yasmin Wijaya[4], Wiwiek Tyasningsih[5], Fedik Abdul Rantam[1,4*]

[1]Virology and Immunology Laboratory, Department of Microbiology, Faculty of Veterinary Medicine, Airlangga University, Surabaya, East Java, 60115, Indonesia.
[2]Division of Veterinary Anatomy, Department of Veterinary Science, Faculty of Veterinary Medicine Airlangga University, Surabaya, East Java, 60115, Indonesia.
[3]Department of Animal Breeding and Genetics, University of Veterinary and Animal Sciences, Lahore, Pakistan.
[4]Research Center for Vaccine Technology and Development, Institute of Tropical Disease, Universitas Airlangga, Surabaya, East Java, 60115 Indonesia.
[5]Bacteriology and Mycology Laboratory, Department of Microbiology, Faculty of Veterinary Medicine, Airlangga University, Surabaya, East Java, 60132, Indonesia.
*E-mail: fedik-a-r@fkh.unair.ac.id

Abstract

Context: The SARS-CoV-2 virus is the cause of the COVID-19 pandemic, which is a severe public health crisis worldwide.

Aims: To analyze the SARS-CoV-2 isolates of Surabaya and predict ORF1ab polyprotein epitopes through the bioinformatics approach for vaccine candidate development.

Methods: Three genomic sequences of Surabaya isolates were obtained from the GISAID, NCBI and PDB Gen-bank databases and MEGA-11 software were used to understand the transformations in the isolates. The IEDB and VaxiJen, AllerTop, and ToxinPred web servers were used to predict B-cell epitopes and analyze their antigenicity, non-allergenicity, non-toxicity, respectively. Moreover, these epitopes were linked by EAAAK for 3D modeling, refinement, and validation through Swiss- Model, Galaxy Refine, and RAMPAGE web tools.

Results: The Surabaya isolates, RSDS-RCVTD-UNAIR-49-A, 54-A, and 42-A, had 10, 20, and 16 mutations in nucleotides and depicted a phylogenetically close relationship to isolates of Egypt, Pakistan, and Bangladesh, respectively. A total of 71 sequential Orf1ab B-cell epitopes were predicted, and only three peptides were found to be antigenic, non-allergenic, and non-toxic. These epitopes were linked with the EAAAK linker to develop a 3D refined and validated structure. This construct was docked with TLR-3 receptor by the Cluspro webserver and found a high affinity of ORF1ab+TLR3 due to 15 hydrogen bonds. The construct demonstrated good humoral and cellular immune responses in the C-ImmSim server, and cloning in the expression vector pET28a (+) yielded a colon of 846bp.

Conclusions: ORF1ab B-cell epitopes could be useful for developing effective vaccines to combat SARS-CoV-2 infection.

Keywords: bioinformatics; epitopes; ORF1ab polyproteins; public health; Indonesia; SARS-CoV-2.

Resumen

Contexto: El virus SARS-CoV-2 es la causa de la pandemia de COVID-19, que es una grave crisis de salud pública a nivel mundial.

Objetivos: Analizar los aislamientos de SARS-CoV-2 de Surabaya y predecir los epítopos de poliproteína ORF1ab mediante el enfoque bioinformático para el desarrollo de candidatos vacunales.

Métodos: Se obtuvieron tres secuencias genómicas de aislamientos de Surabaya de las bases de datos GISAID, NCBI and PDB Gen-bank y el software MEGA-11 para comprender las transformaciones en los aislamientos. Se utilizaron los servidores web IEDB y VaxiJen, AllerTop y ToxinPred para predecir epítopos de células B y analizar su antigenicidad, no alergenicidad y no toxicidad, respectivamente. Además, EAAAK vinculó estos epítopos para el modelado, el refinamiento y la validación en 3D a través de las herramientas web Swiss-Model, Galaxy Refine y RAMPAGE.

Resultados: Los aislamientos de Surabaya, RSDS-RCVTD-UNAIR-49-A, 54-A y 42-A, tenían 10, 20 y 16 mutaciones en nucleótidos y mostraban una relación filogenéticamente cercana con los aislamientos de Egipto, Pakistán y Bangladesh, respectivamente. Se predijeron un total de 71 epítopos de células B Orf1ab secuenciales, y solo tres péptidos resultaron ser antigénicos, no alergénicos y no tóxicos. Estos epítopos se vincularon con el enlazador EAAAK para desarrollar una estructura 3D refinada y validada. Esta construcción fue acoplada con el receptor TLR-3 por el servidor web Cluspro y encontró una alta afinidad de ORF1ab+TLR3 debido a 15 enlaces de hidrógeno. La construcción demostró buenas respuestas inmunitarias celulares y humorales en el servidor C-ImmSim, y la clonación en el vector de expresión pET28a (+) produjo un colon de 846 pb.

Conclusiones: Los epítopos de células B ORF1ab podrían ser útiles para desarrollar vacunas efectivas para combatir la infección por SARS-CoV-2.

Palabras Clave: bioinformática; epítopos; poliproteínas ORF1ab; Indonesia; salud pública; SARS-CoV-2.

ARTICLE INFO
Received: October 25, 2021.
Received in revised form: January 7, 2022.
Accepted: January 14, 2022.
Available Online: February 6, 2022.

INTRODUCTION

COVID-19 epidemic is an important global health catastrophe today. SARS-CoV-2 is an enveloped ssRNA virus, and an early phylogenetic study of its genome suggested that the virus was most closely related to the SARS RaTG13 virus of the horseshoe bat (Liu et al., 2021). The SARS-CoV virus caused the outbreak of Severe Acute Respiratory Syndrome in 2003; therefore, the International Committee on Taxonomy of Viruses is referred to as the SARS-CoV-2 virus (WHO, 2020a).

COVID-19 belongs to the genera *Betacoronaviridae*, known to have frequent cross-species transmission. The coronaviruses name was driven from their envelope-embedded crown-like-spike proteins that determine the viral host/cell-type specificity, receptor recognition and cell membrane fusion process (Tang et al., 2020).

Several coronaviruses can infect humans, and the disease can range from the mild condition like the common cold (HCoV-229E, HCoV-NL63, HCoV-HKU1, and HCoV-OC43) to severe disease or even death (MERS-CoV, SARS-CoV and COVID-19) (Huang et al., 2020a); Dandekar and Perlman, 2005). The SARS-CoV-2 disease pathogenesis may be due to direct viral infection, cytokine dysregulation, or coagulopathy (Cevik et al., 2020). In SARS, antibody to S-protein is protective, but CD8+ T-cell mediated damage is responsible for acute respiratory distress syndrome, multi-organ failure, and cardiac injury (Zheng et al., 2020). ACE2 and dipeptidyl peptidase 4 are two COVID-19 receptors and transducers involved in normal physiological processes such as maintaining glucose homeostasis, renal and cardiovascular physiology, and regulating inflammation. Hence, any disturbance in these receptors leads to many physiological disturbances in many organs (Valencia et al., 2020).

COVID-19 can affect all population's groups, with a higher incidence in men and a higher fatality rate in geriatric and critically ill patients (Huang et al., 2020b). The SARS-CoV-2 virus emerged in Wuhan city, China, in December 2019 and then spread rapidly worldwide. On August 27, 2021, the WHO announced that there are about 214.27 million confirmed cases with 4.27 million deaths in 220 different countries with more prevalence in the United States (38.15 M), India (32.60 M), and Brazil (20.64 M), and that implies high incidence rate of infection with 0.28% of death (WHO, 2021).

COVID-19 has a zoonotic origin and reservoir in bats, so infected animals have transmitted the virus to humans. COVID-19 is highly contagious and can transmit by respiratory droplets (cough/exhale), direct contact of mucous membranes, the fecal-oral route, or contact with any excreta containing the living virus (Yadav et al., 2020). Vertical transmission of SARS-CoV-2 is also reported (Dong et al., 2020). In addition, people can catch COVID-19 by touching contaminated objects (Cai et al., 2020). There is also the awareness that pets can spread the disease (Fritz et al., 2020), or consuming raw or uncooked animal products may be a source of infection (WHO, 2020b).

The median incubation period of COVID-19 is five days, and most patients will develop symptoms on the 11th to 15th day after infection. Therefore, it has been recommended to quarantine those exposed to infection for 14 days (Lewis et al., 2020). The inhaled virus binds to ciliated epithelial cells ACE2 receptor in the nasal cavity and multiply. This binding is mediated by virus S protein and facilitated by the cellular transmembrane serine protease 2 (TMPRSS2 protease).

The anti-inflammatory process is weakened upon this binding, and the angiotensin II function is exaggerated. The person is asymptomatic but infectious, and nasal swabs can detect the virus at this stage. Next few days, the virus migrates down the respiratory tract. It triggers proinflammatory cytokines (interleukin IL-6, IL-10 and TNF-α, granulocyte colony-stimulating factor), accumulation of free radicals, changes in intracellular pH, and accumulation of lactic acid with subsequent hypoxia and cardiopulmonary changes (Han et al., 2020).

