Construction of a full length α-factor secretory signal sequence for human insulin precursor expression in *Pichia pastoris*

N Utami\(^*\), D Nurdiani\(^*\), Hariyatun\(^*\), E W Putro\(^*\), W Kusharyoto\(^*\)

Research Center for Biotechnology, Indonesian Institute of Sciences, Cibinong Science Center, Cibinong-Bogor 16911, Indonesia

Email: dnurdiani@gmail.com, unuruliaty@gmail.com, wien.kyoto@gmail.com

All authors contributed equally (main contributor) to this work

**Abstract.** *Saccharomyces cerevisiae* and *Pichia pastoris* are yeast known as a potential expression system to produce recombinant protein. The full-length α-factor (α-mating factor) secretory signal of *S. cerevisiae* plays an essential role in the secretion and processing of the mature protein of interest. Here, we attempted to construct the full-length α-factor signal sequence of *S. cerevisiae* in the pD902-IP (Insulin Precursor) expression vector for secreted expression of human insulin precursor in *P. pastoris*. We have isolated the full-length α-factor secretory signal sequence in a pTA2 cloning vector. The full-length α-factor was then inserted into the IP-cassette of pD902-IP and transformed into *E. coli* TOP10. The *E. coli* transformants, which were able to grow on the Zeocin selection medium, harbored the full-length α-factor for the IP expression in the pD902 vector validated by PCR and sequencing. Furthermore, the construct electroporation into *P. pastoris* X-33 was done and followed by IP protein expression confirmation visualized with SDS-PAGE.

1. **Introduction**

The discovery of insulin in the 20\(^{th}\) century is one of the great revelations. Before insulin was available today, people with diabetes had a low life expectancy and prognosis. From Genentech, Inc., David Goeddel et al. first produced human insulin precursors from recombinant DNA (rDNA) in 1978 by combining the insulin A chain and the insulin B chain expressed in *Escherichia coli* in the form of cytoplasmic inclusion bodies [1,2]. However, insulin precursors expressed in the *E. coli* system required recovery and refolding process [3]. Genetically modified proteins can be secreted from the host organism by adding N-terminal signaling secretions [4]. *P. pastoris* is often used to secrete recombinant protein with the pre-pro α-factor leader as a secretory signal originating from *S. cerevisiae* [5]. The α-factor secretion signal consists of 85 amino acids (255 bp nucleic acid) divided into two parts. The first part is the N-terminal signal sequence of 19 amino acids that can direct the translocation of proteins into the endoplasmic reticulum (RE), followed by 66 pro-region amino acids that function to mediate the packaging of receptor-dependent proteins into the RE through the transport vesicles [6].

Several studies used the full-length sequence α-factor leader and short C-peptide (AAK) to secrete proinsulin molecules in the culture supernatant and are known could produce about 3.6 g / L of insulin precursor secretion [7–10]. Meanwhile, studies of HRP protein expression using truncated α-factor (deletions in parts (MATα: Δ30-43) and all parts of the last helical (MATα: Δ57-70)) are known to increase the secretion of the target protein [11]. In another study, it found that the use of short C-peptide...
sequences in the form of Asp-Gly-Lys (DGK) was known to produce higher protein secretion yields than strains using short C-peptide (AAK) [12].

The previous study used a *P. pastoris* strain that utilizes α-factor truncated and short C-peptide (DGK) sequences to secret the insulin precursor protein. The insulin precursor cassette's construction using α-factor truncated as a leader in the 5’ terminal position of the precursor insulin expression cassette with Kex2 cutting site (LEKR) in the terminal carboxy part continued with peptide spacers (EEAEAEAEPK), 29 residues amino acids of insulin B-chain, short linker (DGK), and 21 amino acid residues of insulin A-chain. The result showed the construct could produce expression of insulin precursor protein with a size above 6.5 kDa (~7053 Da) as evidenced by SDS-PAGE analysis of 72-hour supernatant culture with methanol induction of 0.5% [13].

This study attempted to subclone the full-length α-factor signal sequence in the pD902-IP (Insulin Precursor) cassette expression vector from previous research, intending to compare the efficiency of IP secretions between truncated sequences and the full-length one for the next examination.

