HAYATI
Journal of Biosciences

Vol. 30 No. 3, May 2023 445-450
DOI:10.4308/hjb.30.3.445-450
ISSN: 1978-3019
EISSN: 2086-4094

Expression of SARS-CoV-2 Nucleocapsid (N) Recombinant Protein Using *Escherichia coli* System

Rizki Aulia Ansari[1], Uus Saepuloh[2], Silmi Mariya[2], Yuliana[2], Rachmitasari Noviana[2], Irma Herawati Suparto[2,3], Huda Shalahudin Darusman[1,2,4*]

[1]*Graduate School Program of Biotechnology, School of Post Graduated Programs, IPB University, Bogor 16680, Indonesia*
[2]*Primate Research Center, IPB University, Bogor 16151, Indonesia*
[3]*Department of Chemistry, IPB University, Bogor 16680, Indonesia*
[4]*School of Veterinary Medicine and Biomedical (SKHB), IPB University, Bogor 16680, Indonesia*

ARTICLE INFO

Article history:
Received March 26, 2022
Received in revised form April 11, 2022
Accepted October 28, 2022

KEYWORDS:
denaturation condition,
E. coli BL21 (DE3),
nucleocapsid,
recombinant protein,
SARS-CoV-2

ABSTRACT

One of the main antigen that can be used for serological testing is the nucleocapsid (N) which is the most abundant viral-derived protein in SARS-CoV-2 where this virus can cause COVID19 disease. The aim of this study was to develop the SARS-CoV-2 N recombinant protein using *Escherichia coli* expression system. A total of 1,089 nucleotides encoding 362 amino acids of SARS-CoV-2 N was cloned to pET-14b vector. The plasmid then expressed in *E. coli* BL21 (DE3) and induced with 1.0 mM IPTG (Isopropyl-β-d-1-thiogalactopyranoside). The cell was harvested using denaturation lysis buffer due to inclusion body formation of SARS-CoV-2 N protein. Dialysis processed and concentrated using PEG-6000 resulted 0.992 mg/ml protein yield. Analysis of SARS-CoV-2 N recombinant protein using SDS-PAGE technique showed approximately 37.0 kDa specific band target protein. Application of this SARS-CoV-2 N recombinant protein to vaccinated and non-vaccinated antibody serum samples using ELISA technique indicated the significant result of optical density mean at 0.603 and 0.135, respectively. This study revealed that the production of SARS-COV-2 N recombinant protein could be carried out in *E. coli* expression system under denatured conditions, therefore the methods are more effective in producing the protein as a basic material in immuno-diagnostic assay.

1. Introduction

The COVID-19 has become a pandemic in worldwide and has a significant impact on human health that affects the economic and social behavior. COVID-19 was pneumonia caused by severe acute respiratory syndrome strain 2 (SARS-CoV-2). The virus was classified as a genus of β-coronavirus that has a positive single ribonucleic acid (RNA) cylindrical shape (Huang *et al.* 2020). The structural protein genome of the SARS-CoV-2 virus consists of nucleocapsid (N), envelope (E), membrane (M), and spike (S) proteins (Li *et al.* 2020). The nuceocapsid (N) protein contained 419 amino acids and has a molecular weight between 35-60 kDa, depending on the post-translational modification (Rosales *et al.* 2021). N protein has a large amount in SARS-CoV-2 virus, therefore it has a good immunogenicity and induce an antibody response in the patient body (Timany *et al.* 2004). During these COVID-19 pandemic, there are some research focus on study of N protein and revealed that the SARS-CoV-2 N protein has similarities to the SARS CoV-N protein (Rosales *et al.* 2021). Several diagnostic test methods confirmed the presence of IgM and IgG antibodies against N protein in blood serum (Minggu *et al.* 2021).This could provide a new strategy for producing SARS-CoV-2 N protein could be used as of serological tests in the diagnosis of COVID-19.

The COVID-19 vaccination program was expected to be able to trigger antibodies (IgM and IgG) response against the SARS-CoV-2 virus. One of the efforts to determine the success of COVID-19 vaccination is by detecting the presence of antibodies after being vaccinated. Enzyme-linked immunosorbent assay (ELISA) as one of serological analysis method can

* Corresponding Author
E-mail Address: hudada@apps.ipb.ac.id

be applied to detect SARS-CoV-2 antibodies from blood serum samples. However, the availability of this test materials especially the antigen is relatively rare, therefore it is needed to produce the in house antigen through recombinant protein technology.

