

## Optimization of the *apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G (APOBEC3G)* gene to enhance its expression in *Escherichia coli*

Rizkyana Avisya,<sup>1</sup> Silvia Tri Widyaningtyas,<sup>2</sup> Budiman Bela<sup>2,3</sup>



pISSN: 0853-1773 • eISSN: 2252-8083  
<https://doi.org/10.13181/mji.oa.202853>  
**Med J Indones.** 2020;29:120–8

**Received:** November 14, 2018

**Accepted:** February 21, 2020

### Authors' affiliations:

<sup>1</sup>Master Program of Biomedical Science, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia, <sup>2</sup>Virology and Cancer Pathobiology Research Center, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia <sup>3</sup>Department of Microbiology, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

### Corresponding author:

Budiman Bela  
 Virology and Cancer Pathobiology Research Center, Faculty of Medicine, Universitas Indonesia, IASTH Building 8th floor, Jalan Salemba Raya No.4, Senen, Central Jakarta 10430, DKI Jakarta, Indonesia  
 Tel/Fax: +62-21-31930353  
 E-mail: budiman.bela@yahoo.com

### ABSTRACT

**BACKGROUND** Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G (APOBEC3G) can abolish HIV infection by inducing lethal mutations in the HIV genome. The HIV protein virion infectivity factor (Vif) can interact with APOBEC3G protein and cause its degradation. Development of a method that can screen substances inhibiting the APOBEC3G-Vif interaction is necessary for identification of substances that potentially used in anti-HIV drug development. In order to increase expression of recombinant APOBEC3G protein that will be used in APOBEC3G-Vif interaction assay, we developed an optimized APOBEC3G gene for expression in *Escherichia coli*.

**METHODS** The gene coding APOBEC3G was codon-optimized in accordance with prokaryotic codon using DNA 2.0 software to avoid bias codons that could inhibit its expression. The APOBEC3G gene was synthesized and sub-cloned into pQE80L plasmid vector. pQE80L containing APOBEC3G was screened by polymerase chain reaction, enzyme restriction, and sequencing to verify its DNA sequence. The recombinant APOBEC3G was expressed in *E. coli* under isopropyl-β-D-thiogalactoside (IPTG) induction and purified by using nickel-nitrilotriacetic acid (Ni-NTA) resin.

**RESULTS** The synthetic gene coding APOBEC3G was successfully cloned into the pQE80L vector and could be expressed abundantly in *E. coli* BL21 in the presence of IPTG.

**CONCLUSIONS** Recombinant APOBEC3G is robustly expressed in *E. coli* BL21, and the APOBEC3G protein could be purified by using Ni-NTA. The molecular weight of the recombinant APOBEC3G produced is smaller than the expected value. However, the protein is predicted to be able to interact with Vif because this interaction is determined by a specific domain located on the N-terminal of APOBEC3G.

**KEYWORDS** APOBEC3G, codon usage, prokaryote gene expression

Apolipoprotein B mRNA editing catalytic polypeptide-like-3 (APOBEC3) refers to the human polynucleotide cytosine deaminase family. APOBEC3 can be incorporated to the HIV-1 virion released from infected cells and catalyze the deamination of cytosine into uracil during reverse transcription of the HIV genome. Deamination leads to hypermutation in the viral genome, which degrades the viral RNA.<sup>1</sup> The virion infectivity factor (Vif) protein of the HIV binds to APOBEC3 type G (APOBEC3G) and E3 ubiquitin ligase

complex, leading to APOBEC3G degradation by the proteasome. Hence, APOBEC3G cannot be incorporated into the new virion and lethal hypermutation of the viral genome cannot occur.<sup>2</sup> The development of antiretrovirals inhibiting the interaction between Vif and APOBEC3G is a promising approach to prevent HIV infection. The function of APOBEC3G as a cytidine deaminase in new virions is maintained by inhibiting the Vif and APOBEC3G interaction, thus, the HIV-1 replication cycle can be disrupted. Some research

on Vif-APOBEC3G interaction inhibitors have been published. Nathans et al,<sup>3</sup> for example, identified a Vif-antagonist molecule that could increase APOBEC3G protein in cells and enhance APOBEC3G incorporation into new virions.

Identification of small-molecule inhibitors by *in vitro* methods requires APOBEC3G protein expression. Some researchers have cited challenges related to the full-length APOBEC3G protein expression and purification in a prokaryotic system, such as low levels of expression due to the genotoxic property of APOBEC3G and the poor solubility of the protein.<sup>4</sup> APOBEC3G expression has been conducted in other expression systems, such as the baculovirus expression system and mammalian cell cultures.<sup>4,5</sup> However, they are relatively more expensive and more difficult to handle.