2. Materials and Methods

2.1. Cloning the full length α-factor secretory signal sequence into *E. coli*

The full-length α-factor secretory signal sequence consisting of 255 bp used in this study was gained from pPICZα A plasmid (Invitrogen™, Cat. no. K1740-01). The fragments of the full-length α-factor were isolated by PCR using HotStarTaq® Master Mix Kit (QIAGEN®, Cat. no. 203443) with α-factor specific primer set ordered from Integrated DNA Technologies, Inc. (IDT) as follows: F primer alpha (5’-AGT CGA ATT CAT GAG ATT TCC TTC AAT TTT-3’) and R Primer alpha (5’-AAA ACT CGA GAG ATA CCC CTT TA-3’) under the following PCR condition: 94°C for 3 min; 35 cycles at 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min; continued with the last period at 72°C for 10 min; then terminated at 4°C for infinite hold. Before cloning, 5μl of the full-length α-factor PCR product was electroporated in 1% agarose gel at 100 volts for 30 min. The gel was soaked in the 0.5 μg/ml ethidium bromide solution (Sigma-Aldrich, E1510) for 15 min and visualized with Gel Doc to ensure full-length α-factor fragments were amplified. PCR product has 3’-overhanging dA at 3’ ends as the result of using the HotStarTaq® Master Mix Kit can be cloned directly into the cloning vector plasmid TArget Clone-Plus TOYOBO (Code No. TAK-101) following the manufacturer’s instructions. The TA cloning vector that inserted of a full-length α-factor (pTA2-α-factor) then cloned in *E. coli* TOP10 by the CaCl₂ heat shock method [14] and plating on blue-white media selection (LB agar plate + Ampicillin 100 μg/ml + IPTG 0.1 mM + X-gal 40 μg/ml). The white colony, which grows on a plate, then picked for colony PCR analysis used α-factor specific primer to confirm that the transformants inserted with the pTA2-α-factor.

2.2. Isolation of the full-length α-factor secretory signal sequence in pTA2 cloning vector

Luria Bertani broth Miller media (Himedia®) with 100 μg/ml Ampicillin (Roche) was used to propagate the transformants overnight at 37°C with shaking. QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany) was used to isolate the pTA2-α-factor. The plasmid was digested with XhoI (NEB, #R0146S) and EcoRI (NEB #R0101S) restriction enzymes to get the compatible full-length α-factor fragments when cloning in a pD902-IP expression vector. After the digest reaction was completed, these were loaded in a 1% agarose gel for electrophoresis to visualize the digested band contained the full-length α-factor. The gel was cut and purified using the QIAquick® Gel Extraction Kit (Qiagen) and quantified the concentration with NanoPhotometer™ (IMPLEN, GmbH).

2.3. Construction of the full-length α-factor secretory signal sequence into a pD902-IP expression vector

The transformants *E. coli* employing the pD902-IP expression vector (3.921 bp) were obtained from the previous study [13]. Firstly, we cultured the pD901-IP clone (*E. coli*) in 10 ml LB broth media supplemented with 100 μg/ml of Zeocin™ at 37°C overnight shaking. Every 5 ml of the culture was
isolated using a QIAprep® Spin Miniprep Isolation Kit to pull the pD9092-IP plasmid. The pD902-IP plasmid has isolated then double digested with XhoI and EcoRI restriction enzymes to separate the backbone part (~3.750 bp) and the truncated α-factor sequence continued by electrophoresis. The agarose gel that holds the backbone pD902-IP fragments are collected and purified with the QIAquick® Gel Extraction Kit (Qiagen), and the concentration was determined. The purified full-length α-factor from the double digest reaction is then linked with the pD902-IP backbone fragment using T4 DNA Ligase (NEB #M0202) follow the manufacturer's ligation protocol.

2.4. Transformation of the pD902-IP-full-length α-factor secretory signal into E. coli TOP10

E. coli TOP10 competent cells are prepared to transform the pD902-IP-full-length α-factor plasmid using the CaCl2 heat shock instructions [14]. About 50–100 µl of the transformants were spread onto Zeocin™ (100 µg/ml) LB Agar plate selection and then incubated for 1–2 days at 37°C. The Zeocin™-resistant transformants are selected. Colony PCR with α-factor specific primer and electrophoresis of the PCR product were conducted due to the importance of confirmation.