According to the complete genome sequence analysis, the evolutionary relationship rate of SARS-CoV 2 with bat SARS coronavirus (SARSr-CoV RaTG13) was 96.00%, respectively (Zhou et al., 2020). This indicates that SARS-CoV-2 may have originated in bats. These Coronaviruses are enveloped viruses with a diameter of 80 to 120 nm. The virion surface is made up of three proteins: spike (S), membrane (M), and tiny membrane protein (E), which gives the virus a crown-like appearance under electron microscopy (Bond et al., 1979; Weiss and Navas-Martin, 2005). The ORF1ab polyprotein, which is found at the viral genome's five prime ends, encodes for 15 or 16 non-structural proteins. It accounts for two-thirds of the viral proteome. Other than structural and non-structural proteins, some accessory proteins were discovered, including ORF7b, ORF3a, orf7a, ORF8, ORF10, and ORF6 (Chan et al., 2020a; Zhang et al., 2020). The non-structural gene ORF1ab, which includes ORF1a and ORF1b, is the most significant gene component of SARS-CoV-2 (Biswas et al., 2020). Papain-like protease (PLpro) and

3C-like protease cleave the replicase ORF1ab (3CLpro). In addition, many non-structural proteins (NSP1-NSP16) are cleaved from ORF1ab (Biswas et al., 2020; Khailany et al., 2020). Furthermore, it has been demonstrated that proteins or protein domains encoded by ORF1ab may play particular roles in virus-cell interactions, virulence, and virus-host response variations (Graham et al., 2008). A recent reverse genetic investigation found that ORF1ab polyprotein proteins play a role in cellular signaling, cellular gene expression modification, and pathogenicity. Furthermore, the ORF1ab polyproteins, particularly NSP3, have been suggested to interact with various structural and non - structural proteins and regulatory regions of viral RNA (Graham et al., 2008). Moreover, it was observed that after SARS-CoV-2 infection, there were increases in Orf protein-based antibodies (Hachim et al., 2020). The SARS CoV-2 ORF1ab polyprotein has the capacity to significantly activate T-cells in COVID19 patients (Gangaev et al., 2020a). The ORF1ab polyprotein of SARS-CoV-2 has been identified as a virulence factor and a key agent in the spread of COVID-19 in a population (Emameh et al., 2020). B-cell epitopes are essential for activating B-cells and generating the initial immune response, leading to antibody production and the establishment of long-time immunity in the form of memory cells (Ghoshal et al., 2021). The current work is intended to analyze Surabaya SARS-CoV-2 isolates molecularly and evaluate the immunogenic characteristics of ORF1ab (non-structural protein) by predicting B-cell epitopes using bioinformatics approach.

MATERIAL AND METHODS

Since the announcement of the SARS-CoV-2 pandemic crisis in Indonesia on March 15 2020, the University of Airlangga has designated a specific laboratory for COVID-19 diagnostic and research in the "Research Center for Vaccine Technology and Development, Institute of Tropical Diseases" (RCVTD-ITD), Laboratory, Biosafety Level-3 (BSL-3) facility. During the pandemic, oropharyngeal and nasopharyngeal swabs were collected in ITD-affiliated hospitals. Swabs were obtained from patients following World Health Organization (WHO) recommendations and soaked in 3 mL of the viral transport medium (VTM) containing a sterile-filtered solution of MEM (Gibco, USA), 1× penicillin-streptomycin (Gibco, USA), and 1× amphotericin B (Gibco, USA) as reported (Chan at el., 2020b), with modification before being sent for SARS-CoV-2 molecular testing. A total of 35 positive nasopharyngeal samples for this study were obtained from Dr. Soetomo-Teaching Hospital in Surabaya, Indonesia, which resulted in six isolates.

Virus SARS-CoV-2 isolation

The isolation of SARS-CoV- 2 was carried out in Vero-E6 of ATCC (United States of America). A T25 flask from Corning USA was used for seeding under the cell count of 2×10^6. Moreover, it was cultured with a minimum essential medium (MEM) (Gibco, USA), supplemented with 10% fetal bovine serum (FBS) (Gibco, USA), until it reached 80% Confluency (37°C and 5% CO_2). Following incubation, 4 mL of fresh MEM (Gibco, USA) was added with 10% BSF (Gibco, USA) (Martin et al., 2020). Cell labeling was conducted as an inoculation culture and was checked daily with an inverted TMS phase-contrast microscope (Nikon, Japan). Blind passage to new monolayer Vero cell culture was performed three days after inoculation by injecting 1 mL supernatant media from infected cells, followed by a one-hour incubation before adding 4 mL of fresh MEM supplemented with 10% FBS medium. The cytopathogenic effect (CPE) and plaque-forming unit (PFU) assays were used to evaluate viral growth (Park et al., 2020).

Extraction of viral RNA and real-time PCR

According to the manufacturer's suggested protocols, the extraction of RNA was carried out using Trizol reagent (Thermo Fisher, USA). According to the kit's manual, the Qubit TM RNA BR Assay Kit and Qubit TM fluorometer (Thermo Fisher Scientific, USA) were accomplished (Falzone et al., 2020). The Goscript® RT-PCR system (Promega, USA) was employed to perform RT-PCR of extracted RNA samples of SARS-CoV-2 through spike gene receptor binding domain primers forward: 5'-CCACAGACACTTGAGATTC-3' and reverse: 5'-GCAACTGAATTTTCTGCACCA-3') under the previously described (Lau et al., 2020).In a Thermal Cycler XP machine, conventional PCR was conveyed from the previously reported RT-PCR cDNA using the GoTaq® green primer and a master mixture (Promega, USA) (Bioer Technology, China). At first, denaturation was performed at 95°C for five minutes, and then amplification was achieved at 45 cycles of ten seconds at 95°C. Annealing was performed at 72°C for ten seconds, and final elongation was completed at 72°C for five minutes. Finally, the termination PCR reaction was carried on at 4°C for 30 minutes.

Whole-genome sequencing (WGS) of SARS-CoV-2

The SARS-CoV2 WGS was carried out on the targeted sequence after gene amplification with ARTIC V3 primer sets RT-PCR, followed by PCR clean up and processing with Nanopore sequencing kits on the GRIDION platform, according to the given protocols (Li et al., 2020a).

Genetika Science Indonesia used EPI2ME software, and Medaka software was used to trim the data. RAMPART software was used to analyze mutations and identify the transient gene and Wuhan Hu-1 as a reference sequence (Mapleson et al., 2015).

Molecular characterization of SARS-CoV-2 isolates

After screening six isolates, only three WGSs of SARS-CoV-2 (1366503, 1366505, 1366509) with B.1 Linage were selected. Fifteen whole-genome sequences of SARS-CoV-2 from other countries: Bangladesh, India, Pakistan, South Africa, Brazil, Egypt, France, Italy, USA, UK, Japan, South Korea, Malaysia, Singapore, UAE were carefully chosen with the same linage of B.1. on June 27 2021 and downloaded from the GISAID EpiCoV database (https://www.gisaid.org). Wuhan/WH04 WGS of SARS-CoV-2 was considered a reference and downloaded from the GISAID EpiCoV database.

Editing, alignment of whole-genome sequences of SARS-CoV-2 and phylogenetic tree construction

The required editing of WGSs of SARS-CoV-2 was completed using Bio edit 7.2 software. Sequences with many ambiguous sites as long stretches of Ns were eliminated and were replaced with reference Wuhan/WH04 whole-genome sequences of SARS-CoV-2. Before analysis, whole-genome sequence (WGS) alignment (EGS) was attained through the CLUSTAL W program in Mega 11 molecular evolutionary genetic analysis software (Saitou and Nei, 1987). The phylogenetic analysis was carried out using the Neighbor-Joining method with a bootstrap of 1000 present in Mega 11 software (Waterhouse et al., 2018).

Translation of WGS into amino acids and confirmation of identified ORF1ab segment of a sequence

The WGS was converted into amino acids (candidate or primary structure proteins) through the ExPASy translator (https://web.expasy.org/translate) (Hernández-Huerta et al., 2021). In addition to this, the selected segments were verified through the NCBI protein BLAST webserver.

Predictions of antigenic, non-toxic, nonallergic ORF1ab B-cell epitopes and physicochemical properties

The IEDB webserver http://tools.iedb.org/bcell/ was employed to predict the B-cell epitopes of B.1 isolates from Indonesian using the default criteria described earlier (Dhanda et al., 2019; Ansori et al., 2020; Adianingsih and Kharisma, 2019). The predicted peptides were then tested by the online webserver VaxiJen v2.0; (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) by applying the default threshold to see if the anticipated epitopes could act as protective antigens in the immune response (Abraham et al., 2021). In this study, the predicted peptide allergenicity prediction was analyzed using the webserver AllerTOP (https://www.ddg-pharmfac.net/AllerTOP/) and default settings. According to Abraham et al. (2021), the predicted peptides were submitted to this web service. The non-toxic protective antigens were then predicted using the ToxinPred webserver (https://webs.iiitd.edu.in/raghava/toxinpred/multi_submit.php and https://crdd.osdd.net/raghava/toxinpred). The standard levels used were described by Gupta et al. (2013) and Larsen et al. (2007). The peptides were then physiochemically predicted using ProtParam's web server (https://web.expasy.org/protparam/) (Gasteiger et al., 2005).

