OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. xx No. xx (xxxx) https://doi.org/10.47352/jmans.2774-3047.368
Check for updates
Research Article

Modification of a Gene Cassette to Express Proglargine in *Pichia pastoris*: Elimination of Glu-Ala Repeats, Spacer and C-peptide Sequences

Uli Julia Nasution, Rika Indri Astuti*, Aris Tri Wahyudi, Dudi Hardianto, and Efrida Martius

Received : August 28, 2025	Revised : November 28, 2025	Accepted : February 20, 2026	Online : March 10, 2026

Abstract
The growing global prevalence of diabetes mellitus has sharply increased the demand for insulin and its analogues. *Pichia pastoris* is a well-established system for recombinant pro-insulin and its analogues production. However, conventional gene constructs often include additional sequences, such as Glu-Ala repeats, spacer peptides, and c-peptides that complicate downstream processing and reduce efficiency. This study aimed to construct and express a *proglargine* (PG) gene cassette lacking the Glu-Ala repeats, spacer, and c-peptide in *P. pastoris* GS115 to obtain a uniform PG protein. The recombinant vector propagated in *Escherichia coli* TOP10F', then expressed in *P. pastoris* GS115. Selected transformants were cultivated in YPG medium, then induced with 1% and 2% methanol daily in BMMY. The optimum methanol concentration further evaluated in ½ BSM induction medium. The result demonstrated that optimal PG expression was achieved with 2% methanol induction in BMMY, producing higher levels than those with ½ BSM. Among the transformants, PG.c2 produced the highest PG protein in BMMY medium induced with 2% methanol. Dot-blot analysis confirmed the secretion of PG, and LC-HRMS analysis demonstrated 100% amino acid sequence coverage of PG, confirming the integrity and completeness of the expressed protein. This study presented a newly modified *PG* gene cassette, inserted into pPICZαA vector, to express uniform secreted PG in *P. pastoris* GS115. By simplifying the precursor structure, a more homogeneous precursor product can be obtained, which is expected to simplify purification and also the downstream enzymatic process of PG into mature insulin glargine.

Keywords: dot-blot analysis, Glu-Ala repeats, *Pichia pastoris* GS115, pPICZαA, spacer and c-peptides

1. INTRODUCTION

Diabetes mellitus is a metabolic disorder characterized by an elevated blood sugar level after carbohydrate consumption. The global prevalence of diabetes escalates annually, results in an increase of insulin demand. Insulin recombinant, the first recombinant protein for human use approved by the FDA in 1982, is an essential life-saving therapeutic agent for type 1 and late-stage type 2 diabetics to control blood sugar levels [1][2]. The advent of biotechnology has resulted in the introduction of rapid and long-acting insulin analogues in the 1990s and early 2000s, respectively [3]-[5]. Among these analogues, insulin glargine is one of the most widely prescribed long-acting formulations for type 2 diabetes care, as it provides stable basal insulin levels and helps reduce fluctuations in blood sugar [6]. Glargine were also developed due to the side effects of nocturnal hypoglycaemia of previous analogue insulins and to avoid multiple injections per day [7]. This makes glargine an important focus for improving production efficiency to support therapeutic needs.

Proglargine (PG) plays a pivotal role as the main key raw material in manufacturing glargine. Following that, PG must be enzymatically processed to generate mature glargine. PG is now produced using recombinant microbes, one of which is *Pichia pastoris*. Methylotrophic yeast of *Pichia pastoris* (reclassified as *Komagataella phaffii*) is a well-established expression system for manufacturing recombinant heterologous proteins. It is also granted as GRAS (Generally Recognised as Safe) [8] and attributed with advantageous characteristics as follows: easy to manipulate the genome, has a tightly and strongly regulated AOX1 promoter to drive heterologous protein expression, secretes a small number of endogenous proteins, and easy to scale-up [9]. The α-mating factor (α-MF) from *Saccharomyces cerevisiae* is widely used as a secretion signal peptide in a plasmid for the production of secreted insulin precursor. α-MF has

Table 1. Media used in this research.

List of Media	Purposes
Luria Bertani (LB)	Growth media for *E. coli*TOP10F'
Yeast Peptone Dextrose (YPD)	Growth media for *P. pastoris* GS115
Low Salt Luria Bertani (LSLB)	Growth and vegetative media for Recombinant *E. coli*TOP10F'
Yeast Peptone Dextrose Sorbitol (YPDS)	Selection media for *P. pastoris* GS115 transformant
Yeast Peptone Glycerol (YPG)	Vegetative media for *P. pastoris* GS115 transformant
Buffered Methanol-complex medium (BMMY)	Induction media for *P. pastoris* GS115 transformant
Half of Basal Salt Medium (½ BSM) added with Trace elements (PTM_1) and Histidine	Induction media for *P. pastoris* GS115 transformant

85 amino acids (AA), comprising pre- and pro-peptides, with 19 AA and 66 AA, respectively. At the end of AA sequences, the pro-peptide consists of a dibasic amino acid Kex2 cleavage site of -KR and two or three peptides of Glu-Ala repeat. Protein -encoding gene construct is usually fused into these -KR and Glu-Ala repeats for the expression [10] [11].

Traditionally, a gene cassette for expressing proinsulin and its analogues was designed using Glu-Ala repeats and/or a spacer at N-terminal B-chain. However, these designs have some limitations, such as the fact that some of the secretory precursors still have Glu-Ala repeats and/or a spacer in different lengths attached to the N-terminus of the B-chain, which may become a burden for enzymatically active glargine processing [12]. Furthermore, insulin precursors were commonly designed with c-peptide as a linker between the A and B-chains, which need to be removed enzymatically during insulin maturation processing. The synthetic *PG* gene can be designed without c-peptide since glargine has two specific residual amino acids in the C-terminus of B-chain that can be connected directly to A-chain and cuttable enzymatically. For this reason, this study focuses on the expression of the *PG* gene, by which the gene cassette is designed without Glu-Ala repeats, a spacer, and c-peptide, in *P. pastoris*. This streamlined construct is intended to simplify downstream maturation and provides a foundation for developing more efficient production routes. Importantly, this strategy points toward the feasibility of large-scale and cost-effective manufacturing of insulin glargine for a critical therapeutic in diabetes care.

2. MATERIALS AND METHODS

2.1. Materials

Media used during this study were shown in Table 1, meanwhile the full composition presented in Supplementary data (S-1). All media used for the research were prepared according to the user manual of EasySelect™ *Pichia* Expression Kit (Invitrogen, USA) [13], except for ½ BSM [14] [15]. All agar media were supplemented with 1.5% agar before sterilization. Zeocin was added to the sterilized media at final concentration of 25 µg/mL

for recombinant *E. coli* TOP10F' and 100 µg/mL for *P. pastoris* GS115 transformants.

2.2. Methods

2.2.1. Plasmids and Strains

The *pPICZαA* vector, *E. coli* TOP10F′, and *P. pastoris* GS115 used in this study were from Invitrogen (Table 2). Recombinant *E. coli* TOP10F′ was used for shuttle vector propagation, while *P. pastoris* GS115 transformants served as the host for protein expression. Parent strains of *E. coli* TOP10F′ and *P. pastoris* GS115 were maintained on LB and YPD media, respectively, while recombinant *E. coli* and *P. pastoris* transformants were maintained on LSLB and YPD media, respectively, supplemented by zeocin as needed [13].

