

EXPRESSION OF GENE CODING ENDOGLUCANASE GH5-4 DERIVED FROM MEAGENOMIC DNA DATA OF BACTERIA IN GOATS RUMEN IN *Escherichia coli*

**Nguyen Hai Dang^{1,†}, Do Thi Huyen^{1,2,†}, Nguyen Thi Kien³, Ha Thi Thuy Hoa¹,
Le Quynh Giang¹, Dao Trong Khoa^{1,2}, Truong Nam Hai^{1,2,*}**

¹Institute of Biotechnology, VAST, Vietnam

²Graduate University of Science and Technology, VAST, Vietnam

³VNU University of Science, Hanoi, Vietnam

Received 02 February 2021, accepted 02 June 2021

ABSTRACT

Glycoside Hydrolase family 5 (GH5) members share a broad range of enzymatic activities on oligosaccharides, polysaccharides and glycoconjugates from a wide range of species. The subfamily 4 (GH5-4) is enriched in some broad-specificity endo- β -1,4-endoglucanases (EG). From metagenomic DNA data of bacteria in Vietnam goats' rumen, a gene GL0361920 coding for endoglucanase GH5-4 was mined and selected for expression in *Escherichia coli*. Firstly, the codons of the gene were optimized and the codon optimized gene (*eg3*) was artificially synthesized and inserted into pET21a(+) at *NdeI-XhoI* positions for expression in *E. coli*. The results of the gene expression in five *E. coli* strains analyzed by SDS-PAGE showed that the recombinant endoglucanase (EG3) coded by *eg3* was the best expressed in BL21, Origami strains, was not expressed in JM109 strain and expressed at very little amount in C43 and Rosetta strains. Among LB, TB, LB, SB, TBD media, the recombinant BL21 strain grew best at TB medium and produced total EG3 higher than the other media. The inducer IPTG gave a negative effect on the growth of the cells but the increase concentration of IPTG from 0.05 mM to 0.9 mM did not increase the negative impact on the cell mass and produced EG3 was nearly the same at the different IPTG concentrations. At 25 °C, in the TB medium containing 0.1 mM IPTG, the recombinant *E. coli* BL21 strain harboring pET21-*eg3* produced about half of endoglucanase EG3 in a soluble form exhibiting activity hydrolyzing CMC substrate. At the OD₆₀₀=10, soluble EG3 accumulated in the recombinant *E. coli* BL21 strain had the activity of 0.22 ± 0.006 U/ml.

Keywords: CBM, endoglucanase, GH5, goats' rumen, metagenomic DNA data.

Citation: Nguyen Hai Dang, Do Thi Huyen, Nguyen Thi Kien, Ha Thi Thuy Hoa, Le Quynh Giang, Dao Trong Khoa, Truong Nam Hai, 2021. Expression of gene coding endoglucanase GH5-4 derived from meagenomic DNA data of bacteria in goats rumen in *Escherichia coli*. *Academia Journal of Biology*, 43(2): 17–26. <https://doi.org/10.15625/15870>

*Corresponding author email: tnhai@ibt.ac.vn

©2021 Vietnam Academy of Science and Technology (VAST)

INTRODUCTION

Sustainable, biological solutions to the rising climate change and energy problems have been of increasing concern to scientists and environmentalists in recent years (de Jong et al., 2012). In comparison to biofuel production from edible polysaccharides, such as starch, recent efforts have concentrated on next generation bioenergy, with higher-energy fuels produced from inedible lignocellulosic biomass contained in a variety of abundant plant materials (Asgher et al., 2013; Ofori-Boateng & Lee, 2013). However, lignocellulose biomasses have complex, rigid and impermeable structures against enzymatic attack. Thus conversion of lignocellulose to bioethanol requires three steps: pretreatment by acid/base or heat, hydrolysis by a mixture of enzymes, and fermentation. Hydrolysis is considered one of the most expensive processing steps in the conversion of lignocellulose to fermentable sugars because this step requires a mass of complex enzymes. This motivates attempts to isolate or design more efficient enzyme mixtures for lignocellulose conversion (Carriquiry et al., 2011).

Among lignocellulose degrading enzymes, endoglucanase is one of the most important enzymes hydrolyzing cellulose into short oligosaccharides for other enzymes such as cellobiose hydrolase, beta-glucosidase to digest further into glucose. Endoglucanases belong to some glycoside hydrolases (GHs) including GH5. The GH5 has more than 50 subfamilies (Do et al., 2014, 2018; Hess et al., 2011; Duan et al., 2009; Elifantz et al., 2008). The GH5-4 subfamily includes several broad-specificity endoglucanases that hydrolyze cellulose, xyloglucan, and mixed-linkage glucans. Endoglucanases GH5-4 are capable of broad specificity without compromising high activity, which helps them to maximize their use in biomass degradation processes (Glasgow et al., 2020).