Linkage and secondary structure analysis of B-cell epitopes

The predicted ORF1ab B-cell epitopes were linked through the EAAAK and the SOPMA tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). SOPMA tool was used to analyze the secondary structure of construct (Geourjon and Deleage, 1995).

Development of ORF1ab B-cell epitopes tertiary structure, refinement and validation

The tertiary structure of Predicted B-cell epitopes was developed by using SWISS-MODEL: https://swissmodel.expasy.org/ (Waterhouse et al., 2018) The developed 3D structure refinement was done through the Galaxy Refine webserver: http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE and the validation of refining structure was done through the RAMPAGE server: https://saves.mbi.ucla.edu/ (Naveed et al., 2021).

Molecular docking of ORF1ab and receptor TLR-3

The selected refined validated 3D structure of ORF1ab B-cell epitopes is subjected to Cluspro-.2 webserver: https://cluspro.org/home.php for analyzing the binding interaction with receptor TLR-3 (Kathwate, 2022).

Host immune simulation and *in silico* cloning method

For the codon optimization, expression development, and reverse translation, the "Java Codon Adaptation Tool" (JCat) has been employed (Grote et al.

2005). The optimization methodology was applied to develop a vaccine via organism *E. coli* as a host (strain K12). The additional options were selected to optimize the Rho-independent transcription, ribosomal binding sites and restriction enzymatic cleavage sites (Abraham et al., 2021). The host's immune response to the vaccine was propagated through the online C-ImmSim webserver (Rapin et al., 2010). The simulation steps were fixed at 1050, and the simulation volume was set to default (Castiglione et al., 2012). The *E. coli* pET-28a (+) expression vector was used for cloning, and its nucleotide sequence was obtained from the Addgene vector database (Naveed et al., 2021). SnapGene v3.2.1 (GSL Biotech LLC., California, USA) was used to predict peptide-based vaccines (Naveed et al., 2021; Singh et al., 2020).

RESULTS

Virus isolation and viral morphology reverse transcription-polymerase chain reaction (RT-PCR)

According to virus isolation and morphology, Reverse transcription-polymerase Chain Reaction (RT-PCR) in Vero monolayer cell culture, SARS-CoV-2 CPE showed rounding, elevation, and detachment (Kim, 2020). Virus passage was repeated every five to six days until the cytopathic effect models were stabilized. 1-Cytopathic effect of infected Vero E6 cell in a 15th blind passage on second post passage day (40× magnification); 2- uninfected Vero E6 cell in 40× magnification; 3- PFU evaluation (Nikon TMS inverted microscope, Japan). The presence of SARS-CoV-2 was verified by the appearance of a 398 base pair size band fragment (Roche, U.S.A.). Viroinformatics structural and phylogenetic analysis of Surabaya Isolates B.1.465 revealed changes in the nucleotide and amino acids throughout the gene sequence of Surabaya B.1.465 isolates (Table 1).

Bioinformatics structural and phylogenetic analysis of Surabaya isolates B.1.465

Regarding phylogenetic results, it is proved that Surabaya isolates RSDS-RCVTD-UNAIR-42-A is very close to Pakistan isolate JRCGR-KHI13/2020, and Bangladesh isolate BCSIR-DU-16/2020. The other Surabaya isolates, RSDS-RCVTD-UNAIR-54-A and RSDS-RCVTD-UNAIR-49-A/2021, are close to the Egyptian isolate MASRI-11/2020 and Pakistan's isolate RCGR-KHI25/2020 (Fig. 1). This study revealed changes in the nucleotides and amino acids throughout the Surabaya B.1.465 isolates gene sequence. By aligning the NCBI blast with the reference strain Wuhan/WHO-04. This study identified that various transformations of in the nucleotide and amino acid such as in isolate RSDS-RCVTD-UNAIR-49-A found 2876-C>T, 8621-T>C, 26628-C>T, 28037-C>T, 27988-A>G, 25456-G>T mutations, 22436-C>-, 27393-T>- deletions, 2469-->T insertion and in amino acid found insertion at 553T- 564I>- respectively. Similarly in isolate RSDS-RCVTD-UNAIR-54-A found 566-C> T , 7148-C> G, 7150-C> G, 11851-C> T, 13343-C> T, 16186-C> T, 17359-A>G, 17362-G>A, 23005-G> T, 25586-C> T mutation, 17354--> C, 17366--> C insertion and 2712-C>-, 10100-C>-, 10102-C>-, 10102-C>-, 11563-C>-, 13343-A>-, 25249-T>-, 25252-T> - deletions in nucleotides. Additionally, in RSDS-RCVTD-UNAIR-42-A isolates found mutation at 187-C>T, 950-C>A, 2310-A>C, 2818-C>T, 5616-G>T, 5619-G>T, 5896-A>T, 8316-C>G, 8482-T>C, 18576-C>T, 27446-C>T, insertion at 951-->T, 951-->T, 5893-->C and deletion at 956-T>-, 20985-C>- in nucleotides.

B-cell epitopes prediction, characterization for antigenicity, allergenicity, toxicity, and physiochemistry properties

A total of 71 sequential B-cell epitopes on ORF1ab polyprotein were predicted, antigenicity, allergenicity, and toxicity were used to weed out the epitopes. The epitopes that were chosen were non-toxic, nonallergic, and antigenic. Three of the twelve epitopes were believed to be efficient B-cell epitopes capable of eliciting B lymphocytes (Table 2). The IEDB conservancy analysis method was used to examine B-cell epitopes for the conservancy. During the conservancy analysis, a total of three epitopes, which were reconfirmed through NCBI Protein Blast as ORF1ab polyprotein, were chosen to be exploited in vaccine development. These epitopes were discovered to be highly conserved, with maximum conservation at 100% coverage and identity. Moreover, it also established the physiochemical prediction for the expected peptides, such as the grand average of hydropathicity (GRAVY), aliphatic index, theoretical pI, molecular weight, and instability index (Table 3).

Table 1. Indonesian B.1.465 isolates extracted from the database (GISAID EpiCoV).

Virus Name	Accession ID	Origin	Specimen source	Sequencing technology
RSDS-RCVTD-UNAIR-49-A	EPI-_ISL_1366505	Surabaya, Indonesia	Nasopharyngeal swab	Nanopore GridION
RSDS-RCVTD-UNAIR-54-A	EPI_ISL_1366503	Surabaya, Indonesia	Nasopharyngeal swab	Nanopore GridION
RSDS-RCVTD-UNAIR-42-A	EPI_ISL_1366509	Surabaya, Indonesia	Nasopharyngeal swab	Nanopore GridION