2.5. Confirmation of constructing plasmid

The confirmation of constructing plasmid used two approaches by restriction enzyme analysis and plasmid sequencing. The E. coli harboring a pD902-IP-full-length α-factor plasmid was cultured in 10 ml LB broth media supplemented with 100 µg/ml Zeocin™ overnight at 37°C with shaking. Plasmid isolation was executed using the QIAprep® Spin Miniprep Kit. One µg of the isolated pD902-IP-full-length α-factor was digested with XhoI and EcoRI enzymes to confirm the constructed plasmid. The isolated plasmids from two colonies were sent to the 1st BASE for sequencing analysis.

2.6. IP expression from construct plasmid in P. pastoris recombinant clones

The pD902-IP-full-length α-factor cassette was then transformed into P. pastoris X-33 strain by electroporation followed the manual instruction of EasySelect™ Pichia Expression Kit (Invitrogen 2010). Several transformants were grown on 2 ml YPD medium [1% (w/v) yeast extract, 2% (w/v) peptone, and 2% (w/v) D-glucose] supplemented with 100 µg/ml Zeocin™ overnight at 37°C with shaking. The culture then centrifugated at 3000 × g for 5 min at 25°C, the cells transferred into 10 ml buffered glycerol complex medium (BMGY) in a 100 ml flask [1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 × 10⁻³% (w/v) biotin, 1% (v/v) glycerol] and grown for 24 h in shaking incubator at 250 rpm and 30°C. The cells harvested by centrifuging at 3000 × g for 5 min at 25°C and placed to an OD₆₀₀ around 10 in 10 ml buffered methanol complex medium (BMMY) in a 100 ml flask [2% (w/v) peptone, 1% (v/v) yeast extract, 1.34% (w/v) YNB, 100 mM potassium phosphate pH 6.0, 4 × 10⁻³% (w/v) biotin, 2% (v/v) methanol]. The BMMY culture was incubated at 30°C with shaking at 250 rpm. To maintain the induction, a 20% methanol was added to a final concentration of 2% methanol every 24 h. After 72 h induction, the culture centrifugated at max speed for 3 min, the supernatant was collected and keep in -20°C for further analysis. SDS-PAGE was conducted for protein analysis using the Tricine buffer system [13,15].

3. Results and Discussion

3.1. Isolation of the full length α-factor secretory signal sequence

Since the pPICZα A plasmid (3593 bp) has been an expression vector for secreted recombinant protein in P. pastoris, it has been facilitated with a native S. cerevisiae α-factor secretion signal sequence (EasySelect™ Pichia Expression Kit Manual MAN0000042, 2010). Therefore, we utilized this plasmid as a template to gain the full-length α-factor fragments by a PCR reaction using an α-factor specific primer set to be subcloned into the pTA2 cloning vector. The 255 bp of the PCR product of full-length α-factor was detected after electrophoresis, as shown in Figure 1(a).
The fragments of the full-length α-factor obtained from the PCR product, then ligated with the pTA2 cloning vector (2981 bp), applying 1:3 of molar ratio between vector and inserts. The transformation into E. coli competent cells were carried out using the CaCl₂ heat shock method followed by blue-white screening selection. Several formed white colonies were further tested using PCR for confirmation. It is confirmed that the full-length α-factor sequence had been fused with the pTA2 cloning vector, as shown in Figure 1(b).

3.2. Isolation and construction of the full-length α-factor secretory signal sequence into a pD902-IP expression vector

The 30 µl of the pTA2-α-factor plasmid was obtained from every 5 ml of the E.coli transformants cultured in ampicillin LB broth media using the QIAprep® Spin Miniprep Isolation Kit. The concentration of the pTA2-α-factor plasmid was measured with a NanoPhotometer™. One µg of the plasmid has been digested with XhoI (NEB, #R0146S) and EcoRI (NEB #R0101S) restriction enzymes with the setup reaction as follows, plasmid DNA 1 µg; 5 µl of the 10X NEBuffer 3.1; 1 µl of XhoI enzyme (10 units); 1 µl EcoRI enzyme (10 units); and the nuclease-free water up to 50 µl.

