Coronavirus is a large family of viruses that cause mild to moderate upper respiratory infections. However, some types of coronavirus can also cause more serious illnesses, such as Middle East respiratory syndrome coronavirus (MERS-CoV), severe acute respiratory syndrome coronavirus (SARS-CoV), and coronavirus disease 2019 (COVID-19) [1]. Up to now, seven coronaviruses (HCoVs) have been identified, namely HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, MERS-CoV, and COVID-19. COVID-19 is a member of the coronavirus family, which by the early of May 2020 has infected more than 3.5 million people and caused almost 250,000 deaths worldwide. The spread of COVID-19 is expanding globally within less than 3 months and causing many losses in various sectors [1]. Severe acute respiratory syndrome (SARS) is an acute respiratory disorder caused by a coronavirus (SARS-CoV). During the global outbreak in 2002/2003, this catastrophic disease resulted in 8,400 cases and 900 deaths according to a report by the World Health Organization [2]. MERS-CoV is an emerging virus that is involved in cases of acute respiratory infections in the Arabian Peninsula, Tunisia, Morocco, France, Italy, Germany, and England. The novel corona-
Virus, which has been contagious in Saudi Arabia since March 2012, has never before been found in the world and has characteristics that are different from the SARS coronavirus that infected 32 countries in the world in 2003 [3].

All types of coronaviruses cause clinical symptoms that can include fever, coughing, acute respiratory distress, pneumonia, fatigue, headaches, dyspnea, lymphopenia, and infrequently cause gastrointestinal symptoms such as diarrhea. Severe COVID-19 infection can be characterized by turbidity in both lung subpleural areas, acute respiratory distress syndrome, and acute cardiac injury. In critical patients occur both local and systemic immune responses, which lead to intense inflammation [1,4].

Vaccination is still the most effective preventive for virus infection. One of the latest vaccine technology developments are peptide-based vaccines or epitope vaccines. Epitope based vaccine is synthesized based on in silico analyzes through an immunoinformatics approach. In silico studies reduce costs and time needed in developing vaccines and construct vaccines with higher efficacy and safety than conventional vaccines [5-7].

Looking at the global pandemic COVID-19, MERS, and SARS caused by coronavirus, it is considered necessary to develop an effective vaccine against all types of coronavirus. Alignment of nine strains of the coronavirus has now been carried out and a highly conserved region of the S2 spike protein has been found. Highly conserved regions can be potential vaccine candidates because they can recognize various strains of the coronavirus.

Spike protein is a surface protein in coronavirus that plays a role in binding with receptors and facilitating membrane fusion. The spike S1 protein plays a role in binding virions to the cell membrane through its interaction with the receptors so that it initiates the infection process. S2 protein facilitates fusion between virions and cell membranes [8,9].

Materials and Methods

Data collection

Vaccine technology developments are peptide-based vaccines or epitope vaccines. Epitope based vaccine is synthesized based on in silico analyzes through an immunoinformatics approach. In silico studies reduce costs and time needed in developing vaccines and construct vaccines with higher efficacy and safety than conventional vaccines [5-7].

Looking at the global pandemic COVID-19, MERS, and SARS caused by coronavirus, it is considered necessary to develop an effective vaccine against all types of coronavirus. Alignment of nine strains of the coronavirus has now been carried out and a highly conserved region of the S2 spike protein has been found. Highly conserved regions can be potential vaccine candidates because they can recognize various strains of the coronavirus.

Spike protein is a surface protein in coronavirus that plays a role in binding with receptors and facilitating membrane fusion. The spike S1 protein plays a role in binding virions to the cell membrane through its interaction with the receptors so that it initiates the infection process. S2 protein facilitates fusion between virions and cell membranes [8,9].