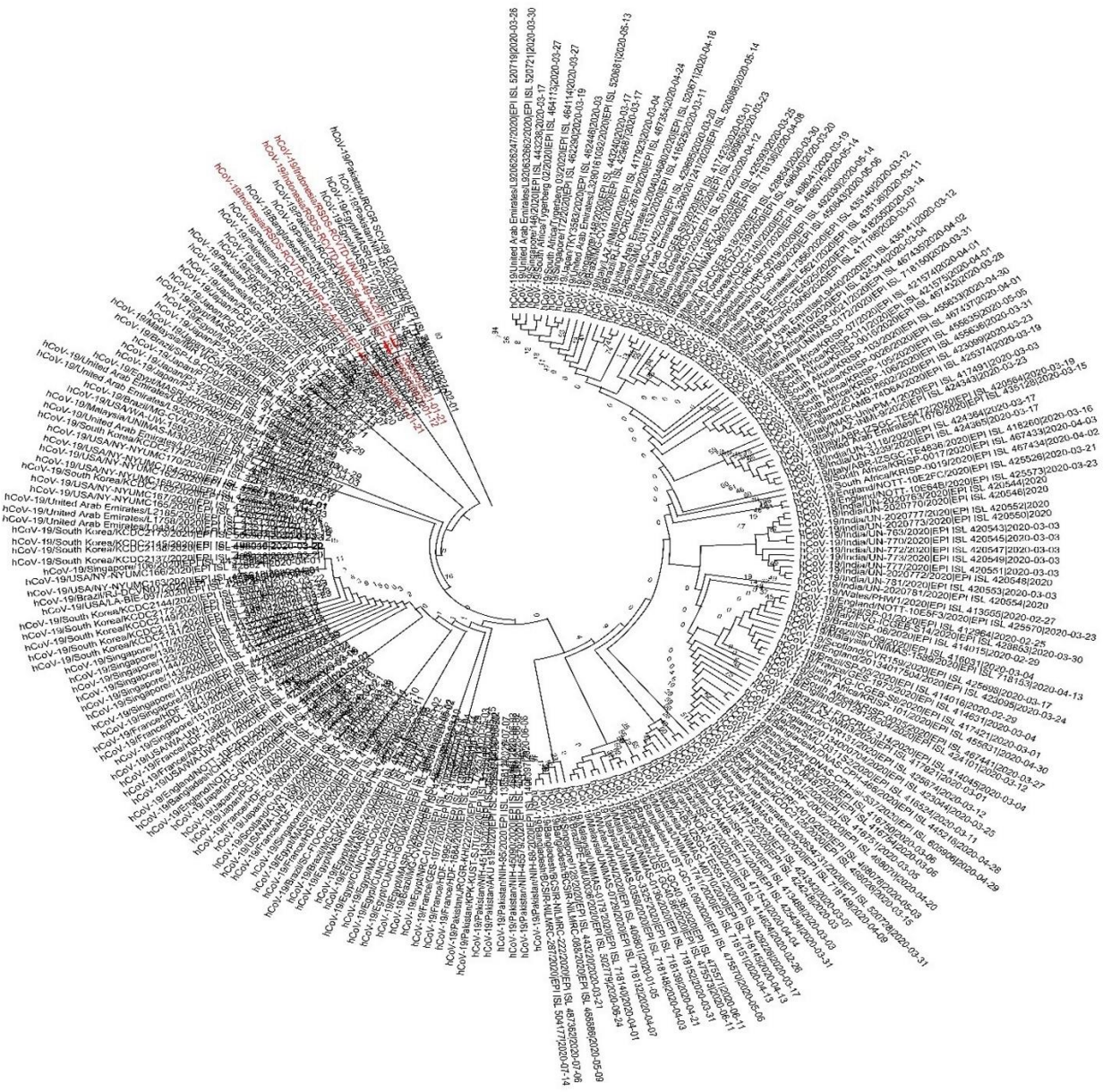

Figure 1. Describing evolutionary relationships of Surabaya Isolates with the world.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree is shown (next to the branches). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. This analysis involved 232 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All ambiguous positions were removed for each pair (pairwise deletion option). There was a total of 29 725 positions in the final dataset. Evolutionary analyses were conducted in MEGA-11 (Kumar et al., 2016).

Linkage development and secondary structure analysis

The predicted ORF1ab polyprotein antigenic, nonallergic, and non-toxic epitopes were linked together by the EAAAK linker and found the sequence of 79 amino acids. Moreover, the SOPMA tool. Pl was used to analyze the secondary structure of construct (Geourjon and Deleage, 1995) and found Alpha helix 36.71%, extended strand 7.59%, random coil 46.84%.

Development of tertiary structure, refinement and validation

The tertiary structure of the ORF1ab polyprotein was developed using the Swiss-Model (Fig. 2A), and the resulting 3D design was refined using Galaxy Refine to minimize error. The Galaxy refinement predicted five refined models. In this study, we chose Model-1 (Fig. 2B) because it had the best scores across all parameters (Rama favored = 100.0, Poor rotamers = 0.0, Clash score = 0.0, MolProbity = 0.500, RMSD =

0.306, GDT-HA = 0.9783). The RAMPAGE server was used to validate the selected model. The rampage predicted that % of residues in our constructed refined structure's favored region indicate the best structure quality (Fig. 2C).

Molecular docking of ORF1ab B-cell epitopic construct with TLR3

The ORF1ab polyprotein was docked with toll-like receptor-3 (TLR3) using the Cluspro.2 webserver for ligand-receptor binding interaction analysis, which yielded ten distinct docking models. We chose model-0 for this study since it has the most cluster members (159) and the lowest energy (-603.4 kcal/mol) (Fig. 3). The PDMsum webserver's analysis of molecular docking results revealed a significant interaction. It showed that the interface area of the construct ORF1ab construct was 872(Å2), while TLR3 had a 749 (Å2). Moreover, the interaction revealed seven salt bridges 127 non-bonded contacts. Because hydrogen bonds are critical for vaccine formation and stability, the design includes 15 hydrogen bonds between chain-9 (vaccine) and chain A (receptor) (Fig. 4).

Immune simulation profile and *in silico* cloning process

The immune simulation was performed using the C-ImmSim server. The immune response was depicted in this method as if it was an actual immunological response. An increase in IgM levels characterized the primary response. Secondary and tertiary responses were characterized by a high B-cell population and found high IgG_1 + IgG_2, IgM, and IgG + IgM levels. Furthermore, the current study revealed cytokine and interleukin production, which depicts the vaccine's efficiency in triggering an immune response. TGF-b, IFN-g, IL-12, IL-4 and IL-2 were also identified in significant concentrations, all of which are vital co-stimulatory for T-cell activation (Fig. 5). The Java Codon Adaptation Tool (JCat) has been used for codon optimization and found an optimized sequence of 237bp with CAI-value = 1.0 and GC-content = 49.78. Subsequently, the pET 28 a(+) expression vector was used to clone the ORF1ab construct between Acc1 and Taq1 restriction sites by using the SnapGene v3.2.1 software and found a colon of 846bp (Fig. 6).

DISCUSSION

Before 2019, six coronavirus strains had been identified as potentially pathogenic to humans. However, the seventh strain of SARS-CoV-2 was discovered in Wuhan, China, in December 2019 (Zhu et al., 2020). About 214.27 million confirmed cases of COVID-19, with 4.27 million deaths, have been reported by WHO since January 2020 (WHO, 2021). Searching for a definitive solution to manage the infection of COVID-19 is an urgent need. Vaccination is an effective way to prevent the spread of Covid-19, but vaccine development is expensive. This burden can be reduced by employing viroinformatic techniques and looking for ways to produce subunit vaccines using viruses' whole genomes and proteomes (Khatoon et al., 2017).

Table 2. B-cell epitopes and other prediction analyses in the Indonesian B.1.465 isolates.

No	Virus Accession No.	Peptide sequence	Length and position	Antigenicity	Allergenicity	Toxicity
1	EPI_ISL_1366505	KNGNKGAGGHSYGADLKSFDLGDELGTDP YEDFQENWNTKHSSGV	45(125-169)	Yes	No	No
		VPGFNEKTH	9(15-13)	Yes	Yes	No
		CHNSEVGPEHSLAEYHNESGLKTILRKG	28(373-400)	Non	Yes	No
		CGETSWQTGDFVKATCEFCGTENLTKEGA TTCG	33(326-358)	Non	Yes	No
2	EPI_ISL_1366503	GVHAGTDLEGNFYGPFVDRQTAQAAGT	27(154-180)	Non	Yes	No
		EDMLNPNYEDLLI	13(31-43)	Yes	Yes	No
		KYNYEPLTQDH	11(220-230)	Yes	Yes	No
		DRDAAMQRK	9(713-721)	Yes	No	No
3	EPI_ISL_1366509	NLHPDSATLVSD	12(70-81)	Yes	Yes	No
		PTDNYITTYPGQGLNGYTVEE	21(129-149)	Non	Yes	No
		AVTAYNGYLTSSSKTPEEH	19(290-308)	Non	Yes	No
		GSYKDWSYSGQSTQL	15(317-331)	Yes	No	No

Table 3. The results of physiochemical prediction in peptides using ProtParam web server.

Virus accession No.	Peptide sequence	Molecular weight	Theorical pI	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
EPI_ISL_1366505	KNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGV	4844.07	4.69	10.13	36.89	-1.236
	VPGFNEKTH	1028.13	6.72	30.29	32.22	-1.089
	CHNSEVGPEHSLAEYHNESGLKTILRKG	3106.42	6.28	64.54	69.64	-0.918
	CGETSWQTGDFVKATCEFCGTENLTKEGATTCG	3475.79	4.25	2.37	26.67	-0.415
EPI_ISL_1366503	GVHAGTDLEGNFYGPFVDRQTAQAAGT	2779.96	4.54	9.47	50.74	-0.396
	EDMLNPNYEDLLI	1578.75	3.43	30.85	120.00	-0.469
	KYNYEPLTQDH	1407.50	5.32	30.40	35.45	-2.018
	DRDAAMQRK	1090.22	8.75	-7.87	22.22	-1.989
EPI_ISL_1366509	NLHPDSATLVSD	1268.35	4.20	16.53	97.50	-0.333
	PTDNYITTYPGQGLNGYTVEE	2332.46	3.57	15.68	50.95	-0.933
	AVTAYNGYLTSSSKTPEEH	2055.19	5.40	35.24	46.32	-0.795
	GSYKDWSYSGQSTQL	1706.79	5.83	33.60	26.00	-1.253