Some modifications have been conducted to express APOBEC3G in prokaryotic systems. Earlier research, for instance, used partially expressed APOBEC3G protein, specifically, its N-terminal or C-terminal domain only, with some mutations to improve its solubility.<sup>4</sup> Li et al<sup>6</sup> claimed to have successfully expressed full-length APOBEC3G fused with a tag protein in the N-terminal of APOBEC3G in a prokaryotic system. However, the effect of tag protein on the activity of APOBEC3G and its interaction with Vif has not been reported. Moreover, the effect of APOBEC3G codon optimization on the ability of APOBEC3G to be expressed in a prokaryotic system has not been described. Plasmids containing the optimized APOBEC3G gene could be used to express and purify APOBEC3G recombinant protein, which is essential in the development and screening of antiretrovirals targeting the APOBEC3G-Vif protein interaction *in vitro*. In the present research, we optimized the APOBEC3G codon to avoid codon bias between organisms and

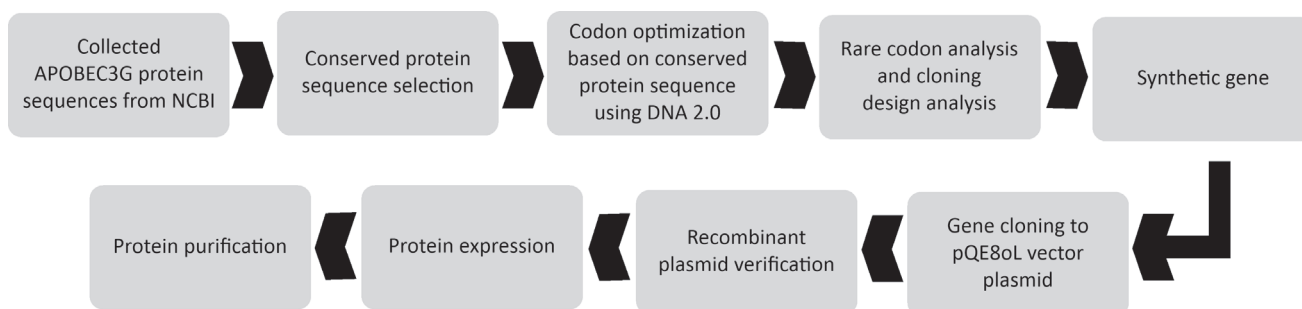
enhance the expression of the protein in *Escherichia coli*. The optimized gene was cloned to the pQE80L plasmid to construct a new plasmid that could express and purify large amounts of recombinant APOBEC3G protein.

## METHODS

This research was conducted at the Virology and Cancer Pathobiology Research Center, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, from March 2017 to December 2017. The study design is described schematically in Figure 1. This research included a bioinformatics study of the protein sequence, codon optimization, gene cloning, protein expression analysis, and purification.

### Codon optimization of the APOBEC3G gene for *E. coli*

APOBEC3G protein sequences were downloaded from the National Center for Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov/protein](http://www.ncbi.nlm.nih.gov/protein)) protein database to begin designing an optimized gene. These sequences were aligned by using the multiple sequence alignment tool Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/>). The identity matrix, a parameter indicating the similarity of amino acids among APOBEC3G proteins, was also measured by using Clustal Omega. The APOBEC3G protein with the highest similarity to other APOBEC3G proteins was chosen. The APOBEC3G proteins were back-translated to DNA using Gene Designer 2.0 software (ATUM, USA). During back-translation, the appropriate codons of APOBEC3G amino acids were optimized for *E. coli*. The DNA sequence obtained from Gene Designer 2.0 was further analyzed by using the rare codon analysis tools provided by GenScript (GenScript Biotech Corporation, USA, available at <https://>



**Figure 1.** Study design of APOBEC3G gene optimization and its expression in *Escherichia coli*. APOBEC3G=apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G; NCBI=National Center for Biotechnology Information

www.genscript.com/tools/rare-codon-analysis). The optimized DNA sequence was selected as the gene candidate under the following conditions: codon adaptation index (CAI) between 0.8 and 1.0, guanine-cytosine (GC) content of 30–70%, codon frequency distribution (CFD) of <30%, and no negative cis or repeat elements. CAI scores were formulated by comparing the relative synonymous codon usage of the codon used with the maximum possible CAI score of the codon for each amino acid.<sup>7</sup> CAI scores ranged from 0 to 1, and a higher score indicates that a recombinant protein is more likely to be expressed in a certain expression system.<sup>8</sup> CAI scores >0.8 indicate that a recombinant protein can be well expressed in a certain expression system. GC content variations between bacterial genomes and coding genes range from 30% to 75%.<sup>9</sup>

Restriction enzyme sites within the optimized DNA were analyzed by using the online tool NEBcutter V2.0 (New England Biolabs [NEB] Inc., USA, available at <http://nc2.neb.com/NEBcutter2/>). Optimized DNA was ordered as a synthetic gene and cloned into a pUC57 universal plasmid at Integrated DNA Technologies, Inc. (IDT, Singapore). Later in this study, the synthetic gene cloned in pUC57 was designated as pUC57-APOBEC3Gopt.