2.2.2. Construction of Recombinant Vector

A codon-optimized nucleotide sequence encoding secretory PG for expression in *P. pastoris* GS115 was synthesized chemically by GenScript. In this research, the codon adaptation index (CAI) and GC content values were 0.84 and 42.65%, respectively. The gene cassette was designed by placing a four amino acid (LEKR), containing the *Xho*I site (LE) and the Kex2 protease cleavage site (KR), downstream of the α-mating factor signal sequence [12][13]. This was followed by the PG B-chain (32 amino acids) and A-chain (21 amino acids) arranged consecutively without an intervening C-peptide. The cassette was cloned in-frame into pPICZαA between *Xho*I and *Not*I to generate the expression construct pPICZαA/PG. A schematic representation of the construct is shown in Figure 1.

2.2.3. Transformation of pPICZαA/PG to Escherichia coli TOP10F'

Competent *E. coli* TOP10F' cells were prepared using $CaCl_2$ method and plasmid transformation was performed using a heat-shock technique according to Rahimzadeh *et al*. [16] and Chang *et al*. [17], respectively. Screening recombinant *E. coli* TOP10F' was performed by growing the cells onto an LSLB medium with zeocin (25 µg/mL) overnight at 37 °C. Parent plasmid pPICZαA was also transformed into *E.* coli TOP10F' as a control. Colonies growing under zeocin selection medium were further confirmed by plasmid sequencing using forward and reverse primers as in Table 2.

2.2.4. Analysis of Nucleotide Sequence of PG Gene

Nucleotide sequence of *PG* gene in the plasmid was analysed before transformed into *P. pastoris* GS115 to verify there was no mutations had occurred during plasmid propagation. The plasmid pPICZαA/PG was propagated in *E. coli* TOP10F', and the recombinant cells were cultured prior to plasmid extraction using a modified Geneaid plasmid isolation procedure as described by Pronobis [18]. The concentration and purity of the isolated plasmid DNA were determined using a NanoDrop™ spectrophotometer. The nucleotide sequence of PG in the isolated pPICZαA/PG was sequenced using AOX1 forward and reverse primers as indicated in Table 2. PCR conditions were conducted as follows: pre-denaturation (94 °C; 2 min), denaturation (98 °C; 10 s), annealing (50 °C; 30 s), extension (68 °C; 90 s), and final extension (68 °C; 7 min). The resulting amplicons

Table 2. Plasmids, microbes and primers used in this study.

Plasmids	
pPICZαA	Parent plasmid
pPICZαA/PG	Recombinant pPICZαA harbouring PG gene
Microbes	
E. coli TOP10F'	Parent host
P. pastoris GS115	Parent host
Primers	
AOX1 forward primer	5'-GACTGGTTCCAATTGACAAGC-3'
AOX1 reverse primer	5'-GCAAATGGCATTCTGACATCC-3'

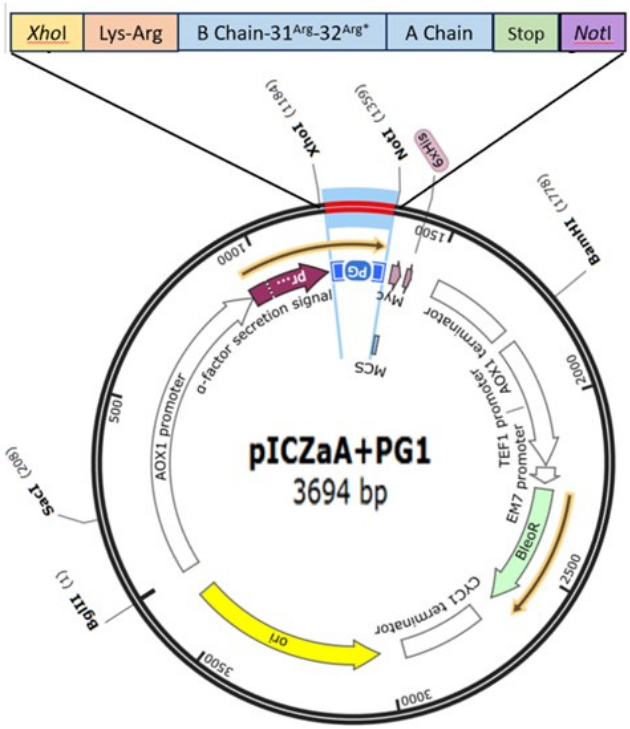


Figure 1. Schematic map of the recombinant yeast-integrative plasmid pPICZαA/PG for PG secretory expression in *Pichia pastoris* driven by the methanol-inducible promoter *AOX*1 used in this study.

were then subjected to sequencing by the First Base company.

2.2.5. Transformation of Plasmid to Pichia pastoris

Competent cells (*P. pastoris* GS115) were prepared using a chemical treatment method and plasmid transformation was carried out using electroporator as described by Cregg *et al.* [19] and Lin-Cereghino *et al.* [20]. Plasmid pPICZαA/PG was linearized with *Sac*I and mixed with 80 µL of competent cells and subjected to electroporation using a Bio-Rad electroporator (1500 V, 200 Ω, 25 µF). Immediately after pulsing, 0.5 mL of cold sorbitol was added, and the cells were incubated at 30 °C for 1 h for recovery, and then 0.5 mL YPD was added into the culture, followed by overnight cultivation at 30 °C with shaking. Transformants were selected on YPDS agar containing zeocin (100 µg/mL) and incubated at 30 °C for 2–4 days. The parental plasmid pPICZαA was transformed in parallel as a negative control.

2.2.6. Phenotypic Analysis of Recombinant Pichia pastoris G115

Mut phenotypic analysis is essential to determine methanol concentration for induction during PG expression since recombination can occur in various ways, generating clonal variation that may impact expression levels. The methanol utilization (Mut) phenotype categorized into three types: Mut^+ strains are able to utilize methanol quickly, whereas Mut^s means that the strain uses methanol more slowly, and Mut^- strains are unable to grow on methanol as a sole carbon source. In this study, three larger colonies, as shown in Table 3, were selected to confirm the Mut phenotype.

Mut phenotype was determined by colony PCR with 5' *AOX*1 forward primer paired with 3' *AOX*1 reverse primer as shown in Table 2, and PCR condition was outlined in section 2.2.4. Colonies that yielded two amplicons of approximately 690 bp and 2.2 kbp were identified as Mut^+ phenotypes, while those showing only a single 690 bp amplicon were classified as Mut^s transformants. Subsequently, the amplicons required additional sequencing to confirm correct integration and absence of mutations. The nucleotide sequence of the PG gene was then sequenced by First Base company.