In this study, we developed the SARS-CoV-2 N recombinant protein as an antigen material for ELISA diagnostic testing. The plasmid Pet-14b carrying the codon optimized *E. coli* BL21 (DE3) expression system. This SARS-CoV-2 N recombinant protein was expected to provide the diagnostic test material in supporting antibody based immunoassay testing.

2. Materials and Methods

2.1. Construction and Codon Optimized of the SARS-CoV-2 N Synthetic Gene

The SARS-CoV-2 N protein gene sequence was obtained from GenBank data base at access code MZ306692.1. The nucleotides that play a role in individual usage codons were optimized using bioinformatics programs (GenScript., USA). The codon optimized SARS-CoV-2 gene were inserted with NcoI and BamHI restriction enzymes to be cloned into the pET-14b *E. coli* expression vector then ordered as synthetic genes at a biotechnology service provider (GenScript., USA).

2.2. Expression and Isolation of N Recombinant Protein in the *E. coli* BL21 (DE3)

The expression process was initiated by transforming the pET-14b plasmid containing the SARS CoV-2 N gene into *E. coli* BL21 (DE3) using the heat shock method. The cells were grown and propagated using 10 ml Luria Bertani broth (Invitrogen) growth medium containing 100 µg/ml of ampicillin (Sigma) and then shaking incubated at 37°C 150 rpm for overnight. *E. coli* was then transferred to 200 ml Lb Broth medium containing 100 µg/ml and then incubated for 3 hours. The overexpressed *E. coli* was then induced with 1.0 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich, St. Louis, MO, USA) and then incubated for 4 hours (checking every 1 hour). *E. coli* cells was then centrifuged at 3000 xg for 15 min. Expressed cell pellets was dissolved using denatured buffer solution (7 M urea, 100 mM NaH_2PO_4, and 100 mM Tris.CI) and incubated in the cold for 1 hour. The solution was then centrifuged for 20 minutes at 16,000 xg and the supernatant was collected as a lysate.

A total of 4.5 ml of the lysed supernatant was inserted into the 14,000 Da cellophane membrane and tied at each end using a clamp, then dialyzed against 1.0 L of phosphate buffered saline (PBS). The buffer was changed every 2 hours for 4 times and the dialysis process was continued overnight. The clear lysate from the dialysis process was then concentrated using PEG-6000 to one tenth of the initial volume. The SDS-PAGE method (BioRad) was used to analyze the the expression results of the N SARS CoV-2 recombinant protein by observing the formation of a specific protein target band of N protein. As much as 15 µL of sample solution was mixed with 15 µL of sample buffer (1 M pH 6.8 Tris, 20% SDS, glycerol, 0.1% bromophenol blue, 1 M DTT,), then heated for 2 min and loaded into the acrylamide gel well and run at 100 V for 60 minutes. The gel was stained with Coomassie blue (Biorad) and incubated using an incubation shaker for 15 minutes. The gel was rinsed with destaining solution (200 ml methanol, 70 ml acetic acid, and 730 ddH_2O) and incubated for 60 min. To measure the concentration of N SARS CoV-2 recombinant protein, the bicinchoninic acid (BCA) technique was used by adding 1 ml of BCA reagent to the protein solution in a test tube and incubated for 30 min at 37°C. Absorbances analysis was carried out using ELISA reader at 595 nm wavelength.

2.3. Interaction of the SARS-CoV-2 N Antigen to the Antibody Using the Enzyme-Linked Immunosorbent Assay (ELISA) Technique

The ELISA test was carried out to determine the activity of the recombinant N SARS CoV-2 protein produced as an antigen that would interact with the SARS CoV-2 antibody. The samples used was human blood serum (ethical approval number 494/IT3.KEPMSM-IPB/SK/2021) who had been vaccinated or had been exposed to the SARS CoV-2 virus and non vaccinated samples. The conjugate used was anti-human IgG peroxidase conjugate reacted with TMB chromogenic substrate. The optical density (OD) was detected using a microplate ELISA reader at 450 nm wavelength.