#### **Cloning of the optimized APOBEC3G gene into a prokaryotic expression system with the pQE80L plasmid**

Plasmids pQE80L (QIAGEN, USA) and pUC57-APOBEC3Gopt (synthesized by IDT) were digested by using HindIII (NEB) and BamHI (NEB) restriction enzymes, respectively. Exactly 10 µg of pQE80L and 10 µg of pUC57 containing the optimized APOBEC3G gene were subjected to HindIII (NEB) restriction for 2–4 hours at 37°C. The digested plasmids were then purified by using a QIAEX II Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions.<sup>10</sup> The purified plasmid was subsequently digested for 2–4 hours of incubation with BamHI at 37°C. The digested plasmids were subjected to electrophoresis on low-melting temperature agarose (1.2 g of low-melting agarose [LMA] method in 1× tris-acetate-ethylenediaminetetraacetic acid buffer containing 0.08% crystal violet). Linearized pQE80L and APOBEC3G DNA fragments were gouged out from LMA and purified by using a QIAEX II Gel Extraction Kit (QIAGEN) according to the manufacturer's

instructions.<sup>10</sup> The purified and linearized pQE80L plasmid and APOBEC3Gopt DNA fragment were combined by using T4 DNA ligase (Thermo Fisher Scientific, USA) with an insert gene-to-vector mass ratio of 1:3 and incubated for 16 hours at 16°C. This combination was transformed into a freshly made *E. coli* TOP10 competent cell. The ligation reaction control was prepared by transforming 20 ng of the remaining digested pQE80L used for ligation to freshly made *E. coli* TOP10 competent cells.

#### **Screening of recombinant pQE80L containing the APOBEC3G gene**

Colony polymerase chain reaction (PCR) was performed to screen recombinant bacteria that bear recombinant pQE80L containing the APOBEC3G gene. The reaction was conducted by using DreamTaq DNA Polymerase (Thermo Fisher Scientific) with universal pQE-forward and reverse primers (IDT). During agarose gel electrophoresis, colonies producing a DNA band of 1,200 base pair (bp) in length were supposed to contain recombinant pQE80L. Colonies were grown in Luria-Bertani (LB) broth, and the recombinant plasmids were isolated by using a QIAprep Spin Miniprep kit (QIAGEN) to verify the plasmid recombinant in each colony. The isolated plasmids were confirmed by using BamHI and HindIII enzymes. The recombinant plasmid containing the APOBEC3G gene produced two DNA bands (pQE80L, 4,700 bp; APOBEC3G gene, 1,200 bp) during 0.8% agarose electrophoresis. Recombinant plasmids containing these two DNA bands were sequenced, and the verified recombinant plasmid was named pQE80L-APOBEC3Gopt. The DNA sequencing results were analyzed by using MEGA 7 (available at <https://www.megasoftware.net>).<sup>11</sup> The DNA sequence was translated into a protein from the start codon of pQE80L until the stop codon of the APOBEC3Gopt gene using the same tool. Finally, the protein sequence was aligned with the expected APOBEC3G protein sequence, which was an APOBEC3G protein sequence with 6×-histidine in the N-terminal of the translated protein.

#### **APOBEC3G expression in *E. coli***

The pQE80L-APOBEC3Gopt plasmid was transformed into chemically competent *E. coli* BL21 strain cells and was spread on LB agar. Two or three colonies grown in LB agar were grown in LB broth

overnight at 37°C in a rotary shaker. The overnight culture was then grown in Terrific broth (TB). The overnight culture volume was 10% of the total culture volume in TB. After 2 hours of incubation, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and 1 ml of the culture was taken every hour over 4 hours of induction. The culture samples were centrifuged at 12,000 rotation per minute (rpm) for 1 min, diluted with 1×SDS-sample buffer, and subjected to SDS-PAGE analysis. Gels were stained with PageBlue™ staining solution (Thermo Fisher Scientific, Singapore). The expression of APOBEC3Gopt protein was manifested by the overexpression of a protein of 46 kDa in size after IPTG induction. Recombinant BL21 colonies expressing the desired protein were subjected to large-scale expression. Then, 1 liter of the selected colony culture was grown in the same manner as in the small-scale expression, and the cell pellet was stored at -20°C. For comparison, a pQE80L recombinant plasmid containing the non-optimized APOBEC3G gene was transformed into *E. coli* BL21-CodonPlus strain for optimal protein expression and subjected to protein expression analysis in the same manner as the optimized gene. The molecular weight of the recombinant protein in SDS-PAGE gel were determined by using GelAnalyzer 19.1 software (<http://www.gelanalyzer.com/>).