In this study, a gene (code GL0361920) encoding endoglucanase GH5-4 was mined from metagenomic DNA data of bacteria in Vietnam goats' rumen. The gene was optimized codon usage for high expression in

E. coli and the optimized gene (*eg3*) was inserted into pET21a(+) for expression of the gene in *E. coli*. Some conditions for recombinant enzyme expression in *E. coli* including expressive host strains, cultured media, induced IPTG concentration were investigated for obtaining active EG3. The recombinant enzyme will be used for enzyme characterization and assessment of the enzyme ability to be used in industry.

MATERIAL AND METHODS

Material

The gene (GL0361920) composed of 1725 nucleotides was mined from metagenomic DNA data of bacteria living in Vietnam goats rumen. In which, 78 nucleotide at 5' terminal coded for a signal peptide. The rest sequence coded for an endoglucanase containing two conserved domains GH5-4 and CBM72. The gene was predicted to be originated from a negative gram *Bacteroidales*.

Methods

Codon optimization and design of expression vector pET21-eg3

The gene was optimized codons for suitable expression in *E. coli* (designated as *eg3*) and artificially synthesized at Biomatik company (USA) then inserted into pET21a(+) at *NdeI-XhoI* position to generate pET21-*eg3*. The *eg3* in the recombinant vector was sequenced by the Sanger method. Then, the pET21-*eg3* was extracted to have amount for transformation of the gene into *E. coli* strains. The obtained plasmid was checked by restriction enzyme *NdeI* and *XhoI*. The total volume of digested reaction was 10 μ l, composed of 1 μ l Tango buffer 2X (10X); 3 μ l DNA (50 ng/ μ l), 0.3 μ l *NdeI/XhoI* (Fermentas, 5 U/ μ l). The reaction was performed at 37 °C for 1 hour then analyzed by electrophoresis on agarose gel 0.8%.

Investigation of some conditions for eg3 expression

The plasmid pET21-*eg3* was transformed into 5 expressive strains of *E. coli* including BL21 (DE3), C43, JM109, Origami and

Soluble to select suitable strains for the expression of endoglucanase EG3.

For expression of EG3, a single recombinant strain was cultured in 5 ml liquid LBA medium (the LB medium complemented with ampicillin 100 µg/ml) overnight at 37 °C, 170 rpm. The overnight cultures were diluted 50 times into LBA medium and continued incubation at the same condition until OD₆₀₀ reached 0.6-0.8 then the cells were induced for EG3 expression by adding 0.5 mM IPTG. The samples were cultured at 30 °C, 180 rpm for 4 hours.

To select a suitable medium for EG3 expression, 5 media were investigated including LB (1% tryptone, 0.5% yeast extract, 1% NaCl), SB (3.2% peptone, 0.5% NaCl, 2% yeast extract), PE (1% yeast extract, 2% peptone), TB (1.2% peptone, 2.4% yeast extract, 72 mM K₂HPO₄, 17 mM KH₂PO₄, 0.4% glycerol) and (5) TBD (1.2% peptone, 2.4% yeast extract, 72 mM K₂HPO₄, 17 mM KH₂PO₄, 0.24% glucose). In addition, the concentrations of IPTG (0, 0.05, 0.1, 0.3, 0.6, 0.9 mM) for induction of *eg3* transcription also were investigated.

After culturing, the cells were harvested by centrifugation at 8,000 rpm for 5 minutes and suspended in water to OD₆₀₀ = 10 then kept at -20 °C until analysis.

For assessment of the EG3 expression in soluble or insoluble fractions, the cells were sonicated (10 pulses, 30 s each at 100 W with 20 s intermission) in the ice bath. The soluble fraction was separated from the pellet by centrifugation at 13,000 rpm for 10 min at 4 °C. The expressed proteins in soluble and insoluble fractions were analyzed by SDS-PAGE in 12.6% polyacrylamide gel. The endoglucanase activity was identified by hydrolyzing 1% carboxymethyl cellulose (CMC) on an agar plate containing 1.4% agar in PBS pH6. Accordingly, the protein fractions (20 µl) or 10 µl of cellulase (0.5 U/ml, Sigma) were dropped in Whatman paper (5 mm diameter) placed on the agar plate. After incubating at 37 °C for about 18 hours to allow the enzyme to diffuse deeply

into the agar and hydrolyze the substrate. The dish was stained with 0.1% Congo red and then washed with 1M NaCl solution until the hydrolysis zones were observed.

Determination of enzyme activity by DNS method

To quantify cellulose hydrolysis activity, Miller's DNS method of quantifying sugar reduction was applied (Lorentz Miller, 1959). The total reaction volume was 0.5 ml containing 225 µl of protein, 25 µl PBS buffer pH6 and 250 µl of 1% CMC. The reaction was performed at 37 °C for 120 min then stopped by adding 500 µl of DNS solution and boiled for 10 minutes. The samples were cooled down at room temperature then measured at 540 nm. An endoglucanase active unit (U) is defined as the amount of the enzyme that catalyzes CMC to reduced sugar equivalent to 1 µmol glucose per minute under experimental conditions.