2.2.7. Shake Flask PG Expression

Three larger Mut^+ colonies were selected for PG

expression screening as shown in Table 3. This PG expression was performed in two phases: vegetative and induction phases. During the vegetative growth, *P. pastoris* harbouring the PG gene was subcultured in fresh YPD agar medium containing 100 µg/mL zeocin and incubated at 30 ℃ for 72 h. This single colony of *P. pastoris* was then inoculated into 50 mL of YPG medium supplemented with 100 µg/mL zeocin, and the culture was incubated in a rotary shaker at 250 rpm and 30 ℃ for 36 h. The cells were then harvested by centrifugation before transferred into induction medium.

The pellet cells were resuspended in either BMMY or ½ BSM medium to an optical density at 600 nm ($OD\lambda_{600nm}$) of 15. The induction process involved the addition of methanol at two concentrations (1% and 2%) to BMMY medium, and methanol was added daily over a 72-h induction period. These conditions were compared to determine the most suitable methanol concentration for optimal expression of PG. Induction with 2% methanol in BMMY resulted in higher PG expression compared to 1% methanol. This optimum methanol concentration determined from this experiment was subsequently applied to another induction medium, ½ BSM.

2.2.8. SDS-PAGE Analysis of PG

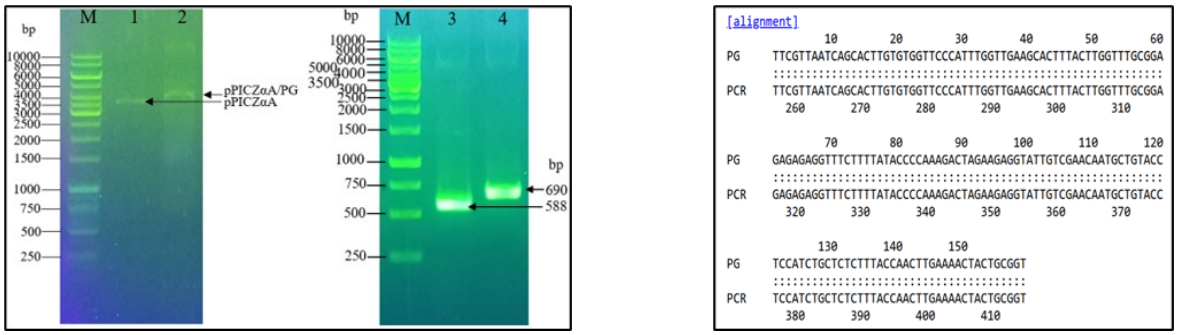
PG analysis in the cell-free supernatant were undertaken using tricine SDS-PAGE with 10% polyacrylamide gel in a tricine buffer system [21], with commercial glargine (Lantus® Solostar®) as a positive control for staining detection. Protein bands were detected using Coomassie Brilliant Blue R-250 or silver staining kit.

2.2.9. Dot-blot Analysis of PG

PG in the supernatant from the most suitable induction condition was characterized using dot-blot analysis [22]. Dot-blot analysis confirmed the presence of PG in the culture supernatant. The supernatant was applied into a nitrocellulose membrane (0.45 µm pore size). Indirect immunodetection was performed using a primary antibody (Mouse anti-Human Insulin Monoclonal Antibody, MyBioresource) followed by a secondary antibody (anti-mouse IgG (H+L) AP conjugate, Promega). The PG indirectly binding to the secondary antibody was detected with NBT/BCIP

Table 3. Transformants used in this study.

Transformants
E. coli TOP10F'/pPICZαA
E. coli TOP10F' transformed with the parent plasmid
E. coli TOP10F'/pPICZαA/PG
E. coli TOP10F' transformed with pPICZαA/PG
P. pastoris GS115/ pPICZαA (A)
P. pastoris GS115 transformed with pPICZαA (negative control)
P. pastoris GS115/pPICZαA/PG.c1 (PG.c1)
P. pastoris GS115 transformed with pPICZαA, designated as clone1
P. pastoris GS115/pPICZαA/PG.c2 (PG.c2)
P. pastoris GS115 transformed with pPICZαA- designated as clone2
P. pastoris GS115/pPICZαA/PG.c3 (PG.c3)
P. pastoris GS115 transformed with pPICZαA- designated as clone3



(**a**) (**b**)

Figure 2. Verification of the recombinant plasmid and *PG* gene sequence (a) Agarose gel electrophoresis analysis of plasmid DNA. Lane M, 1 kb DNA ladder; Lane 2, parent plasmid (pPICZαA); Lane 2, recombinant plasmid (pPICZαA/PG); Lane 3, DNA band without *PG* gene; Lane 4, DNA band contained *PG* gene; (b) Nucleotide sequence alignment of the synthesized *PG* gene (PG line) with the sequence obtained from extracted recombinant plasmid PCR (PCR line).

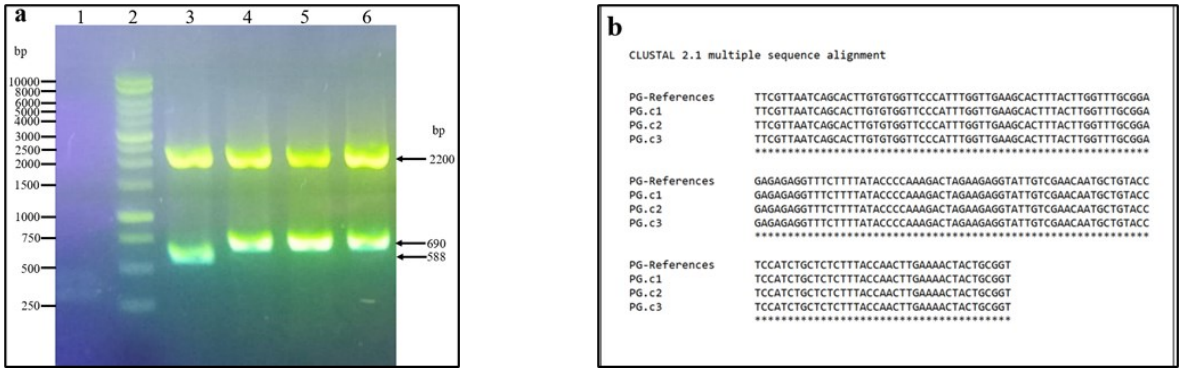


Figure 3. Phenotypic characterisation and *Pichia's* genome-integrated *PG* gene sequence analysis (a) Three selected *P. pastoris* transformants harbouring *PG* gene. Lane 1, negative control of PCR process; lane 2, 1 kb DNA ladder; lane 3, *P. pastoris* GS115 transformed by pPICZαA without *PG* gene (588 bp), lanes 4 to 6, *P. pastoris* transformants Mut$^+$ carrying the *PG* gene (690 bp) and *AOX1* gene (2200bp); (b) ClustalW multiple sequence alignment of PG-cassette sequences amplified from genome's transformants compared to the synthetic *PG* sequence.

solution for 5–15 min.