### Protein purification

The recombinant protein was purified by using immobilized metal affinity chromatography (IMAC) under denaturing conditions. Lysis buffer (NaH<sub>2</sub>PO<sub>4</sub>, 10 mM Tris, 15 mM imidazole, 6 M guanidine HCl, pH 8.0) were added to the cell pellet, and the dissolved pellet was incubated in a rotary incubator shaker at 60 rpm and 4°C for 1 hour and then centrifuged at 12,000 rpm and 4°C. Nickel-nitrilotriacetic acid (Ni-NTA) resin was washed with washing buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0) and added to the supernatant. This mixture was incubated in a rotary shaker at 60 rpm for 1 hour at 4°C. Thereafter, the tube was centrifuged at 12,000 rpm for 1 min at 4°C, and the supernatant was discarded. The Ni-NTA resin was washed five times with washing buffer C (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM Tris, 500 mM NaCl, 8 M urea, pH 6.3). The recombinant protein was eluted five times using buffer D (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM Tris, 500 mM NaCl, 8 M urea, 20% glycerol, pH 5.9) and

then eluted twice with buffer E (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM Tris, 500 mM NaCl, 8 M urea, 20% glycerol, pH 4.5). The buffer composition used in this work were optimized modification based on manufacturer instruction, to fit a condition which the protein were able to purified.

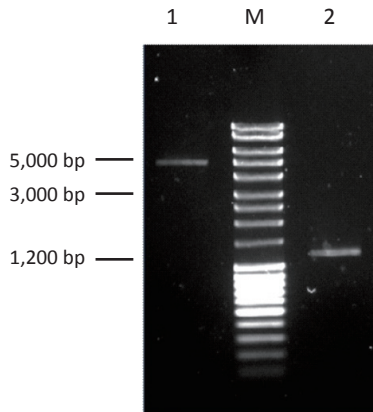
## RESULTS

### Codon optimization

Codon optimization began with selection of the APOBEC3G protein sequence to be used as a template. APOBEC3G protein sequences were obtained from NCBI using the keyword "APOBEC3G." The results showed seven *Homo sapiens* APOBEC3G sequences with accession numbers AGS14892.1, AEA39616.1, NP\_068594.1, EAW60292.1, AAH24268.1, AAZ38722.1, and AGI04219.1. Multiple sequence alignment of the sequences revealed that four of the seven sequences were identical. Therefore, one of the identical sequences, AAZ38722.1, was chosen as the template sequence for codon optimization.

The amino acid sequences of APOBEC3G (AAZ38722.1) were back-translated into codon sequences. Codons that are abundant in *E. coli* and in accordance with the amino acids of APOBEC3G were selected to build an *E. coli* codon-optimized DNA sequence encoding APOBEC3G. The codon-optimized APOBEC3G DNA was determined to have a CAI of 0.9, GC content of 48.41%, with no low-frequency codon and no negative cis nor repeat elements by using online rare codon analysis tools.

In order to facilitate the insertion of APOBEC3G DNA into the multiple cloning site of pQE80L plasmid DNA, HindIII and BamHI restriction recognition sequences were added to form flanking sequences of the APOBEC3G open reading frame. The HindIII restriction recognition sequence was added 5' to the APOBEC3G open reading frame, hence upon translation of the recombinant protein that starts with the histidine tag sequence the HindIII sequence will be translated into amino acid sequence prior to translation of the APOBEC3G sequence, thus becoming part of the recombinant protein. The BamHI restriction recognition sequence was added 3' to the APOBEC3G open reading frame downstream to the stop codon of the APOBEC3G sequence. The codon-optimized APOBEC3G DNA was ordered and then synthesized and cloned into pUC57 by IDT.



**Figure 2.** Linearized *APOBEC3Gopt* DNA fragment and pQE80L vector. Lane 1: linearized and purified pQE80L vector; lane 2: purified *APOBEC3Gopt* gene fragment; and lane M: marker. bp=base pair

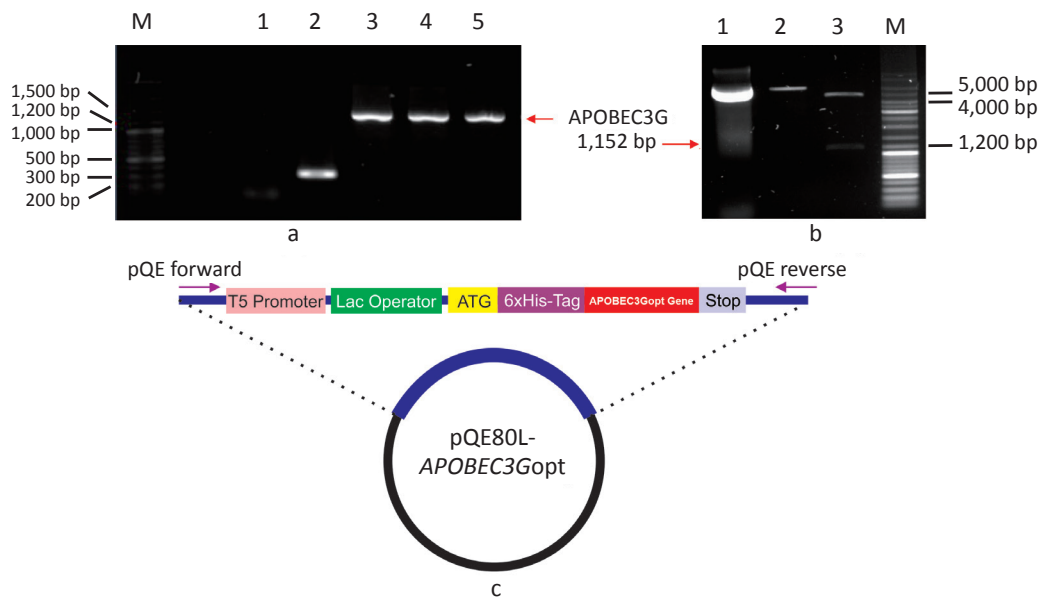
**Cloning of optimized *APOBEC3G* gene into a prokaryotic expression system plasmid**