RESULTS

Codon optimization of the gene and design of expression vector pET21-eg3

The principle for designing a recombinant vector for the expression of a foreign gene is, first of all, the gene sequence inserted into the vector must be in the correct open reading frame to ensure that the amino acid sequence is not altered during transcription. In addition, it is necessary to minimize the addition of extra amino acids coded from nucleotide sequence in polycloning site of vector to the N and C ends of the target protein. So in this case, we used *NdeI* and *XhoI* for cloning the gene into pET21a(+) and intended to use the original signal peptide to introduce proenzyme to *E. coli* periplasm. At the same time, the His-tag sequence in the vector was fused at C terminal of EG3 to keep for convenience of protein purification.

The origin gene (1725 nucleotides) had 50.11% GC content. The analysis of codon usage in *E. coli* showed that the gene had codon adaptation index (CAI) of 0.7. However, the gene employed 4% rare codons that can reduce the efficiency of translation rate of the

gene in *E. coli*. Thus the gene codons were optimized to increase CAI to 0.96, GC content of 53.85 and rare codons reduced to 0%. The sequence of the codon optimized gene (*eg3*) was showed in Figure 1. The *eg3* sequence was synthesized, inserted into pET21a(+) to

generate pET21-*eg3*. The gene *eg3* in the plasmid was confirmed by sequencing. On the other hand, the plasmid was large scale extracted to transformed into different *E. coli* strains for the gene expression. The pET21-*eg3* has checked digestion with *NdeI* and *XhoI*.