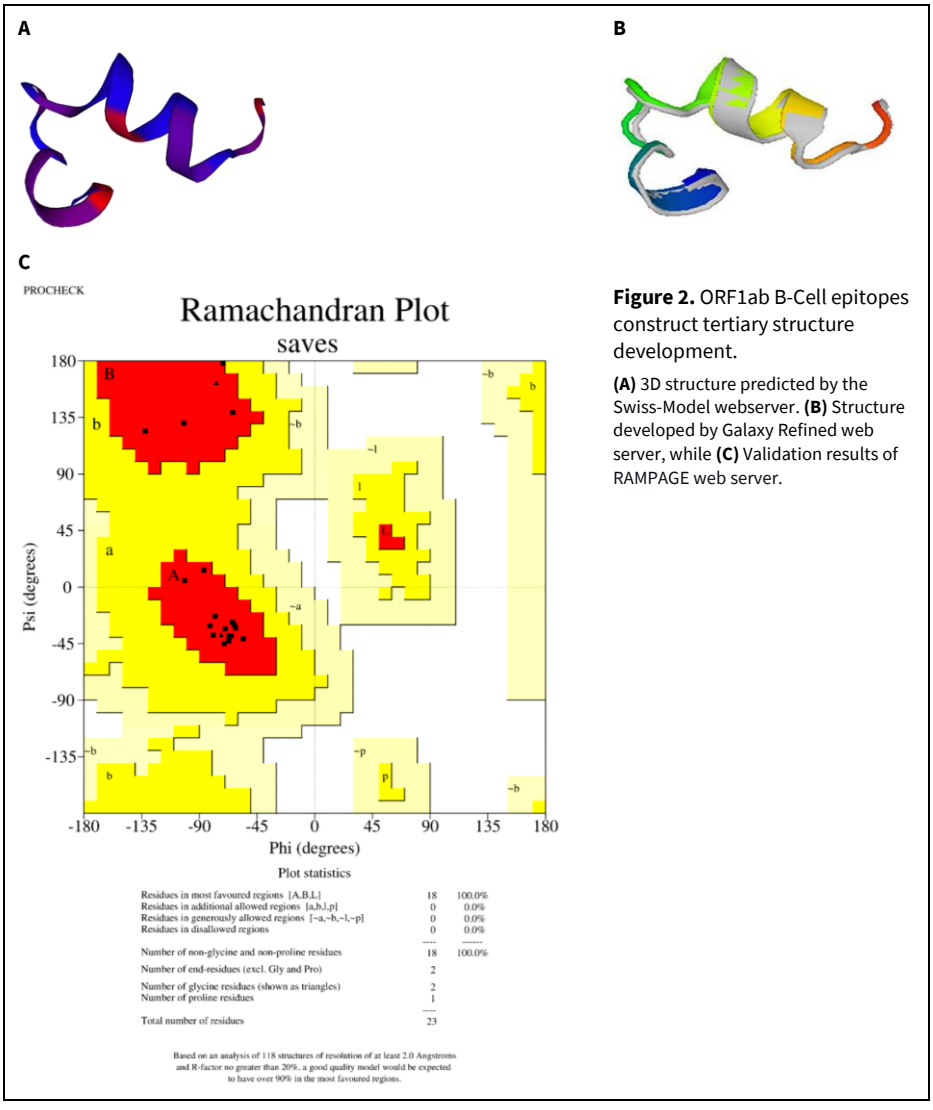

Figure 2. ORF1ab B-Cell epitopes construct tertiary structure development.

(A) 3D structure predicted by the Swiss-Model webserver. **(B)** Structure developed by Galaxy Refined web server, while **(C)** Validation results of RAMPAGE web server.

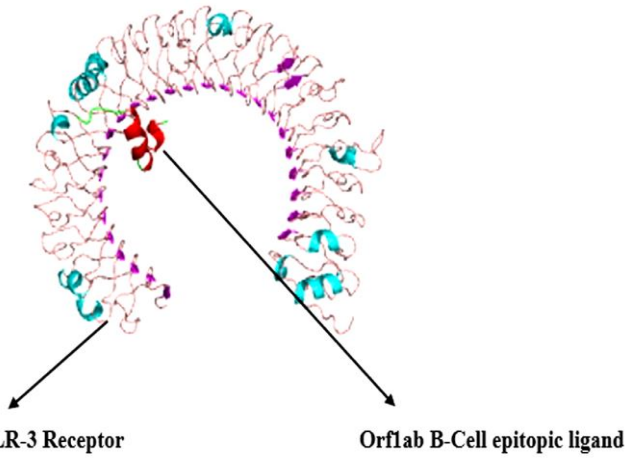

Figure 3. Molecular docking results of ORF1ab B-Cell epitopic construct and TLR-3 receptor developed by Cluspro.2 webserver.

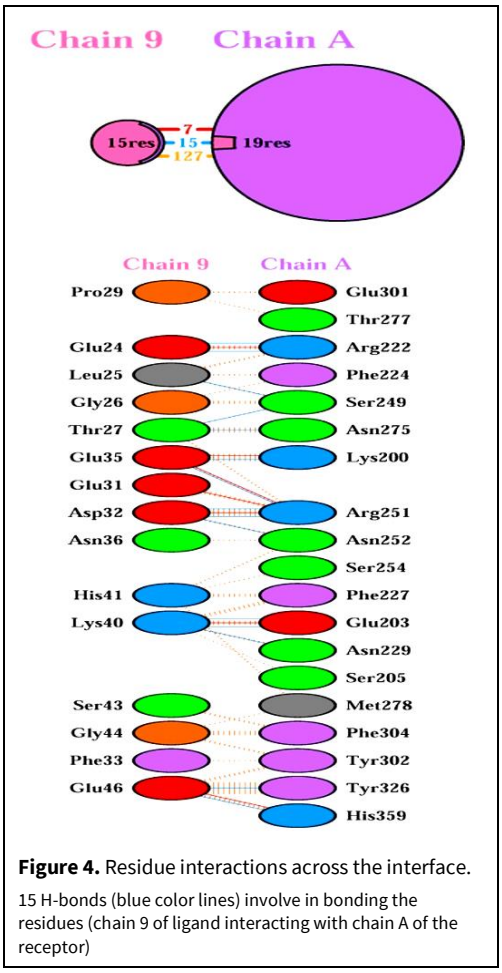

Figure 4. Residue interactions across the interface.

15 H-bonds (blue color lines) involve in bonding the residues (chain 9 of ligand interacting with chain A of the receptor)

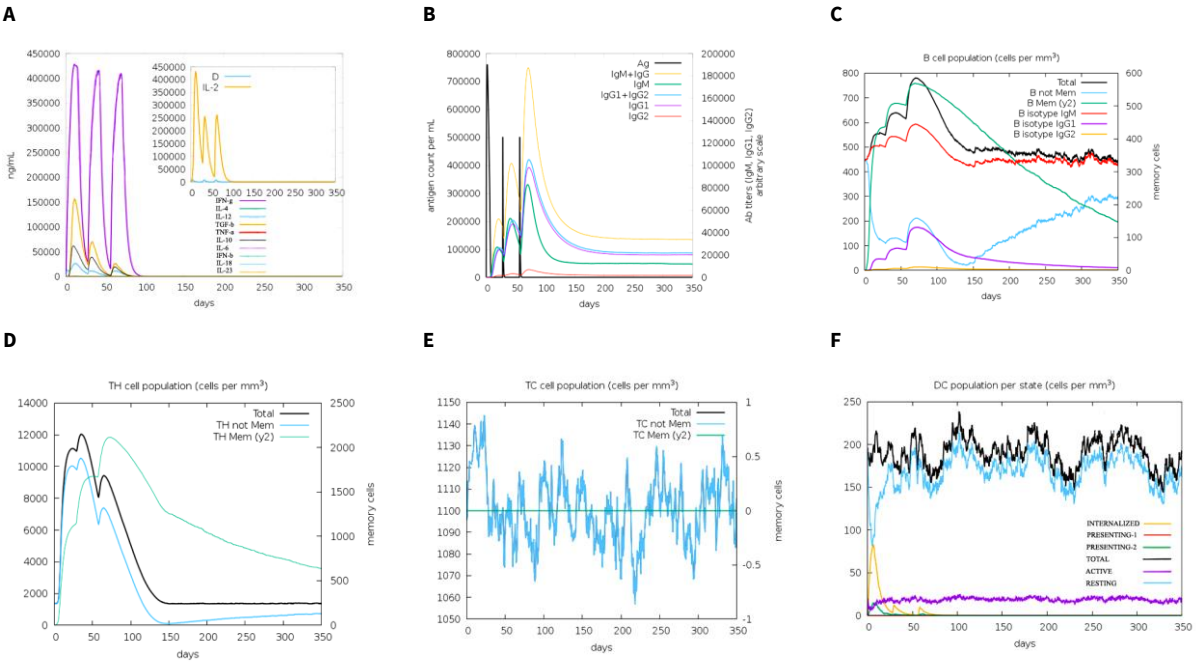

Figure 5. Immunogenic responses of peptides developed through C-Immsim server as **(A)** Concentration of cytokines and interleukins. The inset plot shows the danger signal together with leukocyte growth factor IL-2. **(B)** Virus, immunoglobulins and immunocomplexes. **(C)** B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM, IgG1 and IgG2. **(D)** CD4 T-helper lymphocytes count. The plot shows total memory cell counts. **(E)** CD8 T-cytotoxic lymphocytes count. **(F)** Dendritic cells (DC). DC can present antigenic peptides on both MHC class-I and class-II molecules. Curves show the total number broken down to active, resting, internalized and presenting antigens.