2.2.10. Peptide Mapping Analysis

Peptide mapping of the secreted PG was performed to determine amino acid sequences and verify the primary structure using liquid chromatography–high-resolution mass spectrometry (LC-HRMS) [23], conducted by Corpora Science. LC-HRMS confirmed 100% sequence coverage of PG, verifying its primary structure without mutations or truncations. Liquid chromatography was performed on a Thermo Scientific Vanquish Horizon UHPLC system equipped with a PepMap 100 C_{18} column (150 mm × 1 mm, 3 µm). The column temperature was maintained at 40 °C. The reverse-phase LC eluent was analyzed using a Thermo Scientific™ Orbitrap™ Exploris 240 in Full MS positive ion mode, with ionization via OptaMax™ NG Heated Electrospray Ionization (H-ESI). Data analysis was performed using Proteome Discoverer™ 2.5, SequestHT, and the UniProt database.

3. RESULTS AND DISCUSSIONS

3.1. Plasmid Characterization and PG Gene Sequence Verification in Escherichia coli TOP10F'

Characterization parent and recombinant plasmids, pPICZαA and pPICZαA/PG, respectively, transformed into *E. coli* TOP10F' was important to confirm the intended sequence modification cassette was correctly cloned into the

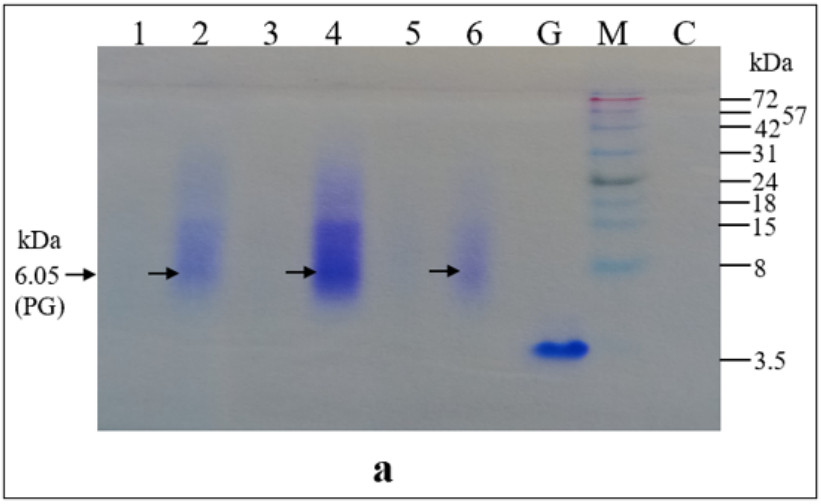


Figure legend picture **a** :
1-2 : supernatant from PG.c1 transformant culture
3-4 : supernatant from PG.c2 transformant culture
5-6: supernatant from PG.c3 transformant culture
G : Positive control (Glargine commercial)
M : Protein Ladder
C : Negative control
1, 3, 5 : cultures induced with 1% methanol in 3 days
2, 4, 6 : cultures induced with 2% methanol in 3 days

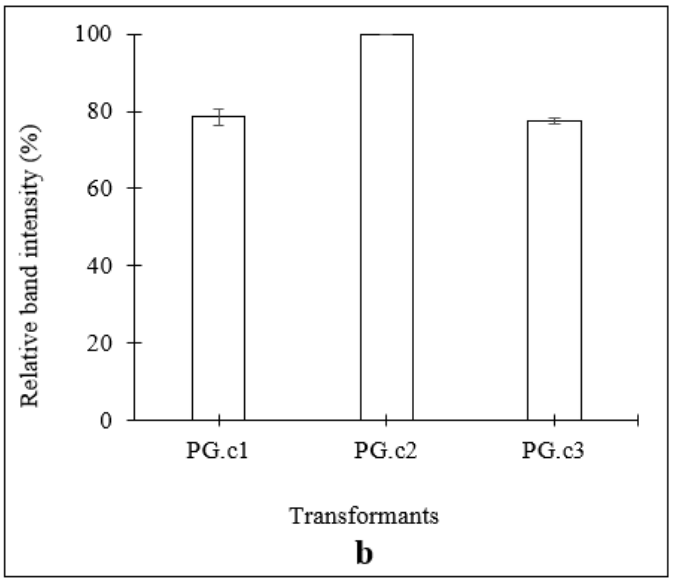


Figure legend picture **b** :
1 : PG.c1 = ~79%
2 : PG.c2 = ~100%
3 : PG.c3 = ~78%
G : Positive control (Glargine commercial)

Figure 4. Cell-free supernatant analysis (a) SDS-PAGE gel analysis of PG expression in BMMY medium induced with 1% or 2% methanol for 3 days; (b) Relative band intensity quantified using ImageJ.

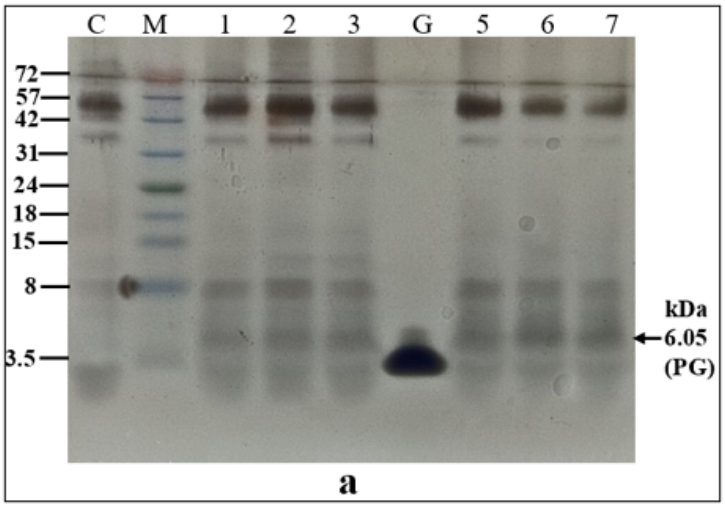


Figure legend picture **a** :
C : Negative control
L : Protein Ladder
1 : Supernatant from PG.c3 transformant
2 : Supernatant from PG.c2 transformant
3 : Supernatant from PG.c1 transformant
G : Positive control (Glargine commercial)
4 : Supernatant from PG.c1 transformant
5 : Supernatant from PG.c2 transformant
6 : Supernatant from PG.c3 transformant
1-6 : Cultures induced with 2 % methanol in 3 davs

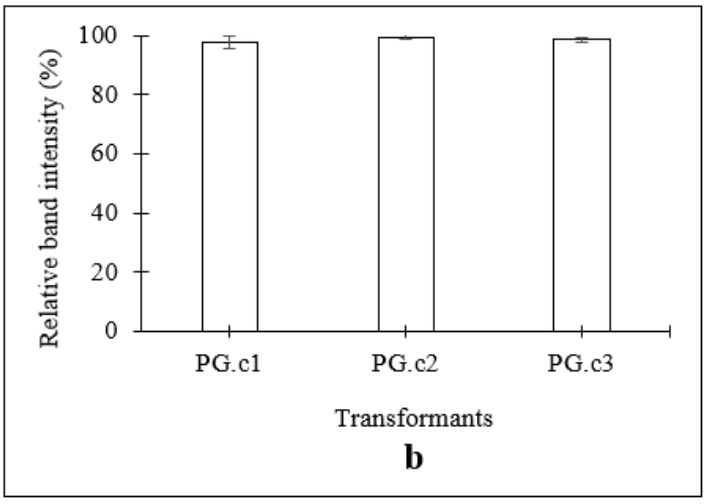


Figure legend picture **b**:
PG.c1 = ~ 98%
PG.c2 = ~ 100%
PG.c3 = ~ 99%

Figure 5. Cell-free supernatant analysis (a) SDS-PAGE gel analysis of PG expression in ½ BSM medium induced with 2% methanol during 3 days; (b) Relative band intensity quantified using ImageJ.