origin	ATG AAG AAA AAA CTA CTA TTA TCA GCG ACT CTT TTG CTG CTT TCA ATG ACT GCA ACC	57
Optimized	CATG ..G ..G C.G AGC ..C ..G C... ..G ..GC ..C ...	
origin	AAT GCA CAG AAT TCC TCC ACA GAC GTA GCT CCT GAC AAC ACG GGT ATG GAT CTT ACA GCC	117
Optimized	..C ..GC AG. AG. ...C ..T ..G ..G ..G ..TC ..CG ..C ..G	
origin	AAG CAG TGG ACG AAG AAT GTG GTG ATG GGT TGG AAC CTC GGA AAC TCA CTA GAG AGC CAA	177
Optimized	..AC ..ATCGCGGAG	
origin	GGT GGC GAG ATG GGA TGG GGA AAC CCT CGT ACC ACA CAG CAG ATG ATT AAG GAC GTA AAG	237
Optimized	..CACCGCCCCT ..G ..A	
origin	GCT CAA GGC TTC AAT GCC ATC CGT ATT CCT GTC AGA TGG ACC GAG CAT CTC TCT GAT AAG	297
Optimized	..G ..GTCG C.TG	
origin	GCC AAT ATG GTT GTC AGC AAC GAT TGG CTG GCG CGT GTC AAG GAA ATC GTC GAC TGG TGT	357
Optimized	..GGG ..GCCCCG ..AT ..G ..TC	
origin	ATG GCT GAA GAT ATG TAT GTC ATC ATC AAT GTT CAT CAC GAG GCG TGG CTG GAC CGT CAT	417
OptimizedCG ..TT ..C ..GCCAGT ..C ..C	
origin	CCG CAG AAG GCC ACC AAG GCG GAG AAC AAC AAG AAG CTG GCA GCC TTG TGG AAA TGC ATA	477
OptimizedAGAATACG C... ..CT	
origin	GCT ACC TAT TTC CGT GAC TAT GAC CAA AAA CTG GCT TTT GCA GGT ACC AAC GAG ACC ATC	537
Optimized	..CTTTGCGCG ..CAT	
origin	TCT CTA GAC GCA AAT GGA CAG GAA AAC TGG GGC GAG CCT ACC GCA GAA TAT CAG GAG GTT	597
Optimized	AGC ..G ..T ..G ..CCTGAGCA ..G	
origin	CAG AAC TCA TAT AAT CAG ACT TTC ATC GAT GCT GTA CGC GCC ACA GGA GGC AAG AAC TAC	657
OptimizedAGCCCTTG ..GCCAT	
origin	TAC CGC AAC CTG GTT GTA CAG ACT TAT GCC TGC AGT GCA TGG AAT GGA TTC AAG GGT TTC	717
OptimizedG ..GCGCGCGCT ..A ..C ..T	
origin	GTC ATT CCA ACC GAC CAG GTA GAG GAG CGC CTT AGC GTT GAG GTT CAT AAC TAT GAC CCT	777
Optimized	..GTAAGGG ..AGCT ..G	
origin	TAC GAG TAT GCC GGC GGC GGA ACC TAT TAC TAT TGG GGT GTC AAA TAT AAG AAC ATG GGT	837
OptimizedACCCCCGCATC	
origin	TAC TCT GTG CCT TCC AGT AAT GAG CAG TCG ATG ATT GAC TAC ATG AAC CGA CTC CGC AAT	897
Optimized	... AGCG AG. ..CAAGCCTCT ..G ...	
origin	ACG TGG AGC AAC AAG GGC CTT GGT GTC GTC ATT GGC GAA TAC GGT GCC ACT TGC CAT TAT	957
Optimized	..CCAG ..CTCCCCT	
origin	ACC GCT GAT AAT AAG CAG GTA CAG ATG GAG AAC CAG CAA TAT TGG TAT CAG ACG ATA GTA	1017
OptimizedCC ..AGAGCCCC ..T ..G	
origin	AGT GCT ATG CGT GAA CGT GGC TTC GCA GGA TTC GTC TGG GAC AAC AAT GCC TTT GGC AAC	1077
Optimized	..C ..GTG ..CTGCTGGT	
origin	GGC ACG GAG AAG TTT GGT ATC TTC CGT CGC TCG GCT ACG GGT ATG ACA GTT GGT AAC GAG	1137
OptimizedCACTCCCCCT ..A	
origin	TAT GCA CTC AAG GGC ATC TGC GAA GGA TCC GGT ACT GAG TAC AAA GAA TCT GGT TCT GGT	1197
Optimized	..C ..C ..G ..ATCC AG. ..C ..C ..ATAGC ..C AGC ..C	
origin	AGC GGT CAG GGT AAC CAG GAC ATC GAT GGT GGC ACC ACG CTA TGG GAG GGT AAC AGC ATG	1257
OptimizedCCTCCCCGACC	
origin	ATG GAC TGG GGC AAC GGT CTC CAG ATT ACT ATT TCT GGC TCC GAC TTC CAG GGC TAT GGC	1317
OptimizedTTTCGCAGCTTC	
origin	AAG GAT GTC GTC ATG CAG CTC AAC TAT ACC CTT GAC TAT ACC GAC TAT AAC ATG ATA CAG	1377
Optimized	..AG ..TG ..TCG ..TCTCT ...	
origin	CTG TTC TAT GGC GAC TGG GGC ACC AAT CCT TCG TTC TTT GTC GAT GGC ACT CCT GTC ACC	1437
OptimizedTTCGCG AGCTCGG	
origin	AAA GAG TTT ACA CCC TCC AGT CTT TAT CCT GTT GGC AAC AAC GAA TCC TGT GTG TCT AAT	1497
OptimizedCT ..G AG. ..C ..G ..C ..G ..GCCAG. ..CAGC ..C	
origin	ATC ACC TTC AGT GAG GAC GTC TAC AAC CAA CTG GTC TCC AAG GGA CTG GCT ATT CAG GGA	1557
Optimized	..TTG ..TTGCG AG. ..ACC	
origin	CAC GGT GTC CGG ATG AAC AAG GTG GTA CTG GGA GCT CCT ACA GGC ATC AGC ACT ATC AGT	1617
Optimized	..TC ..G ..TTGGCGCTC	
origin	ACT TCC TCG CTT GAT GGG GGT AAA TAT TAC ACG CTC AGC GGT ATG CGC AGC AAT ACT CCC	1677
Optimized	..C AG. AGC ..GC ..CCT ..C ..GCCC ..C ..G	
origin	ACT CGT GGC ATC TAT ATC CAG AAC GGA AAG AAA TAC ATG GTC AAA -TA G	1725
Optimized	..CCTCAAA TAC ATG GTC AAA -TA G C.C .AG	

Figure 1. The *eg3* sequences with original and optimized codons. The dots indicate the identical nucleotides at the same site compared to the original gene. The numbers on the right side indicate the nucleotide numbers.

As a result, the pET21-*eg3* was completely digested with *XhoI* to generate a DNA fragment of 7.1 kb observed in agarose gel as expected (Fig. 2). The pET21-*eg3* was incompletely digested with both *NdeI* and *XhoI* to generate 3 DNA fragments of *eg3* (1.7 kb), pET21a(+) (5.4 kb) and opened pET21-*eg3* (7.1 bp) (Fig. 2). Thus, the recombinant plasmid was ready to be transformed into *E. coli* strains for the gene expression.