Figure 1. (a) Electropherogram of the PCR product of the full-length α-factor from pPICZα A plasmid. NC = negative control (water); LD = GeneRuler 100 bp DNA Ladder; 1 = The full-length α-factor (255 bp). (b) PCR confirmation of pTA2- full-length α-factor E. coli transformants. LD = GeneRuler 100 bp DNA Ladder; PC = positive control (pPICZα A); C1-C5 = clone 1-5; NC = negative control (water).

Figure 2. The difference between the PCR product of the truncated and full-length α-factor secretory. LD = GeneRuler 100 bp DNA Ladder; 1 = full-length α-factor (~255 bp); 2 = truncated α-factor (~171 bp); NC = negative control (water).

Figure 3. Electropherogram of the pTA2-α-factor plasmid digested with XhoI and EcoRI. LD = GeneRuler 100 bp DNA Ladder; 1 & 2 = pTA2-α-factor doubled digested with XhoI and EcoRI; 3 = uncut pTA2-α-factor plasmid.
The restriction enzyme reaction was incubated at 37°C overnight. The gel contained a full-length α-factor secretory band after the electrophoresis, then collected and purified with the QIAquick® Gel Extraction Kit (Figure 3). The plasmid isolation, enzyme digestion, and gel extraction processes have been repeated several times until enough results were obtained. The truncated α-factor has the deletion at the region of the second alpha-helix (MATα: Δ30-43) and the entire last helix (MATα: Δ57-70) so that it has a length of about 171 bp [13]. The truncated and full fragments of the α-factor can be distinguished on the PCR band when visualizing the result's electrophoresis, as shown in Figure 2. The previous study conducted by Nurdiani et al. (2018) performed the pD902-IP expression vector that used a synthetic IP-encoding gene constructed in frame with the truncated α-factor signal and a short C-peptide (DGK) linked A-chain and B-chain of human insulin. The truncated α-factor has a Kex2 endoproteinase cleavage site at its carboxy-terminal part (LEKR).

The pD902-IP plasmid is a yeast integrative vector that facilitated with P_AOX1, a stable and tightly regulated methanol inducible alcohol oxidase promoter in *Pichia pastoris* [16] and contained the ZeoR (Zeocin™-resistant) gene plasmid which acts as a selectable marker [17]. This study tried to replace the truncated α-factor with the full-length one. We removed the truncated α-factor from the pD902-IP regarding subcloning of the full-length α-factor secretory sequence into the pD902-IP backbone. The XhoI and EcoRI enzyme have been applied to separate the pD902-IP backbone with the truncated α-factor sequence, as shown in Figure 4 (a).

The double digest of the pD902-IP result shows that the backbone was produced. In the pD902-IP expression vector, the IP expression cassette is in frame with the AOX_1 promoter sequence. In this case, the EcoRI recognition site 5'-G/AATC-3' appears in front of the Methionine (MATα: Δ1), the first amino acid that arranges the truncated α-factor signal. Differently, the XhoI restriction site 5'-C/TCGAG-3' has existed in the region after amino acid Serine (MATα: Δ53) of the truncated α-factor. Consequently, the truncated α-factor was removed, and the pD902-IP backbone (~3.750 bp) with a sticky end containing the EcoRI and XhoI restriction sites formed (Figure 4 (a)).

The full-length α-factor fragments and the pD901-IP backbone were already collected. Because both have been produced from the cutting process using the same two restriction enzymes, we ligated both with the T4 DNA Ligase enzyme with a molar ratio of 1:3 vector to insert for the indicated DNA sizes. Concurrently, we are preparing the *E. coli* TOP10 competent cells for the transformation using the CaCl₂ heat shock method. After the ligation reaction is completed, we transform the pD902-IP-full-length α-
factor into the *E. coli* TOP10 competent cells. The transformants were continued by plating on LB agar with the addition of 100 µg/ml Zeocin™, then incubated at 37°C for 1 - 2 days. The transformation has successfully produced zeocin-resistant clones. Meanwhile, we did a replating of the 16 colonies formed; subsequently, the colonies' PCR confirmation was done and shown in Figure 4 (b). Sixteen colonies showed positive results with a full-length α-factor characterized by DNA bands appearance with a size of ~255 bp, which correspond with the positive control.