construct. The 1% agarose gel electrophoresis result confirmed a recombinant plasmid size of 3694 bp, consistent with insertion of the *PG* gene (Figure 2 (a)). Furthermore, sequencing analysis was performed to ensure the cloning process was executed correctly and verify the accuracy of the inserted *PG* gene. The isolated pPICZαA/PG was applied as a template to amplify *PG* gene's nucleotide sequence with *AOX1* forward and reverse primers (as shown in Table 2) through the PCR method to analyse mutations during plasmid propagation in *E. coli* TOP10F'. The result of sequence analysis confirmed no mutations occurred (Figure 2(b)). Sequence analysis result of the *PG* gene cloned into the pPICZαA plasmid under the control of the AOX1 promoter also confirmed the successful elimination of Glu-Ala repeats and the spacer region. This verification demonstrated that the modified *PG* gene cassette was accurately integrated at the *Xho*I and *Not*I multiple cloning sites.

In this research, codon optimization was performed on the PG encoding gene since some studies demonstrated the increasing translational efficiency of the recombinant protein. However, this codon optimization has not been further validated through subsequent research**.** Some studies have investigated the role of codon optimization by replacing rare codons with those favoured by the host to increase the heterologous protein expressed in *P. pastoris*. Jie *et al*. reported that the *xylanase* gene from *Thermotoga maritima* was codon-optimized by substituting rare codons such as TCG (Ser), CTC (Leu), AGC (Ser), and GCG (Ala), resulting in a 2.8-fold increase in xylanase expression [24].

3.2. Phenotypic Characterization and PG Nucleotide Sequence Confirmation in Recombinant P. pastoris GS115's Genome

Following the transformation, three selected transformants were chosen as shown in Table 2, and

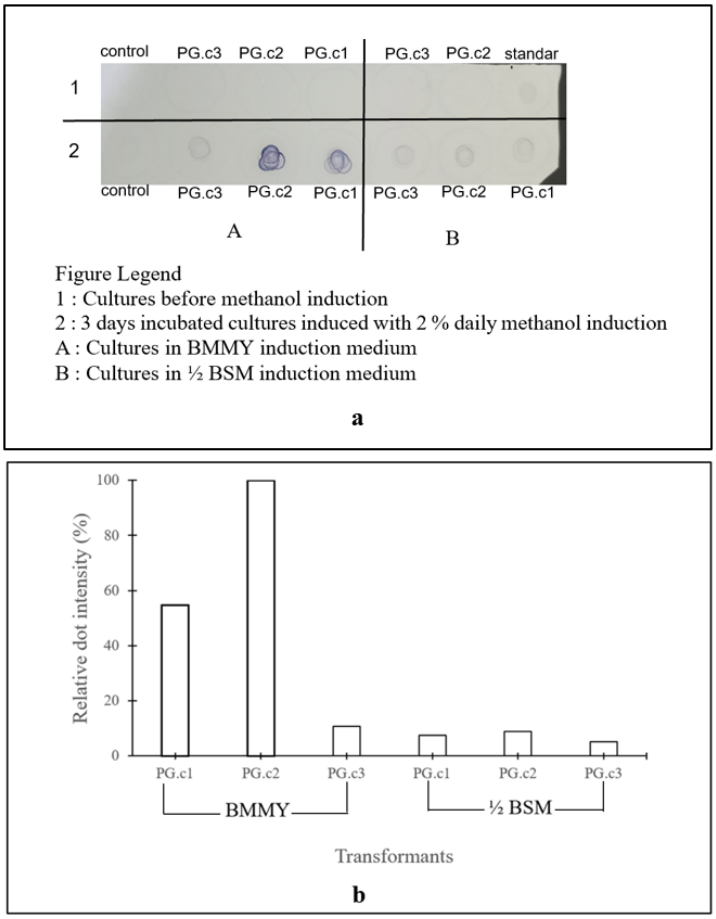


Figure 6. Dot-blot analysis of cell-free supernatant (a) before (1) and after (2) 2% daily methanol induction for 3 days, growth in medium BMMY (A) and ½ BSM (B); (b) Relative dot-blot intensity quantified using ImageJ.

Table 4. Peptide sequence of Proglagine.

Fragment	Peptide sequence	Position of Amino Acid	Retention Time (min)	Hidro-phobicity*	Coverage
I	LYLVCGERGF	B15–B24	10.59	29.10	100%
II	FVNQHLCGSHLVEAL	B1–B15	13.46	35.87	100%
III	GFFYTPKTRRGIVEQCCTSICSLYQLE	B23–A17	19.44	42.31	100%
IV	EALYLVCGERGFFYTPKTRRGIVEQCCTSICSLYQLENYCG	B13–A21	25.86	49.97	100%

*https://www.thermofisher.com/id/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html

their phenotypes were subsequently characterized. Phenotypic analysis was conducted to determine the methanol utilization (Mut) phenotype of recombinant *P. pastoris* strains. The AOX1 promoter gene is naturally present in the *P. pastoris* GS115 genome (2.2 kb) and is also included in the expression plasmid as a 690 bp fragment. In Mut^+ strains, both the genomic AOX1 and the plasmid-derived fragment are present (Figure 3(a)), allowing detection of the 2.2 kb AOX1 gene. In contrast, in Mut^S strains, the plasmid integrates into and disrupts the native AOX1 locus, preventing amplification of the 2.2 kb fragment by PCR. Figure 3(a) revealed that all three selected transformants had Mut^+ phenotype. The 690 bp gene in Mut^+ transformants was then further sequenced to analyse the nucleotide sequence of the *PG* gene. The results demonstrated that the *PG* gene sequence of the three selected transformants was identical to the initial *PG* gene construction as revealed in Figure 3(b).

In this work, the phenotypic characterization of the yeast-transformants was carried out to determine the methanol concentration added during the expression process. The selected transformants were coded as PG.c1, PG.c2, and PG.c3. The induction medium and methanol inducer play a significant role in the PG expression process. Low methanol concentrations may not be sufficient to initiate transcription, and proteolytic degradation may potentially lead to cell starvation; however, excessive methanol concentrations can be harmful to cells [25].