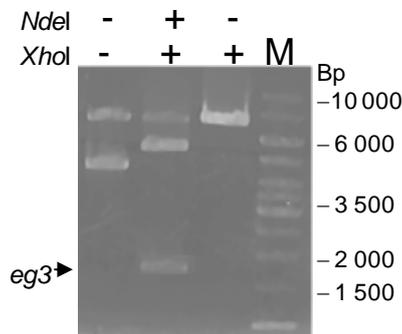


Figure 2. Analysis of DNA fragments from the digestion of pET21-*eg3* by *NdeI*, *XhoI* on agarose gel 1%. M: DNA ladder (1 kb, Fermentas)

Expression of EG3 in *E. coli* strains

E. coli is a popular host for the over-expression of recombinant proteins. There are a number of factors that can influence heterogeneous protein production and careful strain choice can greatly improve the chance of successful expression. As the obtained results (Fig. 3A) EG3 had the size of ~ 63 kDa was observed in the polyacrylamide gel in the lanes of total proteins from BL21 and Origami strains. The EG3 also could be found in the total protein expressed in C43 strain but with very little amount. On the petri dish containing CMC 1%, all protein fractions from five strains harboring pET21a(+) did not exhibit endoglucanase activity (data not shown), while total protein from BL21, Origami and C43 strains containing *eg3* gene exhibited the endoglucanase activity to hydrolyzing CMC to create clear zones around the Whatman papers (Fig. 3B). Surprisingly although enzyme EG3 was not seen in total protein from Soluble strain by SDS-PAGE we also observed enzyme activity on CMC substrate. EG3 was not expressed in strain JM109. Based on this result, the *E. coli* BL21 strain harboring pET21-*eg3* was selected for other studies.

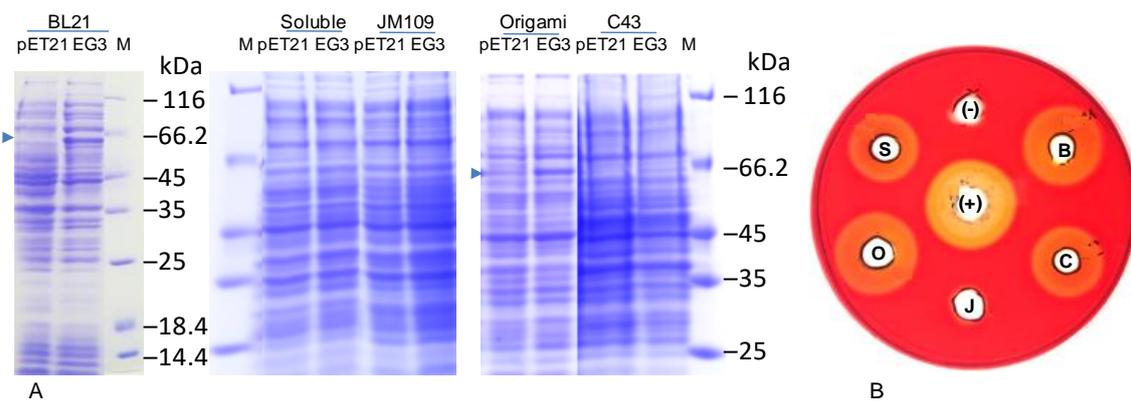


Figure 3. Analysis of EG3 expression in different *E. coli* strains by SDS-PAGE (A) and enzyme activities on CMC 1% (B). (+): cellulase (Sigma); (-): total protein of *E. coli* BL21 harboring pET21a(+); S: *E. coli* Soluble; B: *E. coli* BL21; C: *E. coli* C43; J: *E. coli* JM109; O: *E. coli* Origami. Arrows indicate the position of expressed EG3

Selection of medium for EG3 expression

Each protein will be synthesized to a different degree in culture media with different nutritional components. There are 5 cultural media that were investigated and the cells were induced by 0.5 mM IPTG at 30 °C for 4 hours. Based on the results of the electrophoresis image (Fig. 4A), at the same OD₆₀₀, the expression level of EG3 in five media was nearly the same. All total proteins containing the expressed EG3 in five media also exhibited endoglucanase activity to

hydrolyze CMC to create clear halo (Fig. 4B). In addition, the major part of EG3 was in the insoluble fractions, only small parts of EG3 were soluble. Almost all EG3 expressed in insoluble fractions did not show endoglucanase activity, whereas EG3 expressed in soluble form exhibited very good activity (Fig. 4B). The TB and TBD media are the rich nutrient including nitrogen and carbon sources, which facilitates cell proliferation as well as the over-expression of foreign proteins. Thus TB is the medium of choice due to the most EG3 protein yield.