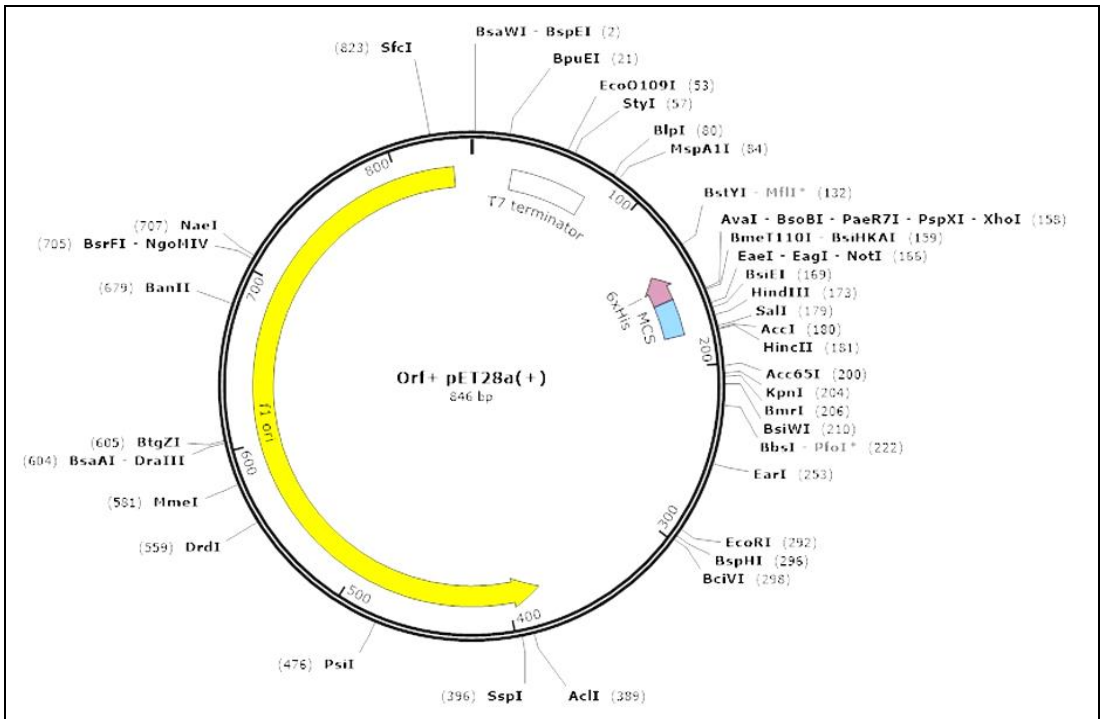

Figure 6. *In silico* cloning's schematic representation of vaccine candidate within pET28 a (+) expression vector.
The total size of the constructed colon is 846 bp, and the gene interest (ORF1ab polyprotein protein).

The current research is based on the three Surabaya B.1.465 isolates generated at RCVTD- ITD of the University of Airlangga Surabaya from the nasopharyngeal swabs of Covid-19 patients. Several studies on the characterization of SARS-CoV-2 and its protein have recently been published (Khailany et al., 2020; Harcourt et al., 2020; Voloch et al., 2021; Watanabe et al., 2020; Weisblum et al., 2020; Li et al., 2020b). SARS-

CoV-2 has various resistance due to a high mutational rate in its genome (Callaway, 2020a); because that, the viral protein of SARS-CoV-2 is the primary focus of our research.

Furthermore, this study also identified that various transformations of in the nucleotide and amino acid such as in isolate RSDS-RCVTD-UNAIR-49-A found 2876-C>T, 8621-T>C, 26628-C>T, 28037-C>T, 27988-A>G, 25456-G>T mutations, 22436-C>-, 27393-T>- deletions, 2469-->T insertion and in amino acid found insertion at 553T- 564I>- respectively. Similarly in isolateRSDS-RCVTD-UNAIR-54-A found 566-C> T , 7148-C> G, 7150-C> G, 11851-C> T, 13343-C> T, 16186-C> T, 17359-A>G, 17362-G>A, 23005-G> T, 25586-C> T mutation, 17354--> C, 17366--> C insertion and 2712-C>-, 10100-C>-, 10102-C>-, 10102-C>-, 11563-C>-, 13343-A>-, 25249-T>-, 25252-T> - deletions in nucleotides. Additionally, in RSDS-RCVTD-UNAIR-42-A isolates found mutation at 187-C>T, 950-C>A, 2310-A>C, 2818-C>T, 5616-G>T, 5619-G>T, 5896-A>T, 8316-C>G, 8482-T>C, 18576-C>T, 27446-C>T, insertion at 951-->T, 951-->T, 5893-->C and deletion at 956-T>-, 20985-C>- in nucleotides.

Because the enzymes that replicate RNA for viruses lack proofreading activity, RNA viruses have exceptionally high mutation rates (Duffy, 2018). Some RNA viruses may have a million times quicker mutation rates than their hosts (Fitzsimmons et al., 2018; Ozono et al., 2021). Virulence modulation and evolvability are linked to a high mutation rate. Factors that influence viral adaptability strike a balance between genetic information integrity and genomic variability (Hu et al., 2020; Korber et al., 2020). More mutations will be discovered as the virus spreads in humans. However, most changes will not affect viral transmission because they do not alter the protein's structure. The accumulation of mutations can determine virus fitness (Callaway, 2020b). According to previous evidence, SARS-CoV-2 propagated swiftly throughout nations, and new mutations had arisen (Callaway, 2020a; Grubaugh et al., 2020).

The phylogenetic analysis explains the evolutionary history and relationships among organisms (Horiike, 2016). The phylogenetic relationship of Surabaya isolates was carried out using Mega 11 software by the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004). According to phylogenetic results, it is proved that Surabaya isolates RSDS-RCVTD-UNAIR-42-A is very close to Pakistan isolate JRCGR-KHI13/2020, and Bangladesh isolate BCSIR-DU-16/2020. The other two Surabaya isolates, RSDS-RCVTD-UNAIR-54-A and RSDS-RCVTD-UNAIR-49-A/2021, are close to the Egyptian isolate MASRI-11/2020 and Pakistans isolate RCGR-KHI25/2020.

The same SARS-CoV-2 lineage B.1.465 circulated India, North America, France, Hongkong, Africa, Spain, Italy, Finland, Germany, UK (GISAD, 2021).

SARS-CoV-2 has various resistance due to a high mutation rate in its genome. Therefore, the protein that covers the virus is the target of this research. This study focused on three structural SARS-CoV-2 protein fusion forms to find the best immunogenic protein to trigger humoral and cellular immune responses.

The most significant barrier to vaccine development is allergenicity (Oany et al., 2014). Some studies were carried out to predict epitopes in different viruses, like in the Zika virus (Adianingsih and Kharisma, 2019). So, in the current study, three amino acid sequences of Surabaya isolates were used to estimate possible B-cell antigenic epitopes. Twelve epitopes with a high antigenic score were chosen for this study. These epitopes were thoroughly examined based on antigenicity, non-toxicity and non-allergenicity. Only three epitopes, "KNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGV", "DRDAAMQRK", "GSYKDWSYSGQSTQL" were determined to meet the already reported criteria for the selection of epitopes for vaccine development (Ansori et al., 2021). The epitopes that are potential to be antigenic, nonallergen and not toxic are immunoreactive for further research (Khan et al., 2014).

In the current study, the selected epitopes' physiochemical properties are in the form of molecular weight, instability index. Theoretical pI, aliphatic index, grand average of hydropathicity (GRAVY) were established (Naveed et al., 2021). They found the strongly thermal stability of peptides in the Instability index regarding the physiochemical properties; the epitopes displayed solid thermal stability. The GRAVY of all epitopes was found to be less than zero. This result proved their hydrophilic properties of them. All epitopes convinced poor heat resistance with Aliphatic index <70, which is essential for SARS-CoV-2 prevention (Gao et al., 2021).

Moreover, these predicted ORF1ab polyproteins were linked by the EAAAK linker to construct a B-cell epitopic sequence. The 3D structure was developed and refining and validation were carried out according to the reported procedure ((Waterhouse et al., 2018; Naveed et al., 2021). The construct has 100% residues in the favorite region because the only 3D structure with >70% of its favored region is considered a quality structure (Chukwudozie et al., 2020).

The molecular docking with the TLR3 has the lowest energy -603.4 kcal/mol, which indicates the strong

binding affinity of protein (Meng et al., 2011). This low energy discovery is consistent with the lowest energy of -602.1 kcal/mol, as Nemati et al. (2021) reported, docking the multi-epitope vaccine with the TRL3 receptor. The current docking revealed 15 hydrogen bonds between ligand and receptor as hydrogen bonds are considered critical for the stability of protein complex (Sen et al., 2021). It is found that when ORF1ab epitopes are added to a multi-epitope vaccine with structural proteins, they demonstrate stronger docking affinity (Safavi et al., 2020). Moreover, the addition of other non-structural proteins like the orf8 protein of SARS-CoV-2's can interact with MHC-I epitopes and causes their downregulation. Furthermore, autophagy specifically targets the ORF8, MHC-epitopes for lysosomal destruction, which results in the SARS-CoV-2-infected cells being substantially more resistant to cytotoxic T lymphocyte lysis. Also, orf8 can damage the antigen presentation mechanism in the body (Zhang et al., 2021). However, ORF1ab epitopes can strongly elicit the T cell epitopes (Safavi et al., 2020).