3.3. PG Expression

The PG expression was evaluated by the SDS-PAGE. The theoretical molecular weight of the PG protein was predicted to be approximately 6.05 kDa and the corresponding band with similar size was observed on the SDS-PAGE gel. The result revealed that PG expression was not detectable under 1% methanol induction, while 2% methanol resulted in clear expression in BMMY medium, with PG.c2 showing the highest expression as shown in Figure 4(a). The 2% methanol condition was then applied to evaluate PG expression in ½ BSM medium. Unlike BMMY, PG expression in ½ BSM could only be detected using silver staining, indicating lower PG accumulation compared to

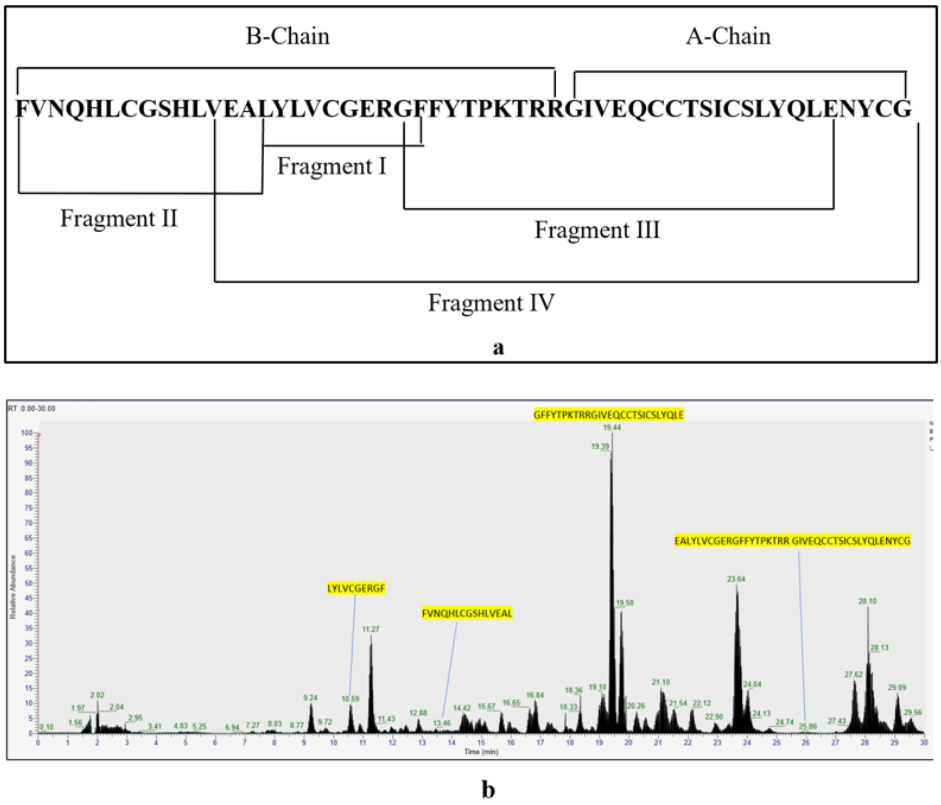


Figure 7. Peptide sequence of PG analyzed by LC-HRMS (a) Amino acid sequence of PG; (b) chromatogram result of the PG analysis.

BMMY (Figure 5(a)).

Band-intensity quantification using ImageJ was performed to compare PG expression among the three selected transformants (PG.c1, PG.c2, and PG.c3). The percentages presented represent **r**elative band intensities normalized to the strongest PG band among the samples (set as 100%). In BMMY induced with 2% methanol, PG.c2 showed the highest expression, while PG.c1 and PG.c3 produced 79% and 78%, respectively (Figure 4(b)). A similar pattern was observed in ½ BSM induced with 2% methanol, where PG.c2 again yielded the highest PG expression (Figure 5(b)).

Induction with 2% methanol resulted in higher PG production than with 1%, suggesting that this concentration more effectively activates the AOX1 promoter to drive *PG* gene transcription and provides sufficient metabolic energy (NADPH and ATP) to support protein biosynthesis [26], while still remaining non-toxic to the cells [27]. Other studies regarding the methanol concentration for induction to express heterologous protein showed similar results. A 2% methanol induction was considered the optimal condition to express recombinant protein xylanase [28] and the multistage tuberculocis vaccine [29] using recombinant *P. pastoris*. PG expression in ½ BSM medium was lower than BMMY. The result was similar with Zitkus *et al*. in which recombinant protein expression was lower in BSM than nutrient (YEPM) media. It is likely because ½ BSM is a minimal medium, requiring cells to synthesize all necessary metabolic intermediates. In contrast, richer media (like BMMY) provides these components that readily for cell growth and product formation, thereby reducing the metabolic burden on the cells [30][31].

3.4. Dot-blot Analysis of PG

Following the expression process, a dot-blot analysis using an anti-insulin monoclonal antibody was performed to detect the presence of PG in the culture supernatants specifically. Dot-blot intensity was semi quantified using ImageJ and the percentages presented represent relative band intensities normalized to the strongest dot-blot signal among the samples (set as 100%). The results revealed a stronger dot signal for PG.c2 compared

to weaker signals for PG.c1 and PG.c3 indicating the highest PG expression in PG.c2 transformant from BMMY medium supernatant induced with 2% methanol, consistent with SDS-PAGE results (Figures 6(a) and 6(b)).

3.5. LC-HRMS Analysis

Peptide mapping of the PG protein was performed through LC-HRMS analysis. The cell-free supernatant was subjected to chromatographic separation using the LC-HRMS method, revealing four distinct peptide fragments eluted at different retention times as shown in Figure 7. They were labelled as fragments I–IV in order of increasing molecular weight corresponding to fragment I (amino acid sequences B15–B24), fragment II (B1–B15), fragment III (B23–A17) and fragment IV (B13–A21) with retention times of 10.59, 13.46, 19.44, and 25.86 min, respectively, as shown in Table 4. The observed retention times suggested differences in the physicochemical properties of the peptide fragments, most probably their molecular weight and hydrophobicity [32]. Peptide fragment I, eluted earliest (10.59 min), was smaller in size with theoretical hydrophobicity value only 29.10 and predicted to be more polar, resulting in weaker interactions with stationary phase of the column. In contrast, peptide fragment IV, which showed the longest retention time (25.86 min), have large molecular weight and the highest theoretical hydrophobicity value of 25.86. These differences may reflect variations in amino acid composition and sequence among the fragments, which in turn affected their retention time.

The peptide mapping sequences identified from these fragments collectively achieved 100% sequence coverage of the PG, confirming the integrity and completeness of the peptide mapping process. This LC-HRMS was also verifying its primary structure without mutations or truncations. Detailed analysis revealed that Glu-Ala repeats and spacer were not detected in the N-terminal of the first amino acid of B-chain PG and the full sequence was successfully mapped into overlapping peptide fragments as in Figures 7(a) and 7(b).

The peptide mapping result suggested that this gene cassette design would have an effect on the industrial scheme in which the downstream bio-enzymatic processing (using Kex2 enzyme) would yield less protein product variants compare to that have design with Glu-Ala repeats and spacer on the upstream region of B-chain of PG DNA and c-peptide that links B- and A-chains. This has a potency to simplify purification process after bio-enzymatic process. A schematic diagram contrasting the traditional construct of precursor form gene cassette with the modified version was illustrated in Figure 8.

By simplifying the precursors, this modification reduces the need for extensive enzymatic processing, thereby potentially lowering production costs and improving yield consistency. From a clinical perspective, this is highly relevant, as insulin glargine remains a most given therapy for type 1 and advanced type 2 diabetes mellitus. Enhancing the efficiency and scalability of PG manufacturing through optimized genetic constructs can contribute to broader accessibility and affordability, particularly in regions where the cost of long-acting insulin analogues limits patient access.