Plasmid pUC57 coding the optimized *APOBEC3G* gene, designated as pUC57-*APOBEC3Gopt*, was transformed into *E. coli* TOP10 cells. Plasmids were then analyzed by using BamHI and HindIII before it used on a larger scale. The expected *APOBEC3Gopt* DNA fragment, which was approximately 1,200 bp in length, was separated from the vector in LMA and purified by using a QIAEX II kit. The linearized pQE80L vector, which is around 4,700 bp in length, and the

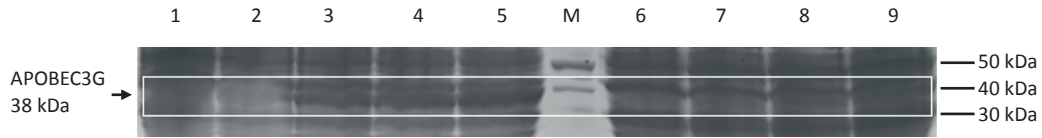
*APOBEC3Gopt* fragment were confirmed by DNA electrophoresis in 0.8% agarose, as shown in Figure 2. The linearized pQE80L vector and *APOBEC3Gopt* DNA fragment were ligated and transformed to chemically competent *E. coli* TOP10 cells.

Screening of recombinant colonies using PCR showed that some colonies produced the expected DNA band of approximately 1,500 bp in length, as seen in Figure 3a. The positions of the universal pQE primers used in colony PCR were 100 bp upstream and 150 bp downstream of the pQE80L multiple cloning site. A colony that appeared to contain the recombinant plasmid based on colony PCR was grown in LB broth, and plasmids were isolated by using QIAprep Spin Miniprep kit and then double-digested using HindIII and BamHI. The expected recombinant plasmid produced two DNA bands of approximately 4,700 and 1,200 bp and produced only one band of around 6,000 bp when digested with HindIII only (Figure 3b). Double-digestion with two restriction enzymes showed that plasmids isolated from those colonies contained the optimized *APOBEC3G* gene ( $\pm 1,200$  bp) in the pQE80L vector (4,700 bp). The gene constructs of the desired recombinant pQE80L-*APOBEC3Gopt* plasmid and the amplified gene in the PCR reaction are shown in Figure 3c.

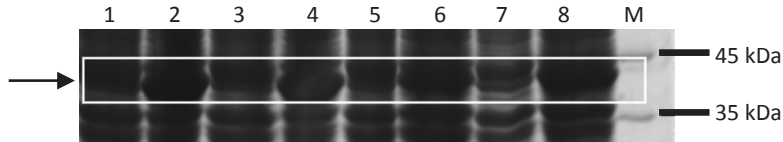
The base accuracy of the *APOBEC3G* gene was confirmed by sequencing, which showed that the



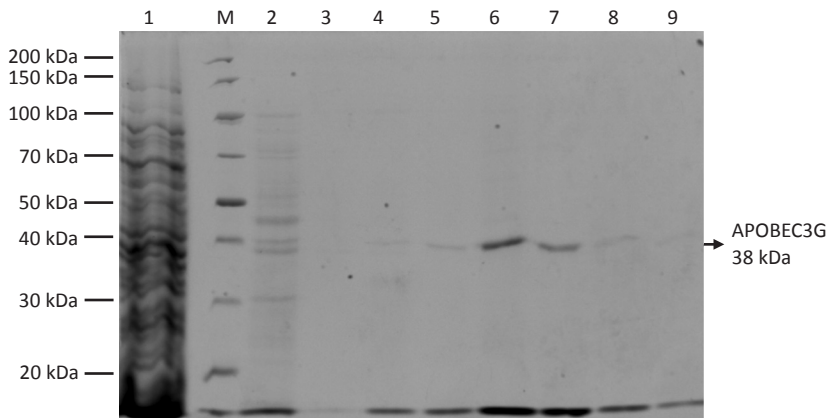
**Figure 3.** pQE80L-*APOBEC3Gopt* recombinant plasmid confirmation. (a) Colony PCR of *Escherichia coli* TOP10 containing ligated plasmid pQE80L-*APOBEC3Gopt*. Lane 1: negative control, lane 2: pQE80L; lane 3: colony 1, lane 4: colony 2, lane 5: colony 3, and lane M: marker; (b) isolated recombinant plasmid confirmation. Lane 1: non-digested pQE80L-*APOBEC3Gopt*, lane 2: pQE80L-*APOBEC3Gopt* digested using HindIII, lane 3: pQE80L-*APOBEC3Gopt* digested using BamHI-HindIII, and lane M: marker; (c) recombinant codon-optimized *APOBEC3G* gene construct in pQE80L-*APOBEC3Gopt*. bp=base pair; *APOBEC3G*=apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G; PCR=polymerase chain reaction