3. Results

3.1. Construction and Codon Optimization of the SARS-CoV-2 N Synthetic Genes

The SARS CoV-2 N target gene that was used in this study was at the position of 172 to 1260 (1089 nucleotides) or at 58 to 419 of the N protein (encoded 362 amino acids) referring to GenBank with the access code MZ306692. 1 (Figure 1). Codon optimized was carried out on the target gene to avoid the bias codons effect and to increase the efficiency of expression due to the presence of unpreferred codon in *E. coli* expression systems. In this study, the codon optimization was succeeded in increasing the codon adaptation index (CAI) from 0.36 to 0.96 and GC content from 46.33 to 59.55%. The SARS CoV-2 N codon-optimized gene was inserted with NcoI (CCATGG) and BamHI (GGATCC) adapted to restriction enzymes at the multiple cloning site (MCS) position in the pET14b vector (Figure 2).

3.2. Expression and Isolation of N Recombinant Protein in *E. coli* BL21 (DE3)

A total of 362 amino acid of the SARS CoV-2 N recombinant protein has expressed indicated with 37.0 kDa moleculer weight of target protein (Figure 3). IPTG was used to induce the target protein expression shown by specific thick protein band compared non induced cells, however only a few target protein was found in soluble condition and most of this protein was in the inclusion bodies. Based on this result, we harvested the target protein from inclusion body cell using denaturation condition with denaturated buffer. Therefore, recombinant SARS-CoV-2 N protein was highly expressed in *E. coli* BL21 (DE3) in the inclusion body form and isolated using denaturation methods.

In order to purify the target protein, we used the affinity chromatography system containing Ni2+-NTA chelating resin, however it did not work, possibly due to 6x histidin amino acid was denaturated and could not bind to the resin. Therefore, we tried to dialyzed against PBS buffer to remove all of residual denaturating buffer (such as urea) and some non-specific small size protein (less than 14 kDa). Concentrating the isolate using PEG6000 was succeeded in increasing the yield measured with BCA assay resulted 0.992 mg/ml.

3.3. Interaction of N SARS CoV-2 Antigen to the Antibodies Using Enzyme-linked Immunosorbent Assay (ELISA) Technique

The activity of SARS-COV-2 N recombinant protein was analyzed by using ELISA technique to determine the interaction with antibody against SARS CoV-2. Referring to Table 1, the in-house SARS-CoV-2 N recombinant protein applied as an antigen reacted with SARS-CoV-2 positive antibodies in the samples indicated with higher optical density (OD) value at the average of 0.603. Meanwhile, the OD value average of negative antibodies (non-vaccinated) were 0.135.

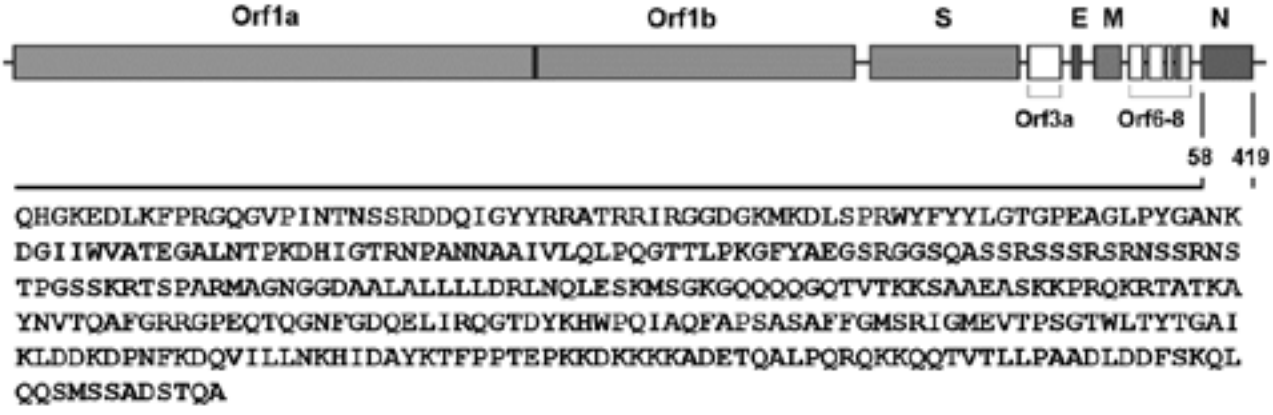

Figure 1. Diagram of the genetic map of the SARS-COV-2 N protein target at positions 58 to 419 (362 amino acids). Orf, Open reading frames; S, spikes; E, Envelopes; M, Membrane; N, Nucleocapsid. (Image modified from Wu *et al.* 2020)