Materials and Methods

Table 1. Coronavirus ID number

| No. | ID  | Organism name                                      |
|-----|-----|----------------------------------------------------|
| 1.  | Q6Q1S2 | Human coronavirus NL63 (HCoV-NL63)                 |
| 2.  | P36334 | Human coronavirus OC43 (HCoV-OC43)                 |
| 3.  | P0DTC2 | Severe acute respiratory syndrome coronavirus 2 (2019-nCoV, SARS-CoV-2) |
| 4.  | P59594 | Human SARS coronavirus (severe acute respiratory syndrome coronavirus, SARS-CoV) |
| 5.  | P15423 | Human coronavirus 229E (HCoV-229E)                 |
| 6.  | K9N5Q8 | Middle East respiratory syndrome-related coronavirus |
| 7.  | Q0ZME7 | Human coronavirus HKU1 (isolate N5) (HCoV-HKU1)   |
| 8.  | Q5MQD0 | Human coronavirus HKU1 (isolate N1) (HCoV-HKU1)   |
| 9.  | Q14EB0 | Human coronavirus HKU1 (isolate N2) (HCoV-HKU1)   |

Fig. 1. Highly conserve region form coronavirus spike protein, amino acid number 945–1,100 was used to predict epitopes.
collected from protein data bank (https://www.uniprot.org/) (Table 1).

Alignment and epitopes prediction
Nine spike protein sequences were aligned using COBALT (constraint-based multiple alignment tools) which is available at https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi.

Highly conservatives’ sequences were chosen and analyzed its B cells epitope using several tools (Emini Surface Accessibility Prediction, Chou and Fasman Beta-turn Prediction, Parker hydrophilicity prediction, Kolaskar Tongaonkar Antigenicity for linear epitopes) and DiscoTope for continuous epitopes. While the T cell epitopes were predicted using NetCTL, Immune Epitope Database (IEDB)-major histocom-

Table 2. T cells epitopes prediction result

| No. | Start | Stop  | Peptide | Methods       | MHC class | Prediction based on | Tool     |
|-----|-------|-------|---------|---------------|-----------|---------------------|----------|
| 1.  | 1038  | 1046  | RVDFCGKGY | ANN           | I         | √                   | NetCTL   |
| 2.  | 1016  | 1024  | AEIRASANL | ANN           | I         | √                   | IEDB-MHC I |
| 3.  | 1015  | 1029  | AAERASANLAKTM | ANN   | II        | √                   | IEDB-MHC II |
| 4.  | 1038  | 1046  | RVDFCGKGY | SM            | I and II  | √                   | MotifScan |
| 5.  | 1041  | 1050  | FCGKGYHLM | QSAR          | I and II  | √                   | MHCPred   |

MHC, major histocompatibility complex; A, quantitative binding affinity; S, supertypes; T, TAP binding; P, proteasomal cleavage; ANN, artificial neural network; IEDB, Immune Epitope Database; SM, sequence motif; QSAR, quantitative structure-activity relationship model.

Table 3. Linear B cells epitope predicted from highly conserved region

| No. | Start | Stop  | Peptide | Tool              | Method               |
|-----|-------|-------|---------|-------------------|----------------------|
| 1.  | 987   | 996   | EADEVSDL | Bepiprep         | Machine learning-   |
| 1034| 1045  |       | LGQSKRDFCGK |                 | decision trees       |
| 2.  | 1067  | 1075  | VPAOENFT  | Emini Surface Accessibility Prediction | Propensity scale |
| 1085| 1090  |       | GKAHFP   |                    |                      |
| 3.  | 1053  | 1059  | POSAPHIG | Chou and Fasman Beta-turn Prediction | Propensity scale |
| 4.  | 960   | 966   | NTLVKQL | Kolaskar Tongaonkar Antigenicity | Propensity scale |
| 972 | 978   |       | ISSVLND  |                    |                      |
| 1004| 1011  |       | LOTTYTQG |                    |                      |
| 1079| 1084  |       | PAICHD   |                    |                      |
| 5.  | 1071  | 1147  | OKEFNTTAICDQGKAFREGVYSNSGINTMITQGWFYQOEFTII  | Ellipro |
|     |       |       | DNTFVSIGNCDVVGIGVINNTYDPLQGELDS |                      |

MHC, major histocompatibility complex; A, quantitative binding affinity; S, supertypes; T, TAP binding; P, proteasomal cleavage; ANN, artificial neural network; IEDB, Immune Epitope Database; SM, sequence motif; QSAR, quantitative structure-activity relationship model.

Fig. 2. Continuous B cells epitope predicted from highly conserved region and its residues (180 residues).
patibility complex (MHC) I, IEDB-MHC II, and MotifScan tools.