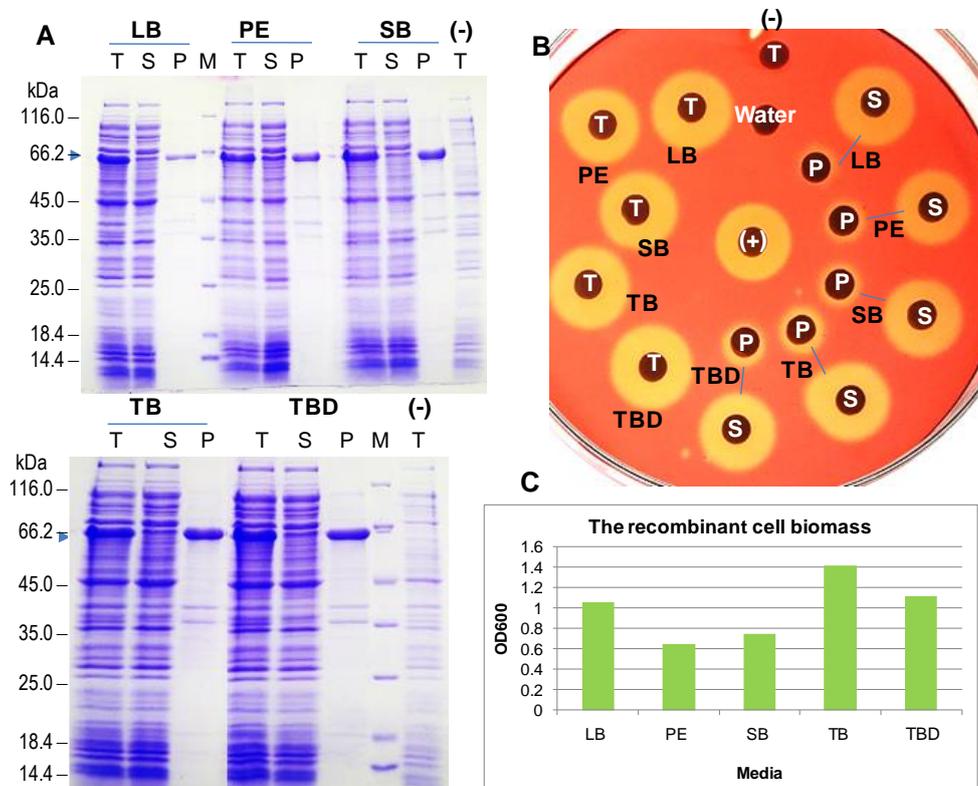


Figure 4. Analysis of EG3 expression in recombinant *E. coli* BL21 strain in different media by SDS-PAGE (A), detection of enzyme activities on CMC 1% (B) determination of cell densities at harvest time at 600 nm wavelength (C). S: soluble fraction; P: pellet/insoluble fraction; (+): cellulase (Sigma); T: total protein (-): total protein of *E. coli* BL21 harboring pET21a (+)

Selection of suitable IPTG concentration for EG3 expression

The transcription of the *eg3* in pET21a(+) was regulated by T7 promoter. This promoter

is induced by the presence of IPTG in the culture medium. High concentrations of IPTG may be toxic to cells, inhibit bacterial growth, but if the concentration too low, it will limit the transcription rate, affecting recombinant

protein expression. Therefore, determining the appropriate concentration of IPTG is essential. In this experiment, after adding IPTG to initiate *eg3* transcription, the cells were culture at 25 °C for 5 hours with expectation to increase EG3 proportion the expressed in soluble form.

The obtained results showed that the recombinant cell biomass in the non-induction sample was the highest (OD₆₀₀=5.75), and the cell biomass in the induced samples was nearly the same (OD₆₀₀ reached 2.2–2.6) at the time of harvest. Out of six concentrations surveyed only a negative control - no addition of IPTG did not synthesize EG3. In the present of

IPTG at different concentrations from 0.05 to 0.9 mM, the same amount of EG3 was expressed. Interestingly, the reduction of temperature from 30 °C to 25 °C, EG3 expressed in the soluble fractions was higher (Figs. 4, 5). We used to reduce culture temperature to 20 °C, however at this temperature, the recombinant *E. coli* strain grew very slowly, the EG3 expression level was reduced. Thus, we will culture the recombinant *E. coli* BL21 strain for expression of EG3 at TB medium complemented with 0.1 mM IPTG at 25 °C for another study. The endoglucanase activity of EG3 in soluble fraction extracted from the cells at OD₆₀₀=10 was 0.22 ± 0.006 (U/ml).