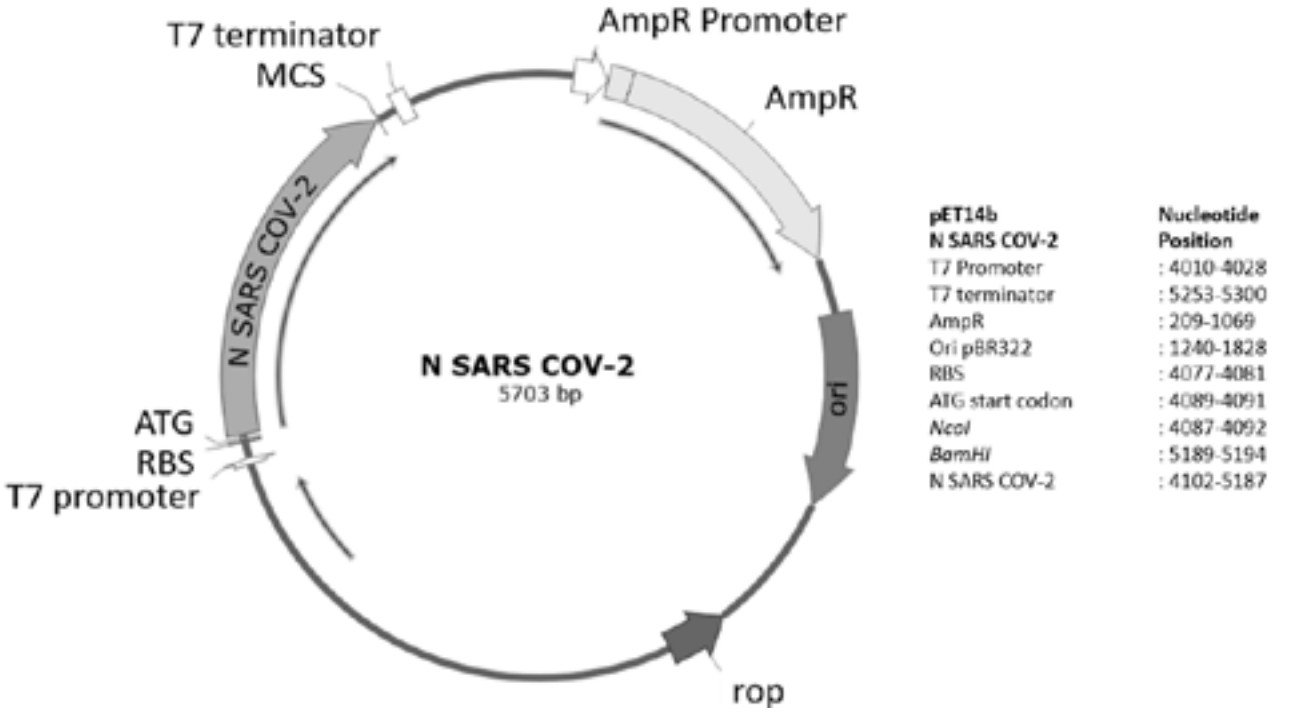

Figure 2. Vector map of pET14b inserted with 1089 SARS-CoV-2 N optimized for codons on the multiple cloning site (MCS) side. RBS, Ribosome binding site; ATG, start codon; AmpR, ampicillin receptor; Ori, origin of replication