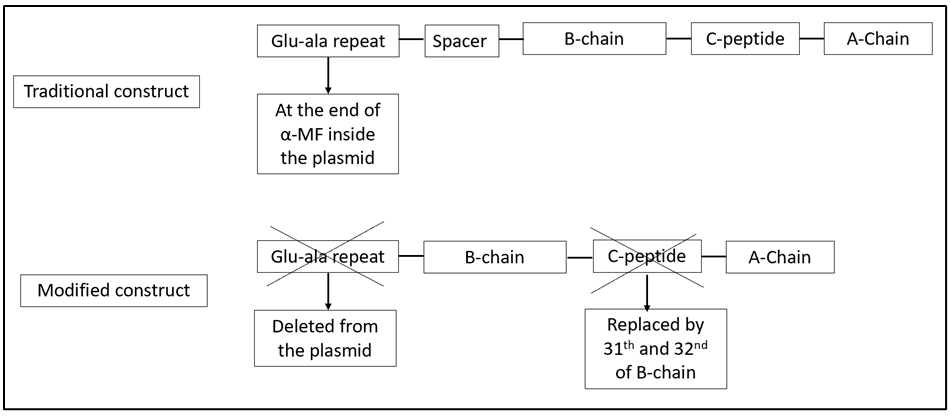


Figure 8. Traditional and modified precursor form of gene cassette construct of insulin and its analogues.

4. CONCLUSIONS

A newly modified *PG* gene cassette inserted into pPICZαA, lacking the Glu–Ala repeats, spacer, and C-peptide regions, was successfully expressed as a secreted PG protein in *Pichia pastoris* GS115. Optimal PG expression was obtained with 2% methanol induction in BMMY, yielding higher levels than those in ½ BSM. Among the transformants, PG.c2 exhibited the highest expression level when induced daily in BMMY medium with 2% methanol. LC-HRMS analysis further confirmed the integrity of the expressed PG product, demonstrating 100% amino acid sequence coverage with phenylalanine (F) as the native N-terminus of the B-chain. Collectively, these findings demonstrated a simplified and efficient system for expressing a more homogeneous PG with promising potential to facilitate purification and reduce variability during further downstream enzymatic maturation of insulin glargine.

AUTHOR INFORMATION

Corresponding Author

Rika Indri Astuti — Department of Biology, IPB University, Bogor-16680 (Indonesia); Biotech Center, IPB University, Bogor-16680 (Indonesia);
orcid.org/0000-0003-1561-6943
Email: rikaindriastuti@apps.ipb.ac.id

Authors

Uli Julia Nasution — Doctoral Program in Microbiology, Department of Biology, IPB University, Bogor-16680 (Indonesia); Research Centre for Applied Microbiology and Bureau of Organization and Human Resources, National Research and Innovation Agency, Tangerang Selatan-15311 (Indonesia);
orcid.org/0000-0003-1547-7883

Aris Tri Wahyudi — Department of Biology, IPB University, Bogor-16680 (Indonesia);
orcid.org/0000-0001-7837-9557

Dudi Hardianto — Research Centre for Drug and Vaccine, National Research and Innovation Agency, Tangerang Selatan-15311 (Indonesia);
orcid.org/0000-0002-5205-1980

Efrida Martius — Research Centre for Drug and Vaccine, National Research and Innovation Agency, Tangerang Selatan-15311 (Indonesia);
orcid.org/0000-0003-4116-7897

Author Contributions

Conceptualization, and Methodology, U. J. N., R. I. A., A. T. W., and D. H.; Software, U. J. N., E. M.; Validation, U. J. N., R. I. A. and E. M.; Formal Analysis, U. J. N., and R. I. A.; Investigation, Resources, Data Curation, Visualization, Writing – Original Draft Preparation, Project Administration, and Funding Acquisition, U. J. N.; Writing – Review & Editing, U. J. N., R. I. A., A. T. W., D. H., and E. M.; Supervision, R. I. A., A. T. W. and D. H.

Conflicts of Interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

This work was partly supported by Indonesia Endowment Fund for Education Agency (*Lembaga Pengelola Dana Pendidikan/LPDP*) from the Ministry of Finance, Republic of Indonesia to U.J.N (KET-2679/LPDP.4/2022 revised by LOG-6581/LPDP.3/2024). The authors also gratefully acknowledge the Deputy for Infrastructure Research and Innovation, National Research and Innovation Agency of Indonesia, for providing access to the laboratory facilities.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, and A. Villaverde. (2009). "Microbial Factories for Recombinant Pharmaceuticals". *Microbial Cell Factories.* **8** : 1-8. 10.1186/1475-2859-8-17.

[2] G. Walsh. (2005). "Therapeutic Insulins and Their Large-Scale Manufacture". *Applied Microbiology and Biotechnology.* **67** (2): 151-159. 10.1007/s00253-004-1809-x.

[3] G. B. Bolli, A. Y. Y. Cheng, and D. R. Owens. (2022). "Insulin: Evolution of Insulin Formulations and Their Application in Clinical Practice Over 100 Years". *Acta Diabetologica.* **59** (9): 1129-1144. 10.1007/s00592-022-01938-4.

[4] I. B. Hirsch, R. Juneja, J. M. Beals, C. J. Antalis, and E. E. Wright. (2020). "The Evolution of Insulin and How It Informs Therapy and Treatment Choices". *Endocrine Reviews.* **41** (5): 733-755. 10.1210/endrev/bnaa015.

[5] C. Mathieu, P. J. Martens, and R. Vangoitsenhoven. (2021). "One Hundred Years of Insulin Therapy". *Nature Reviews Endocrinology.* **17** (12): 715-725. 10.1038/s41574-021-00542-w.

[6] S. Sarkar, J. Heyward, G. C. Alexander, and R. R. Kalyani. (2021). "Trends in Insulin Types and Devices Used by Adults with Type 2 Diabetes in the United States, 2016 to 2020". *JAMA Network Open.* **4** (10): 1-10. 10.1001/jamanetworkopen.2021.28782.

[7] E. K. Sims, A. L. J. Carr, R. A. Oram, L. A. DiMeglio, and C. Evans-Molina. (2021). "100 Years of Insulin: Celebrating the Past, Present and Future of Diabetes Therapy". *Nature Medicine.* **27** (7): 1154-1164. 10.1038/s41591-021-01418-2.

[8] M. Ahmad, M. Hirz, H. Pichler, and H. Schwab. (2014). "Protein Expression in Pichia Pastoris: Recent Achievements and Perspectives for Heterologous Protein Production". *Applied Microbiology and Biotechnology.* **98** (12): 5301-5317. 10.1007/s00253-014-5732-5.

[9] H. Raschmanová, A. Weninger, Z. Knejzlík, K. Melzoch, and K. Kovar. (2021). "Engineering of the Unfolded Protein Response Pathway in Pichia Pastoris: Enhancing Production of Secreted Recombinant Proteins". *Applied Microbiology and Biotechnology.* **105** (11): 4397-4414. 10.1007/s00253-021-11336-5.