The immune simulation of ORF1ab epitopes constructs showed a significant increase in Immunoglobulins IgG_1/IgG_2, IgM, and IgG. Therefore, these epitopes are predicted to be a high antibodies producer with a high level of antigenic binding capacity (Castiglione et al., 2012; Siegrist, 2008). B-lymphocytes showed substantial stimulation, consistent with an already published study. That reported vaccines activated immune effectors are critical for making antibodies from B lymphocytes with the ability to bind toxins and pathogens (Cooper and Nemerow, 1984). CD4 T-helper lymphocytes, CD8 T-cytotoxic lymphocytes, dendritic cells, epithelial cells, cytokines. DCS collects antigen in peripheral tissue and moves to drain lymph nodes. T cells effectively serve as a support system for DCS. They are responsible for providing signals (that additional cause activation of B cells and CD8+ cytotoxic T cells (Siegrist, 2008). T-cell vaccination responses are triggered with B-cell responses and subsequent CD4+ Th-cell responses in the lymph node. These ORF1ab epitope findings are consistent with Gangaev et al. (2020b), who found that ORF1ab proteins are more immunogenic than spike and membrane proteins.

The incompatibility of mRNA codons causes alterations in the foreign gene translation efficiency, which is required for higher expression optimization (Pandey et al., 2018). The CAI value of 1.0 and the GC of 49.78% observed in this study are within the specified limits for the optimum limit that can portray increased expression in the *E. coli* K-12 system.

Moreover, the ORF1ab *in silico* illustrated the 846bp gene of interest generated (Fig. 5). The purpose of peptide *in silico* cloning was to alert genetic engineers and scientists to the expression levels and potential cloning sites in specific expression systems, likewise the K12 system of *E. coli* (Tahir ul Qamar et al., 2020).

In this work, we used Surabaya isolates and compared their molecular variation (mutation) to those of the rest of the world. So that SARS-CoV-2 mutational trends can be investigated. Furthermore, we used a bioinformatics approach to assess the immunogenic characteristics of ORF1ab isotopes, which demonstrated high immunological responses when evaluated computationally against SARS-CoV-2. The suggested options elicited humoral and cell-mediated immune responses according to the required standards. In addition, the proposed peptide was stable, and its receptor binding was found higher. Thus, the immune simulation's results are virtually likely immune in real life. The applied bioinformatics approach gives raw data for *in vivo* and *in vitro* studies to determine the true potential of this COVID-19-fighting strategy.

CONCLUSION

According to this study, the SARS-CoV-2 virus is constantly mutating. These mutations are the most significant barrier to developing an effective vaccine to combat the COVID-19 infection. Moreover, using a bioinformatics approach, the predicted, antigenic, nonallergic, non-toxic B-cell epitopes (KNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTK HSSGV, DRDAAMQRK, and GSYKDWSYSGQSTQL) conserved in ORF1ab polyproteins: elicited strong humoral and cellular immune responses. These findings suggested that incorporating ORF1ab polyprotein B-cell epitopes into developing subunit vaccines against SARS-CoV-2 might result in significant immunological responses. Furthermore, for the sake of humanity, these findings should be tested in the laboratory and in the field to assess actual immunogenic responses.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

This work was supported by the Research Centre for Vaccine Technology and Development, Institute of Tropical Disease, Airlangga University, Surabaya, Indonesia. The authors are grateful to the Research Centre for Vaccine Technology and Development team, Institute of Tropical Disease, Airlangga University, Surabaya, Indonesia, for their technical assistance.

REFERENCES

Abraham Peele K, Srihansa T, Krupanidhi S, Ayyagari V S, Venkateswarulu TC (2021) Design of multi-epitope vaccine candidate against SARS-CoV-2: A *in-silico* study. J Biomol Struct Dyn 39: 3793–3801.

Adianingsih OR, Kharisma VD (2019) Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. J App Pharm Sci 9: 098–103.

Ansori AN, Kusala MK, Normalina I, Indrasari S, Alamudi MY, Nidom RV, Nidom CA (2020) Immunoinformatic investigation of three structural protein genes in Indonesian SARS-CoV-2 isolates. Sys Rev Pharm 11: 422–434.

Ansori AN, Nidom RV, Kusala MK, Indrasari S, Normalina I, Nidom AN, Nidom CA (2021) Viroinformatics investigation of B-cell epitope conserved region in SARS-CoV-2 lineage B. 1.1. 7 isolates originated from Indonesia to develop vaccine candidate against COVID-19. J Pharm Pharmacogn Res 9: 766–779.

Biswas A, Bhattacharjee U, Chakrabarti AK, Tewari DN, Banu H, Dutta S (2020) Emergence of novel coronavirus and COVID-19: Whether to stay or die out? Crit Rev Microbiol 46: 182–193.

Bond CW, Leibowitz JL, Robb JA (1979) Pathogenic murine coronaviruses II. Characterization of virus-specific proteins of murine coronaviruses JHMV and A59V. Virology 94: 371–384.

Cai J, Sun W, Huang J, Gamber M, Wu J, He G (2020) Indirect virus transmission in cluster of COVID-19 cases, Wenzhou, China, 2020. Emerg Infect Dis 26: 1343–1345.

Callaway E (2020a) Making sense of coronavirus mutations. Nature 585: 174–177.

Callaway E (2020b) The coronavirus is mutating--does it matter? Nature 585: 174–178.

Castiglione F, Mantile F, De Berardinis P, Prisco A (2012) How the interval between prime and boost injection affects the immune response in a computational model of the immune system. Comput Math Methods Med 2012: 842329.

Cevik M, Kuppalli K, Kindrachuk J, Peiris M (2020) Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ 371: m3862.

Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY (2020a) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microb Infect 9(1): 221-236.

Chan JFW, Zhang AJ, Yuan S, Poon VKM, Chan CCS, Lee ACY, Yuen KY (2020b) Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in a golden Syrian hamster model: implications for disease pathogenesis and transmissibility. Clin Infect Dis 71: 2428–2446.

Chukwudozie OS, Chukwuanukwu RC, Iroanya OO, Eze DM, Duru VC, Dele-Alimi TO, Okinedo EU (2020) Attenuated subcomponent vaccine design targeting the SARS-CoV-2 nucleocapsid phosphoprotein RNA binding domain: *In silico* analysis. J Immunol Res 2020: 2837670.

Cooper NR, Nemerow GR (1984) The role of antibody and complement in the control of viral infections. J Investig Dermatol 83: S121–S127.

Dandekar AA, Perlman S (2005) Immunopathogenesis of coronavirus infections: Implications for SARS. Nat Rev Immunol 5: 917–927.

Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Peters B (2019) IEDB-AR: immune epitope database—analysis resource in 2019. Nucleic Acids Res 47: W502–W506.

Dong L, Tian J, He S, Zhu C, Wang J, Liu C, Yang J (2020) Possible vertical transmission of SARS-CoV-2 from an infected mother to her newborn. JAMA 323: 1846-1848.

Duffy S (2018) Why are RNA virus mutation rates so damn high? PLoS Biol 16: e3000003.

Emameh RZ, Nosrati H, Taheri RA (2020) Combination of biodata mining and computational modelling in identification and characterization of ORF1ab polyprotein of SARS-CoV-2 isolated from oronasopharynx of an Iranian patient. Biol Proced Online 22: 8.

Falzone L, Musso N, Gattuso G, Bongiorno D, Palermo CI, Scalia G, Stefani S (2020) Sensitivity assessment of droplet digital PCR for SARS-CoV-2 detection. Int J Mol Med 46: 957–964.

Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, Lauring AS (2018) A speed–fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS Biology 16: e2006459.

Fritz M, Rosolen B, Krafft E, Becquart P, Elguero E, Vratskikh O, Denolly S, Boson B, Vanhomwegen J, Gouilh MA, Kodjo A, Chirouze C, Rosolen SG, Legros V, Leroy EM (2020) High prevalence of SARS-CoV-2 antibodies in pets from COVID-19+ households. One Health 11: 100192.

Gangaev A, Ketelaars SL, Isaeva OI, Patiwael S, Dopler A, Hoefakker K, De Biasi S, Gibellini L, Mussini C, Guaraldi G, Girardis M (2020a) Identification and characterization of an immunodominant SARS-CoV-2-specific CD8 T cell response. Res Sq [Preprint] DOI: 10.21203/rs.3.rs-33197/v2

Gangaev A, Ketelaars SL, Patiwael S, Dopler A, Isaeva OI, Hoefakker K, Kvistborg P (2020b) Profound CD8 T-cell responses towards SARS-CoV-2 OFR1ab in COVID-19 patients. Res Sq [Preprint] DOI: 10.21203/rs.3.rs-33197/v1

Gao T, Gao Y, Liu X, Nie Z, Sun H, Lin K, Wang S (2021) Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. BMC Microbiol 21: 58.

Geourjon C, Deleage G (1995) SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 11: 681–684.

Ghoshal B, Ghoshal B, Swift S, Tucker A (2021) Uncertainty estimation in SARS-CoV-2 B-cell epitope prediction for vaccine development. In: International Conference on

Artificial Intelligence in Medicine. Springer, Cham, p. 361–366.

GISAD (2021) GISAID Database. Available from https://www.epicov.org/epi3/frontend#6d403 [Accessed July 27, 2021].

Graham RL, Sparks JS., Eckerle LD, Sims AC, Denison MR (2008) SARS coronavirus replicase proteins in pathogenesis. Virus Res 133: 88–100.

Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33: W526–W531.

Grubaugh ND, Hanage WP, Rasmussen AL (2020) Making sense of mutation: What D614G means for the COVID-19 pandemic remains unclear. Cell 182: 794–795.

Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, Raghava GP (2013) *In silico* approach for predicting toxicity of peptides and proteins. PloS One 8: e73957.

Hachim A, Kavian N, Cohen CA, Chin AW, Chu DK, Mok CK, Tsang OT, Yeung YC, Perera RA, Poon LL, Peiris JM (2020) ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. Nat Immunol 21: 1293–1301.

Han H, Ma Q, Li C, Liu R, Zhao L, Wang W, Xia Y (2020) Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. Emerg Microbes Infect 9: 1123–1130.

Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, Thornburg N J (2020) Isolation and characterization of SARS-CoV-2 from the first US COVID-19 patient. BioRxiv [Preprint] DOI: 10.1101/2020.03.02.972935

Hernández-Huerta MT, Pérez-Campos Mayoral L, Romero Díaz C, Martínez Cruz M, Mayoral-Andrade G, Sanchez Navarro LM, Matias-Cervantes CA (2021) Analysis of SARS-CoV-2 mutations in Mexico, Belize, and isolated regions of Guatemala and its implication in the diagnosis. Med Virol 93: 2099–2114.

Horiike T (2016) An introduction to molecular phylogenetic analysis. Rev Agri Sci 4: 36-45.

Hu J, He CL, Gao Q, Zhang GJ, Cao XX, Long QX, Huang AL (2020) The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. BioRxiv [Preprint] DOI: 10.1101/2020.06.20.161323

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cao B (2020a) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395: 497–506.

Huang Y, Yang C, Xu XF, Xu W, Liu SW (2020b) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin 41: 1141–1149.

Kathwate GH (2022) *In silico* design and characterization of multi-epitopes vaccine for SARS-CoV2 from its spike proteins. Int J Pept Res Ther 28: 37.

Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. Gene Reports 19: 100682.

Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, Seraj ZI (2014) *In silico* predicted mycobacterial epitope elicits *in vitro* T-cell responses. Mo.l Immunol 61: 16–22.

Khatoon N, Pandey RK, Prajapati VK (2017) Exploring leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Sci Rep 7: 8285.

Kim JI (2020) New potential for healing the trauma of Maori from Brain Education. [Interview] Dr. Lily George, Director of Education, New Zealand Headquarters of ECO. IBREA Report 12: 3–7.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Montefiori DC (2020) Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. Cell 182: 812–827.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870-1874.

Lau SY, Wang P, Mok BWY, Zhang AJ, Chu H, Lee ACY, Chen H (2020) Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect 9: 837–842.

Lewis DM, Leibrand S, Leibrand H (2020) A test-based strategy for safely shortening quarantine for COVID-19. MedRxiv [Preprint] DOI: 10.1101/2020.11.24.20238287

Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Wu J (2020a) Coronavirus infections and immune responses. J Med Virol 92: 424–432.

Li J, Wang H, Mao L, Yu H, Yu X, Sun Z, Wang X (2020b) Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. Sci Rep 10: 17492.

Liu K, Pan X, Li L, Yu F, Zheng A, Du P, Han P, Meng Y, Zhang Y, Wu L, Chen Q (2021) Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. Cell 184(13): 3438–3451.

Mapleson D, Drou N, Swarbreck D (2015) RAMPART: A workflow management system for de novo genome assembly. Bioinformatics 31: 1824–1826.

Martin A, Nateqi J, Gruarin S, Munsch N, Abdarahmane I, Zobel M, Knapp B (2020) An artificial intelligence-based first-line defence against COVID-19: Digitally screening citizens for risks via a chatbot. Sci Rep 10: 19012.

Meng XY, Zhang HX, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7: 146–157.

Naveed M, Tehreem S, Arshad S, Bukhari SA, Shabbir MA, Essa R, Khan I (2021) Design of a novel multiple epitope-based vaccine: An immunoinformatics approach to combat SARS-CoV-2 strains. J Infect Public Health 14: 938–946.

Nemati AS, Tafrihi M, Sheikhi F, Tabari AR, Haditabar A (2021) Designing a new multi epitope-based vaccine against COVID-19 disease: An immunoinformatic study based on reverse vaccinology approach. Res Sq [Preprint] DOI: 10.21203/rs.3.rs-206270/v1

Oany AR, Emran AA, Jyoti TP (2014) Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an *in silico* approach. Drug Des Devel Ther 8: 1139.

Ozono S, Zhang Y, Ode H, Seng TT, Imai K, Miyoshi K, Tokunaga K (2021) Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry. Nat Commun 12: 848.

Pandey RK, Bhatt TK, Prajapati VK (2018) Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci Rep 8: 1125.

Park WB, Kwon NJ, Choi SJ, Kang CK, Choe PG, Kim JY, Oh MD (2020) Virus isolation from the first patient with SARS-CoV-2 in Korea. J Korean Med Sci 35: 10–14.

Rapin N, Lund O, Bernaschi M, Castiglione F (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PloS One 5: e9862.

Safavi A, Kefayat A, Mahdevar E, Abiri A, Ghahremani F (2020) Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches. Vaccine 38: 7612–7628.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.

Sen Gupta PS, Biswal S, Singha D, Rana MK (2021) Binding insight of clinically oriented drug famotidine with the identified potential target of SARS-CoV-2. J Biomol Struct Dyn 39: 5327–5333.

Siegrist CA (2008) Vaccine Immunology. Vaccines. Saunders Elsevier, p. 17–36.

Singh A, Thakur M, Sharma LK, Chandra K (2020) Designing a multi-epitope peptide based vaccine against SARS-CoV-2. Sci Rep 10: 16219.

Tahir ul Qamar M, Shahid F, Aslam S, Ashfaq UA, Aslam S, Fatima I, Chen LL (2020) Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against SARS-CoV-2. Infect Dis Poverty 9: 132.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101: 11030–11035.

Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S (2020) Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antivir Res 178: 104792.

Valencia I, Peiró C, Lorenzo Ó, Sánchez-Ferrer CF, Eckel J, Romacho T (2020) DPP4 and ACE2 in diabetes and COVID-19: therapeutic targets for cardiovascular complications? Front Pharmacol 11: 1161.

Voloch CM, da Silva Francisco Jr R, de Almeida LG, Cardoso CC, Brustolini OJ, Gerber AL, de Vasconcelos ATR (2021) Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. Virol J 95: e00119-21.

Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. Science 369: 330–333.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46: W296–W303.

Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Bieniasz PD (2020) Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife 9: e61312.

Weiss SR, Navas-Martin S (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol Mol Biol Rev 69: 635–664.

WHO (2020a) World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technicalguidance/naming-the-coronavirus-disease-(covid-2019)-and-thevirus-that-causes-it. [Accessed 2 March, 2020].

WHO (2020b) World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 32. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200221-sitrep-32-covid-19.pdf [Accessed January 05, 2022].

WHO (2021) World Health Organization. Coronavirus disease. Available from https://covid19.who.int/ [Accessed August 27, 2021].

Yadav V, Rajput M, Diwakar RP, Kumar R (2020) An Overview on transmission of diseases in special reference to COVID-19 and potential Tar-gets to control this pandemic. J Adv Microbiol Res 4: 015.

Zhang J, Zeng H, Gu J, Li H, Zheng L, Zou Q (2020) Progress and prospects on vaccine development against SARS-CoV-2. Vaccines 8: 153.

Zhang Y, Chen Y, Li Y, Huang F, Luo B, Yuan Y, Xia B, Ma X, Yang T, Yu F, Liu J (2021) The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. Proc Natl Acad Sci U S A 118: e2024202118.

Zheng YY, Ma YT, Zhang JY, Xie X (2020) COVID-19 and the cardiovascular system. Nat Rev Cardiol 17: 259–260.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382: 727–733.

AUTHOR CONTRIBUTION:

Contribution	Shehzad A	Kuncorojakti S	Tacharina MR	Ahmad HI	A'la R	Wijaya AY	Tyasningsih W	Rantam FA
Concepts or ideas	x	x	x	x	x	x	x	x
Design	x	x	x	x	x	x	x	x
Definition of intellectual content	x	x	x	x	x	x	x	x
Literature search	x	x	x			x		x
Experimental studies	x	x	x	x	x	x	x	x
Data acquisition	x	x	x	x	x	x	x	x
Data analysis	x	x	x	x	x	x	x	x
Statistical analysis	x	x	x	x	x	x	x	x
Manuscript preparation	x	x	x	x	x	x	x	x
Manuscript editing	x	x	x					x
Manuscript review	x	x	x	x	x	x	x	x

Citation Format: Shehzad A, Kuncorojakti S, Tacharina MR, Ahmad HI, A'la R, Wijaya AY, Tyasningsih W, Rantam FA (2022) Molecular characterization and prediction of B-cell epitopes for the development of SARS-CoV-2 vaccine through bioinformatics approach. J Pharm Pharmacogn Res 10(3): 429–444. https://doi.org/10.56499/jppres21.1262_10.3.429