3.3. Confirmation of construct plasmid

Although the PCR colony showed positive results for the full-length α-factor, further confirmation of the plasmid construction is essential. Therefore, we take two confirmation approaches, first, with restriction enzyme digestion, and secondly, with an analysis of plasmid sequencing results. For enzyme digestion, from 16 colonies, we chose six clones as representatives, namely C1, C2, C3, C11, C12, and C13 colonies. Each colony was cultured in LB media with the addition of 100 µg/ml zeocin; then, plasmid isolation was performed. Restriction enzyme reaction using XhoI and EcoRI enzymes applied to each plasmid. Figure 5 shows the results of cutting each plasmid with the addition of the truncated PCR product, full-length α-factor, pTA2-α-factor, and pD902-IP plasmids as controls. It knows that C1, C2, C3, C11, C12, and C13 plasmids produce two fragments after being cut using XhoI and EcoRI enzymes.

![Figure 5](image_url). The double digest confirmation of the representative's plasmid using XhoI and EcoRI enzyme. LD = GeneRuler 1 kb DNA Ladder; NC = negative control (PCR product of truncated α-factor ~171 bp); PC = positive control (PCR product of full-length α-factor ~255 bp); A = uncut pTA2-α-factor plasmid; A+ = digested pTA2-α-factor; B = uncut pD902-IP plasmid; B+ = digested pD902-IP; 1 = uncut C1 plasmid; 2 = digested C1; 3 = uncut C2 plasmid; 4 = digested C2; 5 = uncut C3 plasmid; 6 = digested C3; 7 = uncut C11 plasmid; 8 = digested C11; 9 = uncut C12 plasmid; 10 = digested C12; 11 = uncut C13 plasmid; 12 = digested C13.
Figure 6. The multiple sequence alignment of C3 and C13 using the ClustalW. The truncated α-factor has the deletion at the second alpha-helix region (MATα: Δ30-43) and the entire last helix (MATα: Δ57-70), showed in color letters.

The resulting pieces are 255 bp in size and backbone at ~3.750 bp. Although the 255 bp fragment resulting from the cutting does not look very clear on the electropherogram, the larger piece seen in the lane 2, 4, 6, 8, 10, and 12 after being cut has the same size as the pD902-IP backbone fragment that has digested (B+). The 255 bp fragment is parallel to the small fragment produced from the digested plasmid pTA2-α-factor (A+) (Figure 5.). To ensure that transformants properly contain pD902-IP-full-length α-factor plasmid, we sent C3 and C13 plasmid samples for sequencing using AOX_1 forward primer. The expression cassette's position is in the control of the AOX_1 promoter in the pD902 plasmid construct. Then the results were analyzed with the ClustalW multiple alignment sequence. From the alignment result with the truncated and full-length α-factor reference sequence, it was observed that both the C3...
and C13 plasmids had a 100% similarity in the full α-factor region, which had 255 bp identical nucleotides, as shown in Figure 6.

3.4. IP Expression in P. pastoris

To verify that the construct of pD902-IP-full-length α-factor was integrated into the genome of P. pastoris and could express the human insulin precursor protein, we selected eight Zeocin resistant recombinant clones for expression analysis. The single colony from each clone was grown in a YPD medium for 48h in order to the general growth of the cells, then transferred into the BMGY medium to generate biomass for 24h. The BMGY culture was centrifuged to collect the pellets. The cells were then resuspended in BMMY medium with 2% methanol induction every 2h. The 72h post-induction supernatant from 8 clones were electrophoresed in 15% SDS-PAGE gel using a tricine buffer system (Figure 7).

The SDS-PAGE showed that the representative of eight recombinant clones was confirmed to secrete IP protein into the culture supernatant, a size of ~7 kDa. Figure 7. showed that the pD902-IP-full-length α-factor construct integrated into the P. pastoris genome could express the same insulin precursor into the supernatant culture as the pD902-IP construct. In this study, we chose 2% methanol concentration instead of 0.5% [13] for induction because, based on our follow-up studies, it was found that 2% methanol concentration was also able to express insulin precursors [18,19]. This construct can be used for future studies to compare the efficiency of insulin precursor expression between the truncated α-factor and the full one.