Results

Highly conserved region from coronavirus spike protein
Spike protein sequences from nine strains of coronavirus which infected human were collected. Alignment result showed a highly conserved region in amino acid number 945-1100 from severe acute respiratory syndrome coronavirus 2 (2019-nCoV, SARS-CoV-2) spike protein (Fig. 1). This region was used to predict the T and B cells epitopes.

T cells and B epitopes
Several tools to predict T cells epitopes identified epitopes that presented by MHC class I and II (Table 2). While, the B cells linear epitopes prediction was presented in Table 3, the continuous B cells epitopes is demonstrated in Fig. 2. In summary, all of the epitopes identified in highly conserved region is revealed in Fig. 3.

Discussion

Vaccination is one of the most effective approaches to prevent viral infections. However, the development of vaccines requires a long time and high costs since it is required for the screening of large arrays of potential epitope candidates. Using the in-silico predictions method, it can dramatically reduce the cost for vaccine development. The immune system recognizes antigens through the mechanism of humoral and cellular immune systems, each of which is mediated by B cells and T cells. Both types of immune cells recognize the antigen not as a whole but only in a portion of the pathogenic components called antigens. The introduction of B cell antigens and T cells requires a different process [10].

We predict epitopes from spike glycoprotein (S protein) since this protein has been studied as the most antigenic part of the virus [11]. Prior to epitope prediction, sequencing of S protein sequences of nine strains of the coronavirus was carried out. From this alignment, it is obtained that the highly conserved region is from amino acid residue number 945–1100.

From the highly conserved region, epitope prediction is carried out; both B cell epitope and T cell epitope. Epitope prediction is performed in the highly conserved area with the intention that the vaccine can be used for a variety of coronavirus strains, including it is expected that if a new type of virus strain develops in the future, the area this is conserved and vaccination remains effective. Our findings provide a sequence from highly conserved region of S2 protein which can help guide new experimental efforts to develop coronavirus vaccine candidate.

B cell epitope prediction is performed to predict both linear and continuous epitopes. From the prediction of linear epitopes in the highly conserved region it was found that the area contained several potential epitopes. Prediction of continuous epitopes has similar results with the presence of epitopes that is recognized by B cells in the spike protein. T cell epitopes prediction in highly conserved region also has similar results. The conclusion of these predictions is the presence of epitopes in the highly conserved region so that they can be developed as vaccine candidates.

The results of this study can be a reference for the next stage of coronavirus vaccine development. A delivery strategy that can be useful in the development of the coronavirus vaccine is by the mucosal pathway using live bacteria vector as a career. Live bacteria become an important career because they can induce the mucosal immune system in addition to the systemic immune system [12], the mucosal immune system is very important to defense against viral infections that attack the respiratory tract.
References

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109:102433.
2. Xu RH, He JF, Evans MR, et al. Epidemiologic clues to SARS origin in China. Emerg Infect Dis 2004;10:1030-7.
3. Chen X, Chughtai AA, Dyda A, MacIntyre CR. Comparative epidemiology of Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia and South Korea. Emerg Microbes Infect 2017;6:e51.
4. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med 2020;8:420-2.
5. Correia BE, Bates JT, Loomis RJ, et al. Proof of principle for epitope-focused vaccine design. Nature 2014;507:201-6.
6. El-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. Immunome Res 2010;6 Suppl 2(Suppl 2):S2.
7. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chem Biol Interact 2008;171:165-76.
8. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 2020;12:254.
9. Wu Y. Strong evolutionary convergence of receptor-binding protein spike between COVID-19 and SARS-related coronaviruses [Internet]. Huntington, NY: bioRxiv, Cold Spring Harbor Laboratory; 2020 [cited 2020 Jun 2]. Available from: https://doi.org/10.1101/2020.03.04.975995.
10. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T- and B-cell epitope prediction. J Immunol Res 2017;2017:2680160.
11. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. Candidate targets for immune responses to 2019-novel coronavirus (nCoV): sequence homology- and bioinformatic-based predictions [Internet]. Rochester, NY: SSRN; 2020 [cited 2020 Jun 2]. Available from: http://dx.doi.org/10.2139/ssrn.3541361.
12. Yurina V. Live bacterial vectors: a promising DNA vaccine delivery system. Med Sci (Basel) 2018;6:27.