**Figure 4.** Optimized *APOBEC3G* gene expression. Lanes 1–5: *Escherichia coli* BL21 pQE80L-*APOBEC3Gopt*; 1: before IPTG induction, 2: 1 hour after induction, 3: 2 hours after induction, 4: 3 hours after induction, and 5: 4 hours after induction. Lanes 6–9: *E. coli* BL21 pQE80L; 6: before IPTG induction, 7: 1 hour after induction, 8: 2 hours after induction, and 9: 3 hours after induction. *APOBEC3G*=apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G; IPTG=isopropyl β-D-1-thiogalactopyranoside



**Figure 5.** Comparison of protein expression profile between optimized and non-optimized recombinant *APOBEC3G* proteins. Lanes 1 and 2: *Escherichia coli* BL21 pQE80L-*APOBEC3Gopt* colony 1 before and after IPTG induction, respectively; lanes 3 and 4: *E. coli* BL21 pQE80L-*APOBEC3Gopt* colony 2 before and after IPTG induction, respectively; lanes 5 and 6: *E. coli* BL21 CodonPlus pQE80L-*APOBEC3G* non-optimized codon, colony 1, before and after IPTG induction, respectively; lanes 7 and 8: *E. coli* BL21 CodonPlus pQE80L-*APOBEC3G* non-optimized codon, colony 2, before and after IPTG induction, respectively, and lane M: marker. *APOBEC3G*=apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G; IPTG=isopropyl β-D-1-thiogalactopyranoside



**Figure 6.** *APOBEC3G* protein purification SDS-PAGE result. Lane 1: flow through; lanes 2–3: protein purification with buffer C; lanes 4–5: protein purification with buffer D; lanes 6–9: protein elution with buffer E; and lane M: marker. *APOBEC3G*=apolipoprotein B mRNA editing enzyme catalytic polypeptide-like-3G

*APOBEC3G* gene cloned in pQE80L was identical to the designated gene. Moreover, no mutation that may affect the amino acid sequence was found. Protein translation analysis based on the recombinant plasmid sequencing result using MEGA demonstrated an amino acid sequence identical to that of the desired *APOBEC3G* protein with six units of histidine as the protein tag on the N-terminal.

**Optimized *APOBEC3G* gene expression in *E. coli***

Codon-optimized recombinant *APOBEC3G* was expressed in *E. coli* BL21. IPTG was used to induce the expression of the recombinant protein. The protein expression profile is shown in Figure 4. The expression of recombinant *APOBEC3G* was manifested by the

overexpression of a protein weighing around 38 kDa; pQE80L-containing BL21 and wild-type BL21 did not overexpress this 38 kDa protein. However, the molecular weight of recombinant *APOBEC3G* was slightly different from the predicted value (48.3 kDa). The expression of the codon-optimized *APOBEC3G* gene was more abundant than that of *APOBEC3G* that was not codon-optimized for prokaryotic expression systems (48 kDa), as seen in Figure 5.

**Protein purification**

Recombinant *APOBEC3G* protein was purified by using IMAC with a Ni-NTA matrix under denaturing conditions. During protein purification, 8 M urea buffers with gradually decreasing pH were used. The

recombinant protein began to elute when buffer D (pH 5.9) was added to Ni-NTA, and most of the protein was eluted out by using buffer E (pH 4.5), as shown in Figure 6.

## DISCUSSION

APOBEC3 is an important natural antiviral protein in the human body. The APOBEC3 family consists of several types encoded A–H.<sup>12</sup> Among the APOBEC3 members, APOBEC3G and APOBEC3F possess the highest potential activity toward HIV-1 replication; more importantly, APOBEC3G protein purification SDS-PAGE result has 10–50 times higher activity compared with APOBEC3F.<sup>13</sup> Therefore, APOBEC3G is preferred as a target in novel antiretroviral research compared with APOBEC3F.

Determination of the abundance of a recombinant protein is important in *in vitro* studies. A large amount of pure recombinant protein is needed to produce antibodies in biochemical research and protein interaction studies. The protein expressed in this research could be used to develop a new method to screen potential APOBEC3G-Vif interaction-inhibiting substances as new HIV-1 drugs.

We have cloned a full-length APOBEC3G amplified from humans in a prokaryotic system previously, but the expression level of APOBEC3G protein was very low when expressed in *E. coli* BL21 strain CodonPlus (data not published). Hence, in this study, we attempted to establish a plasmid containing an optimized APOBEC3G gene designed for higher expression in the prokaryotic system (*E. coli*). pQE80L was chosen as the vector because the tag for protein purification is relatively small. Therefore, tag polypeptides should not affect the structure of the protein and its interaction with Vif.