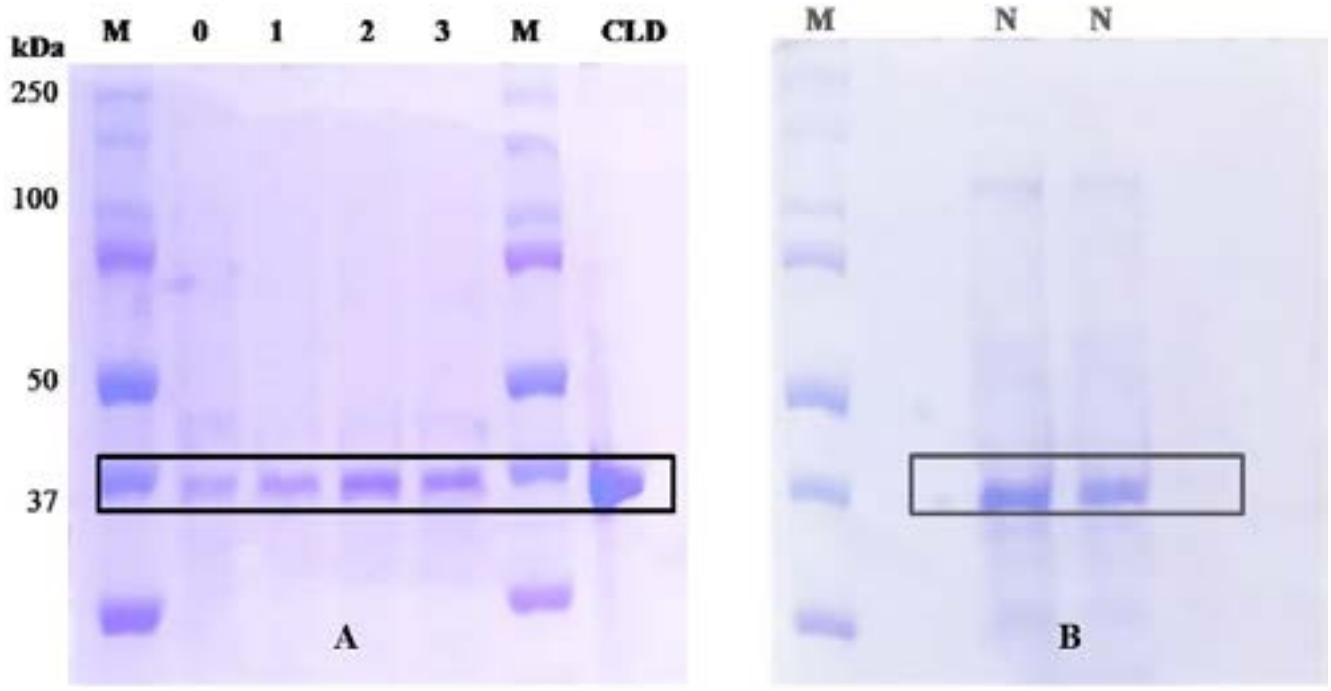

Figure 3. Electrophorogram analysis of the recombinant SARS-CoV-2 N protein using SDS-PAGE. Panel A. Specific protein bands after induction with IPTG at 0, 1, 2, and 3 hr; Panal B. The target protein was lysed by denaturing buffer followed by dialysis and concentration. (M. Protein marker, 0-3–duration of IPTG induction, CLD. Denatured lysate, N. Lysate dialysis)

Table 1. Confirmation test results for the SARS-CoV-2 antigen binding to antibodies against vaccinated and unvaccinated serum samples using the ELISA method

Sampel ID	Sampel ID			
	Test I	Test II	Average	Samples group
S1a	0.814	0.871	0.842	Vaccinated
S2a	0.708	0.773	0.740	
S3a	0.231	0.223	0.227	
S1b	0.018	0.016	0.017	Unvaccinated
S2b	0.070	0.060	0.065	
S3b	0.070	0.050	0.060	

4. Discussion

In this study, we performed the codon optimization from native SARS-CoV-2 N template gene in order to increase the expression level in heterologous host. The optimization result was succeeded in increasing the GC percentage were in accordance with the ideal value at 30 to 70%. A high GC value (>70%) can reduce efficiency and inhibit translation process, while a low GC value (<30%) can causes a delay in the transcriptional elongation process (Rosano and Ceccarelli 2009). According to (Gustafson *et al.* 2014; Merkl 2003; Silaban and Subroto 2016), the differences in the percentage of GC content between the target gene and host expression causes codon bias, where a high codon bias can reduce expression levels and was thought to cause translation errors and produce inactive complex aggregate proteins that cause degradation into different fragments. The codon optimization of SARS-CoV-2 N synthetic gene increased the CAI value to 0.96, theoretically the CAI value >0.8 was predicted to be suitable for expression in *E. coli* hosts (Carbone *et al.* 2003; Xia 2007). CAI was a value that measures the similarity between the use of gene and reference codons (Sharp and Li 1978; Yuliana *et al.* 2021). Several researchers reported that CAI values could be used to predict the expression level of heterologous proteins (Carbone *et al.* 2003; Fox and Erill 2010; Xia 2007). According to Brown (2002) in Yuliana *et al.* (2021), codon optimization was highly correlated with the level of gene expression and the frequency of use of codons encoding amino acids has different variations for each organism.