4. Conclusions

Confirmation using restriction enzyme and sequencing results analysis on plasmids convinced that full-length α-factor has been successfully subcloned into pD902 plasmids, which employ precursor insulin cassette that express and secrete the human IP protein into the supernatant.

Acknowledgment

This research is partially supported by Prioritas Nasional (PN) Penyediaan Sumber Bahan Baku Obat Berbasis Protein Rekombinan dan Obat Herbal Terstandar (OHT) untuk Penanganan Penyakit Menular dan Tidak Menular 2019 Research Grant from the Indonesian Institute of Sciences (LIPI) and Pendanaan Riset Inovatif Produktif (RISPRO) Mandatori Lembaga Pengelola Dana Pendidikan (LPDP) bertema Prioritas Riset Nasional (PRN) 2020 Pengembangan Research Cell Bank (RCB) untuk Ekspresi Protein Rekombinan Human Insulin dan Analognya.
References

[1] Goeddel D V, Kleid D G, Bolivar F, Heyneker H L, Yansura D G, Crea R, Hirose T, Kraszewski A, Itakura K, and Riggs A D 1979 USA. Proc Natl. Acad. Sci. 76(1) 106–10

[2] Quianzon C C and Cheikh I 2012 J Community Hos Inter Med Perspect 2(18701) 1–3

[3] Singh A, Upadhyay V, Upadhyay A K, Singh A M, and Panda A K 2015 Microb. Cell. Fact. 14(41) 1–10

[4] Otte S and Barlowe C 2004 Nat. Cell. Biol. 6 1189–94

[5] Barrero J J, Casler J C, Valero F, Ferrer P, and Glick BS 2018 Microb. Cell. Fact. 17(161) 1–13

[6] Kurjan J and Herskowitz I 1982 Cell 30 933–43

[7] Kjeldsen T, Pettersson A F, and Hach Morten 1999 Biotechnol. Appl. Biochem. 29 79–86

[8] Xie T, Liu Q, Xie F, Liu H, and Zhang Y 2008 Prep Biochem Biotechnol 38(3) 308–17

[9] Gurramkonda C, Adnan A, Gäbel T, Lünsdorf H, Ross A, Nemani S K, Swaminathan S, Khanna N, and Rinas U 2009 Microb. Cell. Fact. (8) 13 1–8

[10] Polez S, Origi D, Zahariev S, Guarnaccia C, Tisminetzky S G, Skoko N, and Baralle M 2016 PLoS ONE 11(12) 1–15

[11] Lin-Cereghino G P, Stark C M, Kim D, Chang J, Shaheen N, Poerwanto H, Agari K, Moua P, Low L K, Tran N, et al. 2013 Gene 519(2) 311–17

[12] Kjeldsen T 2000 Appl. Microbiol. Biotechnol. 54 277–86

[13] Nurdiani D, Haryati H, and Kusharyoto W 2018 Indones. J. Biotechnol. 23(2) 102–8

[14] Sambrook J and Russel D 2001 Molecular Cloning: A Laboratory Manual (NY: Cold Spring Harbor Laboratory Press) pp 6.4–6.11

[15] Haider S R, Reid H J, and Sharp B L 2012 Protein Electrophoresis: Methods and Protocols, Methods in Molecular Biology vol 869, ed B T Kurien and R H Scofield (Springer Science+Business Media, LLC) chapter 8 pp 81–91

[16] Wang J, Wang X, Shi L, Qi F, Zhang P, Zhang Y, Zhou X, Song Z, and Cai M 2017 Sci. Rep. 7(41850) 1–12

[17] Higgins D R, Busser K, Comiskey J, Whittier P S, Purcell T J, and Hoeffler J P 1998 Methods. Mol. Biol. 103 41–53

[18] Nurdiani D, Haryati H, Utami N, Putro E W, and Kusharyoto W 2020 IOP Conf. Ser.: Earth Environ. Sci. 439(012017) 1–8

[19] Nurdiani D, Haryati H, Utami N, Putro E W, and Kusharyoto W 2020 IOP Conf. Ser.: Earth Environ. Sci. 572(012008) 1–8