Codon optimization is usually used to counter problems in protein expression in different organisms due to bias codons. In this study, codon optimization was conducted by using the *E. coli* codon table and human codons in the APOBEC3G protein were changed in accordance with the codons frequently found in *E. coli*. Low-frequency codons often lead to low expression levels in *E. coli*.<sup>8</sup> The ability of an optimized codon to be expressed in a particular expression system could be predicted by measuring its CAI, CFD, GC content, and presence of negative cis and repeat elements. The CAI of the codon-optimized APOBEC3G gene in this study was 0.9, which indicates that the

gene could be expressed straightforwardly in *E. coli*. The optimized gene does not contain low-frequency codons and includes only 3% slightly low-frequency codons. GC content also affects the expression level of a gene because it can affect the stability of mRNA. High GC contents in a coding sequence are positively correlated with DNA and mRNA stability and the length of the coding sequence.<sup>9</sup> The optimized APOBEC3G gene had 48.41% GC content, which is the optimum range for gene expression in *E. coli*.

The optimized APOBEC3G gene cloned in the pQE80L plasmid vector was expressed in *E. coli* BL21 strain under the regulation of the coliphage T5 promoter. The codon-optimized APOBEC3G was cloned into the BamHI and HindIII restriction sites, which are the first and last restriction sites, respectively, of the pQE80L multiple cloning site, to reduce the number of additional amino acids included in the protein. The expression result of APOBEC3G in BL21 with 1 mM IPTG induction shows that the protein is expressed at a relatively high level, however, the molecular weight of this protein was lower than expected. Theoretically, the molecular weight of the codon-optimized APOBEC3G protein fused with the histidine tag is 48.3 kDa while the actual result was only 38 kDa. Furthermore, comparison with the protein expression profile of the non-codon-optimized recombinant APOBEC3G protein in *E. coli* BL21-CodonPlus strain showed that the codon-optimized APOBEC3G protein is expressed more abundantly. Interestingly, the protein molecular weights obtained, at around 38 kDa, were identical. This finding indicates that the discrepancies in molecular weight are not due to codon optimization. The sequencing analysis results of the purified pQE80L-APOBEC3Gopt showed no mutation compared with the designed sequence. Sequence translation analysis using bioinformatics tools showed no mutation, premature stop codons in the gene, or other gene alterations that could change the protein expressed from the plasmid. Plevoda et al<sup>4</sup> reported that the recombinant APOBEC3G produced in Sf9 cells by using the pFastBac vector showed an approximately identical result. The full-length recombinant APOBEC3G protein expressed with the 4×his-tag has a molecular weight of approximately 37–40 kDa in SDS-PAGE, which is lower than the predicted molecular weight. Li et al<sup>6</sup> successfully cloned full length APOBEC3G to pET32 vector plasmid and produced the recombinant APOBEC3G in *E. coli*.

The SDS-PAGE result of their recombinant APOBEC3G protein fused with thioredoxin was not clear enough to confirm the molecular weight of APOBEC3G. Hence, it is not assured that the APOBEC3G were expressed in full length, or the molecular weight of APOBEC3G expressed were identical as the theoretical molecular weight.

Discrepancies in the predicted molecular weight and SDS-PAGE results could be attributed to several factors. Since the SDS-PAGE method is based on denaturation of the protein 3D structure using SDS as a detergent, therefore, protein structure and SDS-binding to the protein has an important role in protein movement in SDS-PAGE gel. Some intrinsic properties of the protein that could affect protein migration in SDS-PAGE include acidic amino acid residues, post-translational modification, specific amino acid domains and regions in the protein, and the SDS-binding capacity of the protein, which is influenced by the hydrophobicity and hairpin, tertiary, and quaternary structures of the examined protein.<sup>14-16</sup>

APOBEC3G protein, previously known as CEM15, is 46 kDa protein coded by chromosome 22 of the *Cbx6* and *Cbx7* genes. As a member of the APOBEC3 family, APOBEC3G possesses the zinc-binding motif (H/C)XE(X)<sub>23-28</sub>PCXXC.<sup>12</sup> APOBEC3G has two zinc-binding motifs or cytidine deaminase domains located in its N- and C-terminals. APOBEC3G undergoes homooligomerization via its C-terminal domain in amino acid residues 209–336, which must remain catalytically active. The N-terminal domain is only a pseudo-catalytic domain, it does not undergo oligomerization and provides an interaction site with viral proteins, such as Gag, for incorporation into virions. This domain also interacts with the Vif protein of HIV1 to block APOBEC3G's activity.<sup>2,4</sup>

The protein was purified using Ni-NTA resin. Six units of the histidine tag allow the protein to attach tightly to the resin. Hence, the protein remained bound to Ni-NTA while other proteins were washed away by the washing buffer. The purified protein was eluted at certain conditions, such as low pH and ionic strength. In the pQE80L plasmid, the histidine tag was located in the N-terminal of the expressed protein. The N-terminal of APOBEC3G interacts with Vif. Thus, we attempted to purify the protein to examine whether the N-terminal of the expressed APOBEC3G remains intact and is not digested by the protease.