Plasmid pET-14b carrying 1089 nucleotides of SARS CoV-2 N synthetic gene encoding 362 amino acids has been successfully expressed using *E. coli*. The pET-14b vector has a T7 promoter capable of binding to T7 RNA polymerase. The T7 expression system will be active in the presence of the T7 RNA polymerase enzyme, where T7 RNA polymerase was not naturally founded in bacteria. However, *E. coli* BL21 (DE3) has been designed to carry the gene encoding RNA polymerase in its DNA segment. IPTG was used to induce recombinant protein expression that has a function to prevent the lac repressor from binding to the lac operator thereby inducing gene expression to form an active protein. In this study, IPTG induction was attempted to express the target protein, but only a few target proteins were found in soluble form and most of them were in the inclusion body. The description of IPTG induction results in the SARS-CoV-2 N protein expression process was shown in Figure 3.

SARS CoV-2 nucleocapsid protein was sensitive to *E. coli* proteases and has the ability to self-proteolyze in the original condition and there was no post-translational modification in the *E. coli* expression system. Therefore, it often experiences folding error and cause the protein not to be expressed as a soluble fraction but formed the aggregates as insoluble proteins or inclusion bodies. The occurrence of inclusion bodies in the target protein required a process of modification or alteration of the protein. Denaturation was a biochemical process that can be used to modify protein structures without causing the breaking of convalent bonds. Proteins can form a zwitter ion structure and have an iso-electric point with the same number of positive and negative ionic charges on the protein (Saepuloh *et al.* 2020; Singh *et al.* 2015), so that in this situation the protein can be denatured. The mechanism of denaturation took place when a substitution reaction occurs between negative and positive ions. In this study, the buffer solution for the denaturation process consisted of chemical compounds such as urea, NaH_2PO_4, Tris.CI. These chemical compounds were compound that can break hydrogen bonds that cause protein denaturation because they can break hydrophobic interactions and increase the solubility of hydrophobic groups in water (Singh *et al.* 2015). The denatured clear lysate obtained and confirmed the presence of the SARS-CoV-2 N target protein was 37.0 kDa molecular weight (Figure 3). The presence of a good and clear target protein band from the denaturation resulted proves that the target protein is appropriate.

The dialysis process of a solution was determined by the pH and the strength of the solute ionization needed to balance the molecules being dialyzed. In this study, the solvent used was PBS solvent which is an isotonic solution and is non-toxic to cells and has the ability to maintain osmolarity. The use of PBS buffer in the dialysis process aims to remove non-specific proteins (Rosales *et al.* 2021). Dialysis lysate molecules with small sizes that were inserted into the cellophane membrane will

exit through the membrane pores, while molecules with larger sizes will be retained in the membrane sac. The dialysis lysate was then concentrated using PEG6000 which has the ability to bind to the target protein. The results of SDS-PAGE characterization showed that after dialysis and concentration using PEG6000 to remove other proteins that were not specific, the target protein bands were cleaner and thicker (Figure 3B). Measurement of SARS-CoV-2 N recombinant protein concentration was measured using the BCA method resulted 0.992 mg/ml.

ELISA was one of the serological tests to detect the presence of antibody or antigen components from the sample being tested. The advantages of using the ELISA test are good reagent stability factors as well as high safety and sensitivity (Voller *et al.* 1976). Application of this SARS CoV-2 N recombinant protein to vaccinated and non-vaccinated antibody serum samples using ELISA technique indicated the significant result of optical density mean at 0.603 and 0.135, respectively. This study revealed that the production of SARS COV-2 N recombinant protein could be carried out in *E. coli* expression system under denatured conditions, therefore the methods are more effective in producing the protein as a basic material in immuno-diagnostic assay. The procedure was relatively simple and inexpensive, therefore it could be carried out on a large scale to obtain antigen sources that can be used in the development of serological diagnostic methods for COVID-19.