[10] M. Merkaš, N. Grujicic, M. Geier, A. Glieder, and A. E. Augustin. (2025). "The MFα Signal Sequence in Yeast-Based Protein Secretion: Challenges and Innovations". *Applied Microbiology and Biotechnology.* **109** : 138-157. 10.1007/s00253-025-13532-z.

[11] C. Zou, L. Lu, S. Wang, C. Zhang, X. Chen, Y. Lin, and Y. Huang. (2022). "The α-Mating Factor Secretion Signals and Endogenous Signal Peptides for Recombinant Protein Secretion in Komagataella phaffii". *Biotechnology for Biofuels and Bioproducts.* **15** (1): 1-10. 10.1186/s13068-022-02243-6.

[12] S. Polez, D. Origi, S. Zahariev, C. Guarnaccia, S. G. Tisminetzky, N. Skoko, and M. Baralle. (2016). "A Simplified and Efficient Process for Insulin Production in Pichia Pastoris". *PLoS One.* **11** (12): 1-15. 10.1371/journal.pone.0167207.

[13] I. Palanikumar, S. Katla, N. Tahara, M. Yui, R. Zhang, A. Ebihara, and S. Sivaprakasam. (2019). "Heterologous expression, purification, and functional characterization of recombinant ovine angiotensinogen in the methylotrophic yeast Pichia pastoris". *Biotechnology Progress.* **35** (5): e2866. 10.1002/btpr.2866.

[14] C. H. Luna-Flores, Y. Weng, A. Wang, X. Chen, B. Peng, C. Zhao, L. Navone, J. Von Hellens, and R. E. Speight. (2023). "Improving Phytase Production in Pichia Pastoris Fermentations Through De-Repression and Methanol Induction Optimization". *Biotechnology and Bioengineering.* **120** : 3276-3287. 10.1002/bit.28510.

[15] J. Wu, G. Gong, S. Han, W. Zhang, Y. Hu, and L. Xie. (2019). "Expression, Purification, and Characterization of the Degludec Precursor DesB30". *Protein Expression and Purification.* **161** : 28-39. 10.1016/j.pep.2019.04.010.

[16] M. Rahimzadeh, M. Sadeghizadeh, F. Najafi, S. S. Arab, and H. Mobasheri. (2016). "Impact of Heat Shock Step on Bacterial Transformation Efficiency". *Molecular Biology Research Communications.* **5** (4): 257-261.

[17] A. Y. Chang, V. W. Chau, J. A. Landas, and Yvonne. (2017). "Preparation of Calcium Competent Escherichia coli and Heat-Shock Transformation". *Journal of Experimental*

Microbiology and Immunology Methods. **1** : 222-225.

[18] M. I. Pronobis, N. Deuitch, and M. Peifer. (2016). "The Miraprep: A Protocol That Uses a Miniprep Kit and Provides Maxiprep Yields". *PLoS One.* **11** (8): 1-12. 10.1371/journal.pone.0160509.

[19] J. M. Cregg, I. Tolstorukov, A. Kusari, A. J. Sunga, K. Madden, and T. Chappell. (2018). "Expression of Recombinant Genes in the Yeast Pichia pastoris". *Current Protocols in Essential Laboratory Techniques.* **17** (1): 1-17. 10.1002/cpet.25.

[20] J. Lin-Cereghino, C. A. Naranjo, and G. P. Lin-Cereghino. (2022). In: "Methods in Molecular Biology". New York: Humana Press. 113-120. 10.1007/978-1-0716-2399-2_7.

[21] S. R. Haider, H. J. Reid, and B. L. Sharp. (2019). In: "Electrophoretic Separation of Proteins". New York: Humana Press. 151-160. 10.1007/978-1-4939-8793-1_15.

[22] V. Mishra. (2022). "Dot-Blotting: A Quick Method for Expression Analysis of Recombinant Proteins". *Current Protocols.* **2** (9): 1-19. 10.1002/cpz1.546.

[23] J. C. Hsu, B. Y. Lai, W. L. Hsieh, D. Y. Wang, D. Y. C. Shih, and C. F. Lo. (2012). "Identification of Recombinant Insulin Analogues by Peptide Mapping Method". *Journal of Food and Drug Analysis.* **20** (4): 957-962. 10.6227/jfda.2012200427.

[24] H. Jia, G. Fan, Q. Yan, Y. Liu, Y. Yan, and Z. Jiang. (2012). "High-Level Expression of a Hyperthermostable Thermotoga maritima Xylanase in Pichia pastoris by Codon Optimization". *Journal of Molecular Catalysis B: Enzymatic.* **78** : 72-77. 10.1016/j.molcatb.2012.02.009.

[25] M. Karbalaei, S. A. Rezaee, and H. Farsiani. (2020). "Pichia pastoris: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins". *Journal of Cellular Physiology.* **235** (9): 5867-5881. 10.1002/jcp.29583.

[26] H. Shi, H. Chen, and T. Tan. (2025). "Enhance Protein Secretion Pathway and Energy Metabolism to Improve Protein Activity in Recombinant Pichia pastoris". *Biochemical Engineering Journal.* **225** : 1-11. 10.1016/j.bej.2025.109935.

[27] S. Macauley-Patrick, M. L. Fazenda, B. McNeil, and L. M. Harvey. (2005). "Heterologous Protein Production Using the Pichia pastoris Expression System". *Yeast.* **22** (4): 249-270. 10.1002/yea.1208.

[28] C. C. De Queiroz Brito Cunha, A. R. Gama, L. C. Cintra, L. A. M. Bataus, and C. J. Ulhoa. (2018). "Improvement of Bread Making Quality by Supplementation with a Recombinant Xylanase Produced by Pichia pastoris". *PLoS One.* **13** (2): 1-14. 10.1371/journal.pone.0192996.

[29] S. Soleimanpour, H. Farsiani, A. Mosavat, K. Ghazvini, M. R. A. Eydgahi, M. Sankian, H. Sadeghian, Z. Meshkat, and S. A. Rezaee. (2015). "APC Targeting Enhances Immunogenicity of a Novel Multistage Fc-Fusion Tuberculosis Vaccine in Mice". *Applied Microbiology and Biotechnology.* **99** (24): 10467-10480. 10.1007/s00253-015-6952-z.

[30] C. B. Matthews, A. Kuo, K. R. Love, and J. C. Love. (2017). "Development of a General Defined Medium for Pichia pastoris". *Biotechnology and Bioengineering.* **115** (1): 103-113. 10.1002/bit.26440.

[31] E. Žitkus, E. Čiplys, M. Žiaunys, A. Sakalauskas, and R. Slibinskas. (2025). "Development of an Efficient Expression System for Human Chaperone BiP in Pichia pastoris: Production Optimization and Functional Validation". *Microbial Cell Factories.* **24** : 1-15. 10.1186/s12934-025-02679-z.

[32] L. Moruz and L. Käll. (2016). "Peptide Retention Time Prediction". *Mass Spectrometry Reviews.* **36** (5): 615-623. 10.1002/mas.21488.