The purification process occurred only in denaturing conditions with the use of urea as a mild denaturant. During the purification process, we found that the APOBEC3G could not be purified and eluted in the regular denaturing buffer. Thus, we optimized a denaturing buffer with high ionic strength that could wash most contaminant proteins and elute the desired protein. The molecular weight of the purified protein was similar to that obtained in the previous step. Hence, the expressed protein was successfully purified.

Purification of the recombinant APOBEC3G protein by using Ni-NTA revealed that the six units of the histidine tag located in the amino terminal of the recombinant protein maintained its function. Hence, the 38 kDa recombinant APOBEC3G protein retains its N-terminal, which is the domain for Vif interaction. This finding indicates that the protein could still be used in interaction studies of APOBEC3G and Vif, screening of small-molecule inhibitors of the Vif-APOBEC3G interaction, and antibody production. However, the purification process involved the use of a protein denaturant, which could interfere with the 3D structure of the protein. Therefore, the protein obtained must be further purified from its buffer containing denaturant. The protein must be dialyzed so that its 3D structure can return to its normal form.

Further research on screening of the potential inhibitor towards Vif-APOBEC3G interaction *in vitro* by using the recombinant APOBEC3G and Vif protein is expected, thus, a new antiretroviral drug featuring sensitivity not affected by viral mutations is expected in the near future. However, the purified protein needs to be confirmed by Western blot. In this research, the protein were not confirmed by Western blot method yet, since the protein were not able to be transferred neither to polyvinylidene fluoride nor nitrocellulose membranes during protein transfer from SDS-PAGE gel to the membrane in Western blot process (data not shown). Furthermore, this study did not evaluate the activity of the recombinant APOBEC3G protein, the ability of the recombinant APOBEC3GB to catalyze enzymatic deamination, and the ability of the protein to interact with recombinant Vif protein. Further research is necessary to screen the potential APOBEC3G-Vif interaction inhibitor completely.

In conclusion, codon optimization of the APOBEC3G gene was successfully carried out. The codon-optimized recombinant protein expressed in the prokaryotic



system is more abundant than the non-optimized APOBEC3G gene.

#### Conflict of Interest

The authors affirm no conflict of interest in this study.

#### Acknowledgment

We thank the *Insentif Riset Sistem Inovasi Nasional (Insinas)*–Ministry of Research, Technology, and Higher Education for supporting this study.

#### Funding Sources

This study was funded by *Insentif Riset Sistem Inovasi Nasional (Insinas)*–Ministry of Research, Technology, and Higher Education.

## REFERENCES

- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell*. 2003;113(6):803–9.
- da Costa KS, Leal E, dos Santos AM, Lima E Lima AH, Alves CN, Lameira J. Structural analysis of viral infectivity factor of HIV type 1 and its interaction with APOBEC3G, EloC and EloB. *PLoS One*. 2014;9(2):e89116.
- Nathans R, Cao H, Sharova N, Ali A, Sharkey M, Stranska R, et al. Small-molecule inhibition of HIV-1 Vif. *Nat Biotechnol*. 2008;26(10):1187–92.
- Polevoda B, McDougall WM, Bennett RP, Salter JD, Smith HC. Structural and functional assessment of APOBEC3G macromolecular complexes. *Methods*. 2016;107(1):10–22.
- Iwatani Y, Takeuchi H, Strebel K, Levin JG. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J Virol*. 2006;80(12):5992–6002.
- Li L, Yang YS, Li ZL, Zeng Y. Prokaryotic expression and purification of HIV-1 Vif and hAPOBEC3G, preparation of polyclonal antibodies. *Virolog Sin*. 2008;23(3):173–82.
- Oliver JL, Marin A. A relationship between GC content and coding-sequence length. *J Mol Evol*. 1996;43(3):216–23.
- Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987;15(3):1281–95.
- Muto A, Osawa S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A*. 1987;84(1):166–9.
- Qiagen. QIAEX® II Handbook for DNA extraction from agarose and polyacrylamide gels and for desalting and concentrating DNA from solutions. 2015.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
- Jónsson SR, Andrésdóttir V. Host restriction of lentiviruses and viral countermeasures: APOBEC3 and Vif. *Viruses*. 2013;5(8):1934–47.
- Zennou V, Bieniasz PD. Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. *Virology*. 2006;349(1):31–40.
- Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A*. 2009;106(6):1760–5.
- Iakoucheva LM, Kimzey AL, Masselon CD, Smith RD, Dunker AK, Ackerman EJ. Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci*. 2001;10(7):1353–62.
- Guan Y, Zhu Q, Huang D, Zhao S, Lo LJ, Peng J. An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Sci Rep*. 2015;5:13370.