Acknowledgements

This research is funded by Excellence Basic Research of Higher Education (PDUPT, 1938/IT.L1/TN/2021), Ministry of Education,Cultute, Research, and Technology, Republic of Indonesia and cooperation with Primate Research Center (PSSP), IPB University, Indonesia.

References

Brown, T., 2002. *Genomes*, second ed. Wiley-liss, Oxford.

Carbone, A., Zinovyev, A., Kepes, F., 2003. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics.* 19, 2005-2015. https://doi. 10.1093/bioinformatics/btg272

Fox, J.M., Erill, I., 2010. Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Research.* 17. 185-196. https://doi.10.1093/dnares/dsq012

Gustafson, D.J., Gibson, D.J., Nickrent, 2004. Competitive relationships of A*ndropogon gerardii* (Big Bluestem) from remnant and restored native populations and select cultivated varieties. *Functional Ecology*. 18, 451-457. https://doi.org/10.1111/j.0269-8463.2004.00850.x

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., *et al.* 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 15, 497-506. https://doi.10.1016/S0140-6736(20)30252-X

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *The New England Journal of Medicine.* 382, 1199-1207. https://doi.10.1056/NEJMoa2001316

Merkl, R.A., 2003. Survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. *Journal of Molecular Evolution.* 57, 453–466. https://doi.10.1007/s00239-003-2499-1

Minggu, R.B., Rumbajan, J.M., Turalaki, G.L.A., 2021. Struktur genom severe acute respiratory syndrome coronavirus-2(SARS-CoV-2). *Journal of Biomedik.* 13, 233-240. https://doi.org/10.35790/jbm.13.2.2021.31996

Rosales, J.D., Quintero, W., Cruz, J., Perdomo, B., Quintero, M., Bastidas, M., Lugo, J.D., Rodriguez, K.R., Freites-Perez, J.C., Castillo, A.C., 2021. Expression and novel alternative purification of the recombinant nucleocapsid (N) protein of SARS-CoV-2 in *Escherichia coli* for the serodiagnosis of COVID-19. *BioRvix.* https://doi .10.1101/2021.11.10.467990

Rosano, G.L., Ceccarelli, E.A., 2009. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb Cell Fact.* 8, 1-9. https://doi.org/10.1186/1475-2859-8-41

Saepuloh, U., Iskandriati, D., Pamungkas, J., Solihin, D.D., Mariya, S.S., Sajuthi, D., 2020. Construction of a preliminary three-dimensional structure simian betaretrovirus serotype-2 (SRV-2) reverse transcriptase isolated from Indonesian cynomolgus monkey. *Tropical Life Sciences Research.* 31, 47–61.

Sharpl, P.M., Li, W., 1987. The codon adaptation index – a measure of directional synonymous codon usage bias and its potential application. *Nucleic acids Res.* 15, 1281-1295. https://doi. 10.1093/nar/15.3.1281

Silaban, S., Subroto, T., 2016. Kajian ekspresi gen pretrombin-2 manusia sintetik pada *Escherichia coli* secara *in silico* untuk produksi trombin sebagai komponen lem fibrin. JPKIM. 8, 58-64. https://doi.org/10.24114/jpkim.v8i1.4425

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M., Panda, A.K., 2015. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb Cell Fact*. 14, 41. https://doi.org/10.1186/s12934-015-0222-8.

Timani, K.A., Ye, L., Ye, L., Zhu, Y., Wu, Z., Gong, Z., 2004. Cloning, sequencing, expression, and purification of SARS-associated coronavirus nucleocapsid protein for serodiagnosis of SARS. *Journal of Clin Virol.* 30, 309–312. https://doi.10.1016/j. jcv.2004.01.001

Voller, A., Bidwell, D.E., Huldt, G., Engvall, E.O., 1974. A microplate method of enzyme-linked immunosorbent assay and its application to malaria. *Bull World Health Organ.* 51, 209–211.

Xia, X., 2007. An improved implementation of codon adaptation index. *Evolutionary Bioinformatics.* 3, 53-58. https://doi.org/10.1177/117693430700300028

Yuliana, Saepuloh, U., Suryani, 2021. Development of a recombinant Taq DNA polymerase enzyme expressed using a synthetic gene and its comparison with a commercial enzyme. *Asia Pacific Journal of Molecular Biology and Biotechnology*. 29, 43-50. https://doi.org/10.35118/apjmbb.2021.029.2.05