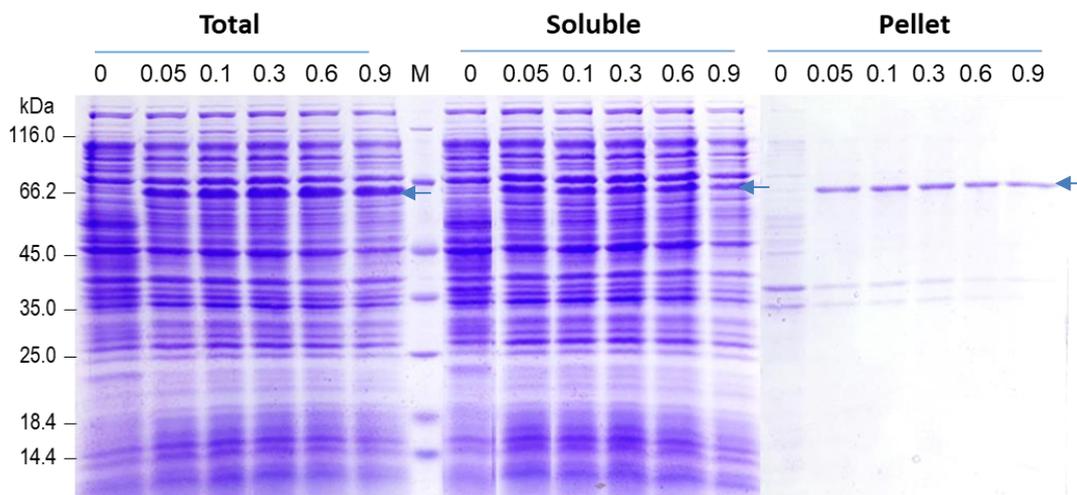


Figure 5. Analysis of EG3 expression in different IPTG concentrations by SDS-PAGE

DISCUSSIONS

Lignocellulose is mainly composed of 40–50% cellulose, 25–30% hemicellulose and lignin. This biomass has been being gained much attention from scientists to converse to high-value added products for the development of bioeconomy and also for the production of biofuel to substitute the depleted fossil fuel. However, the conversion of this biomass into sugars is not effective because of the rigid and complicated lignocellulose structure. So the lignocellulose conversion is required an effective mixture of fast-active lignocellulases. In recent years,

many novel genes coding for steadfast and efficient enzymes were exploited from uncultivable/cultivable microbiome by the Metagenomics approach (Dash & Das, 2018). In this study, the gen GL0361920 coding for endoglucanase GH5-4 was mined from metagenomic DNA data of bacteria in Vietnam goats' rumen. With the target to have effective endoglucanase to make a cocktail enzyme for the conversion of lignocellulose, the codon optimized gene (*eg3*) was investigated to be expressed in *E. coli*.

E. coli is the most popular host for the expression of many recombinant proteins used

in the biopharmaceutical industry. Until now there are many gene-modified *E. coli* strains designed for heterogenous gene expression. The choice of the suitable strains plays a major role in the protein expression, solubility and yield. Every strain has advantages for recombinant protein expression. For example, BL21 strain is deficient in the proteases Lon and OmpT thus this strain has the ability to increase protein stability. Origami strain markedly enhances disulfide bond formation in the cytoplasm. Rosetta strain has more Arg, Pro, Gly, Leu and Ile tRNAs than other strains and C43 strain is one of the high expressing variants. In this study, at 30 °C, EG3 was successfully expressed in *E. coli* BL21 and Origami strains and failed to be expressed in JM109 strain. Very little amount EG3 was expressed in C43 strain. The EG3 activities were observed in all four total proteins from 4 strains BL21, Soluble, Origami and C43 (Fig. 3). The reduction of temperature from 30 °C to 25 °C, a significant part of EG3 was expressed in soluble form. Cultivation at reduced temperatures has been recommended to reduce *in vivo* aggregation of recombinant proteins (Sørensen & Mortensen, 2005), thus temperature effect was seen in many studies improving the solubility of the recombinant protein expressed in *E. coli* (Hasan & Shimizu, 2008; de Groot & Ventura, 2006; Gadgil et al., 2005). Beside temperature, IPTG concentration and media also have been investigated to increase recombinant protein solubility. The higher IPTG concentration used, the higher the transcription rate of recombinant genes. Thus with the protein affecting the host metabolism (for example to inhibit *E. coli* growth), the high IPTG concentration will give a strong negative effect on the growth of the recombinant cells. In this study, the final cell masses at the harvest were different from uninduced (OD600 of 5.7) to induced recombinant cell groups (OD600 of 2.2–2.6). The induction showed a strong effect on the growth of recombinant BL21 strain harboring *eg3* gene. However, the changing of IPTG concentration from 0.05 mM to 0.9 mM, the harvested cells had no significant differences. At 0.05 mM

IPTG, EG3 was expressed with the same levels obtained from the higher concentration. This result is in agreement with another study (Mühlmann et al., 2017). The IPTG concentration of 0.1 mM was chosen to produce EG3, and also seen to be selected for production of other proteins such as leptospiral protein of *Leptospira* (Larentis et al., 2014). At 25 °C, in the TB medium containing 0.1 mM IPTG, the recombinant *E. coli* BL21 strain harboring pET21-*eg3* produced endoglucanase EG3, while about half of EG3 existed in soluble form and exhibit activity hydrolyzing CMC substrate.

CONCLUSION

The gene GL0361920 coding for endoglucanase GH5-4 from metagenomic DNA data of bacteria in goats rumen was codon optimized and expressed successfully in *E. coli* strains. In the TB medium containing 0.01 mM IPTG at 25 °C for 4–5 hours, BL21 strain harboring pET21-*eg3* expressed half of endoglucanase EG3 soluble fraction exhibiting activity to hydrolyse CMC substrate.

Acknowledgements: This study was conducted with the financial support of the National Project code NĐT.50.GER/18 and equipments of National Key Laboratory of Gene Technology, Institute of Biotechnology, VAST, Vietnam.

†The authors contributed equally to this work

REFERENCES

- Asgher M., Ahmad Z., Iqbal H. M.N., 2013. Alkali and enzymatic delignification of sugarcane bagasse to expose cellulose polymers for saccharification and bio-ethanol production. *Ind. Crops Prod.* 44:488–495. <https://doi.org/10.1016/j.indcrop.2012.10.005>
- Carrquiry M. A., Du X., Timilsina G. R., 2011. Second generation biofuels: Economics and policies. *Energy Policy* 39 (7):4222–4234. <https://doi.org/10.1016/j.enpol.2011.04.036>

- Dash H., Das, S., 2018. Molecular methods for studying microorganisms from atypical environments, 1st ed. Elsevier Ltd. <https://doi.org/10.1016/bs.mim.2018.07.005>
- De Groot N. S., Ventura S., 2006. Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* 580 (27):6471–6476. <https://doi.org/10.1016/j.febslet.2006.10.071>
- De Jong E., Higson A., Walsh P., Wellisch M., 2012. Bio-based chemicals: value added products from biorefineries. IEA Bioenergy Task 42 Biorefinery.
- Do T. H., Le N. G., Dao T. K., Nguyen T. M. P., Le T. L., Luu H. L., Nguyen K. H. V., Nguyen V. L., Le L. A., Phung T. N., van Straalen N. M., Roelofs D., Truong N. H., 2018. Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats' rumen. *J. Gen. Appl. Microbiol.* 64(3): 108–116. <https://doi.org/10.2323/jgam.2017.08.004>
- Do T. H., Nguyen T. T., Nguyen T. N., Le Q. G., Nguyen C., Kimura K., Truong N. H., 2014. Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *J. Biosci. Bioeng.* 118(6): 665–671. <https://doi.org/10.1016/j.jbiosc.2014.05.010>
- Duan C. J., Xian L., Zhao G. C., Feng Y., Pang H., Bai X. L., Tang J. L., Ma Q. S., Feng J. X., 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* 107(1): 245–256. <https://doi.org/10.1111/j.1365-2672.2009.04202.x>
- Elifantz H., Waidner L. A., Michelou V. K., Cottrell M. T., Kirchman D. L., 2008. Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean. *FEMS Microbiol. Ecol.* 63(3): 316–327. <https://doi.org/10.1111/j.1574-6941.2007.00429.x>
- Gadgil M., Kapur V., Hu W. S., 2005. Transcriptional response of *Escherichia coli* to temperature shift. *Biotechnol. Prog.* 21(3): 689–699. <https://doi.org/10.1021/bp0496301>
- Glasgow E. M., Kemna E. I., Bingman C. A., Ing N., Deng K., Bianchetti C. M., Takasuka T. E., Northen T. R., Fox B. G., 2020. A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *J. Biol. Chem.* 295(51): 17752–17769. <https://doi.org/10.1074/jbc.RA120.015328>
- Hasan C. M. M., Shimizu K., 2008. Effect of temperature up-shift on fermentation and metabolic characteristics in view of gene expressions in *Escherichia coli*. *Microb. Cell Factories*, 7: 35. <https://doi.org/10.1186/1475-2859-7-35>
- Hess M., Szczyrba A., Egan R., Kim T. W., Chokhawala H., Schroth G., Luo S., Clark D. S., Chen F., Zhang T., Mackie R. I., Pennacchio L. A., Tringe S. G., Visel A., Woyke T., Wang Z., Rubin E. M., 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016): 463–467. <https://doi.org/10.1126/science.1200387>
- Larentis A. L., Nicolau J. F. M. Q., Esteves G. dos S., Vareschini D. T., de Almeida F. V. R., dos Reis M. G., Galler R., Medeiros M. A., 2014. Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in *E. coli* using shaking flasks and microbioreactor. *BMC Res. Notes*, 7(1): 671. <https://doi.org/10.1186/1756-0500-7-671>
- Lorentz Miller G., 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.*, 31(3): 426–428.

- Mühlmann M., Forsten E., Noack S., Büchs J., 2017. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microb. Cell Factories*, 16. <https://doi.org/10.1186/s12934-017-0832-4>
- Ofori-Boateng C., Lee K. T., 2013. Sustainable utilization of oil palm wastes for bioactive phytochemicals for the benefit of the oil palm and nutraceutical industries. *Phytochem. Rev.* 12(1): 173–190. <https://doi.org/10.1007/s11101-013-9270-z>
- Sørensen H. P., Mortensen K. K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Factories*, 4(1): 1. <https://doi.org/10.1186/1475-2859-4-1>